Demystifying thermal comfort in smart buildings : an interpretable machine learning approach

Zhang, Wei; Wen, Yonggang; Tseng, King Jet; Jin, Guangyu

2020

Zhang, W., Wen, Y., Tseng, K. J. & Jin, G. (2020). Demystifying thermal comfort in smart buildings : an interpretable machine learning approach. IEEE Internet of Things Journal, 8(10), 8021-8031. <https://dx.doi.org/10.1109/JIOT.2020.3042783>

<https://hdl.handle.net/10356/152739>

<https://doi.org/10.1109/JIOT.2020.3042783>

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: <https://doi.org/10.1109/JIOT.2020.3042783>.

Downloaded on 26 Apr 2023 20:58:32 SGT

Demystifying Thermal Comfort in Smart Buildings: An Interpretable Machine Learning Approach

Wei Zhang, *Member, IEEE*, Yonggang Wen, *Fellow, IEEE*, King Jet Tseng, and Guangyu Jin

Abstract—Thermal comfort is a key consideration in smart buildings and a number of comfort models are available nowadays to evaluate the comfort level of occupants. However, the models are often complex and hardly interpretable for the developers and operators. Indeed, the model interpretations are beneficial in multifold such as for system inspection and optimization. In this paper, we propose an interpretable thermal comfort system to introduce interpretability to any black-box comfort models. First, we focus on the relationship between a model's input features and output comfort level. The feature impact on comfort is investigated and the impact patterns are shown to be diverse for different features. Second, we unveil the model mechanisms about the data processing inside the model by building the model surrogates based on the interpretable machine learning algorithms. The surrogates offer outstanding fidelity for simulating the actual model mechanisms and the interpretations based on the surrogates are intuitive and informative. Our interpretable comfort system can be integrated with the existing building management systems. Accordingly, we can ease building owner's concerns about adopting new black-box technologies and enable various smart building applications like smart energy management.

Index Terms—Thermal comfort; smart building; smart city; deep learning; interpretable machine learning.

I. INTRODUCTION

THERMAL comfort is an essential consideration of smart buildings. The impact of thermal comfort can be either positive or negative. Satisfied comfort ameliorates the occupant's well-being, happiness, and working efficiency [1], [2], while discomfort results in the sick building syndrome, absenteeism, and degraded cognitive performance [3]. Especially, modern people spend as much as 90% of their time in buildings [4]. As such, the Internet of things (IoT) system

with sensing and analyzing capabilities is deployed in smart buildings to monitor the comfort-related features and generate data. Here, the comfort models come in to utilize the feature data to calculate and evaluate the degree of occupant comfort. Various comfort models are available nowadays, based on either the traditional statistical analysis [1] or the latest machine learning (ML) technologies [5], [6], [7].

Despite the well progressed and established research on thermal comfort modeling, the comfort models are often mystery boxes to end-users. Consider the possibly most popular one, the predicted mean vote (PMV) comfort model. It has been widely used for decades and adopted by the standards like ASHRAE 55 [8]. The Python-based PMV toolkit [9] takes around a hundred lines of code full of equations and numerical values only for the basic version. The implementation is barely interpretable except for a small group of domain experts. The later amendments and revisions, i.e., supporting a wider range of air velocity [10], further inflate the code size manifold and deteriorate the model's interpretability. The phenomenal is not atypical and the deficient interpretability is generic for the other comfort models as well, i.e., the ones based on the black-box neural networks.

Interpretability should be a fundamental requirement for comfort models. It introduces transparency to the models and enables human-in-the-loop development and operation. Interpretability unveils the model mechanisms to the model developers, who accordingly can inspect the model correctness, detect potential design flaws, and enhance the imperfect parts. The advantage also extends to the building operators, who can thereby access the intuitive model causality, understand the rationality, and optimize the building control. To unleash the potential of interpretability for smart buildings, a critical mission is to explore the beneficial but missing interpretability of the existing comfort models.

Recent advancement of interpretability largely emerges from the ML community [11], partially due to the fact that many ML models are black-box in nature and thus uninterpretable. One option is to target the model input/output (I/O) to interpret the feature impact on the model output. The interpretation can be either coarse-grained global feature importance or fine-grained feature interaction and condition-sensitive feature impact [12]. Another option is to dive into the model internal to investigate the model dynamics and mechanisms, i.e., how data is flowed and processed in the model. While the model itself is uninterpretable, the key concept is to simulate the model with an interpretable surrogate and then interpret the surrogate instead, given the surrogate is accurate enough [13]. In this paper, we aim to radiate the interpretable

Manuscript received XXX; revised XXX; accepted XXX. This research is funded by National Research Foundation (NRF) via the Green Buildings Innovation Cluster (Grant NO.: NRF2015ENC_GBICRD001-012), administered by Building and Construction Authority (BCA) Singapore. In addition, this research is sponsored by National Research Foundation (NRF) via the Behavioural Studies in Energy, Water, Waste and Transportation Sectors (Grant NO.: BSEWWT2017_2_06), administered by National University of Singapore (NUS). Moreover, this research is funded by Nanyang Technological University (NTU) via the Data Science & Artificial Intelligence Research Centre @ NTU (Grant NO.: DSAIR@NTU). (Corresponding author: W. Zhang.)

W. Zhang is with the Infocomm Technology Cluster, Singapore Institute of Technology (SIT), Singapore 567739 (e-mail: wei.zhang@singaporetech.edu.sg.)

Y. Wen is with the School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore 639798 (e-mail: ygwen@ntu.edu.sg.)

K. Tseng is with the Engineering Cluster, Singapore Institute of Technology (SIT), Singapore 579700 (e-mail: kingjet.tseng@singaporetech.edu.sg.)

G. Jin is with the Building and Construction Authority (BCA), Singapore 579700 (e-mail: jin_guang_yu@bca.gov.sg.)

ML advancement to the thermal comfort interpretations, and our methodology is briefly described as follows.

We propose an interpretable thermal comfort system to interpret comfort models about the model I/O and internal mechanisms. For the former, we use the partial dependence plot (PDP) [12] to assess the marginal effect of every feature on thermal comfort, where each plot illustrates the comfort change with varying feature input. We also adopt the emerging SHAP values [14], derived from the game-theory based Shapley values [15], to evaluate each feature's localized contribution to the comfort output. Thereafter, we move to the internal model mechanisms. We use two prevalent interpretable ML algorithms, linear regression (LR) and decision tree (DT), to construct the surrogates of the comfort models. We elevate the surrogate fidelity through local subspace learning and subsequently, we interpret the surrogates accurate enough to derive the interpretations of the comfort models. Overall, this paper achieves the following main contributions.

- We show that the feature impact to thermal comfort is diverse and it can be positive or negative, linear or nonlinear for different features. Also, the impact varies under different ambient and vital conditions.
- We reveal that surrogate approximation in the global feature space is hard. Meanwhile, local subspace learning is effective for achieving superb surrogate fidelity with the R^2 of up to 0.99.
- We present the interpretations of the surrogates through visualization and description. We also demonstrate the optimized building control inspired by the interpretations.

Our interpretable comfort system can be integrated with the existing building management system to enhance occupant comfort and facilitate system development and operation. First, interpretability is vital to ease the building owner's concern about the adoption of any new black-box technologies. Also, the comfort system is a pillar to support other building components such as facade design and energy-efficiency control. Our interpretability concept for thermal comfort also sheds light on other building applications like energy consumption modeling and air conditioner maintenance. In general, our work can propel the wide-scale adoption of interpretable modules into building systems.

The rest of this paper is organized as follows. We describe the system architecture and datasets in Section II. Section III investigates the feature impact on the comfort. Section IV presents the LR-based comfort model surrogates and the corresponding interpretations. In Section V, we also discuss the DT-based surrogates together with the surrogate interpretations. Finally, Section VI concludes this paper and suggests several future research directions.

II. SYSTEM ARCHITECTURE AND DATASET

In this section, we first present the system architecture of the interpretable comfort system. Then, we introduce the comfort model and the datasets for this research.

A. System Architecture

The system architecture of the interpretable thermal comfort system is illustrated in Fig. 1. The system consists of three

Fig. 1. The system architecture of the interpretable thermal comfort system. The three modules in this system are data generation, interpretable learning, and smart building applications.

major modules. The first one is for dataset generation, where the feature data (e.g., temperature and humidity) either synthetic or collected by the smart building IoT facilities serves as the input of an existing comfort model to be interpreted. The comfort model digests the feature data and produces the corresponding comfort index (e.g., hot and cold). Combining the input feature data and the output comfort index, we can generate the data for the interpretable ML study, to be covered in the following module.

The key research focus of this paper is on the middle module for the interpretable ML. With the generated data, this module works on interpreting the system mechanisms of the obscure comfort model. The interpretability module can interpret the comfort model both in terms of the model I/O and internal mechanisms. The former analyzes how each feature impacts the comfort and the latter answers how features are processed inside the comfort model. Candidate technologies can be PDP and SHAP values for feature impact analysis and interpretable ML-based surrogate analysis using LR and DT for model mechanisms analysis. Finally, this module provides insights and interpretations of the comfort model to various types of building applications in the next module.

The last module refers to the model interpretations to support various smart building applications, i.e., decision making and system optimization to improve energy efficiency. Besides, the achievements of this module can be returned to the first module to update and refine the comfort model and form an iterative optimization loop of the whole system.

B. Comfort Model

Our proposed interpretable thermal comfort system is model agnostic and can interpret any given models. Here, to facilitate the description and demonstration of our system, we choose to interpret the classic PMV comfort model [16], which should be the most popular and widely-used comfort model for decades. The PMV model assumes that thermal comfort is determined by six comfort features, including four ambient ones, air temperature, mean radiant temperature (MRT), relative humidity (RH) and air velocity¹, and two vital ones, metabolic rate and cloth insulation. Given a vector of the feature values, the PMV model calculates the comfort index within $[-3, 3]$, where

¹For simplicity, we use temperature and air temperature interchangeably in the rest of the paper, and so do velocity and air velocity.

TABLE I

DATASETS WITH DIFFERENT FEATURE INPUT RANGES, WITH CELSIUS FOR TEMPERATURE AND MRT, PERCENTAGE FOR RH, M/S FOR AIR VELOCITY, MET FOR METABOLIC RATE, AND CLO UNIT FOR CLOTH INSULATION.

Name	Temp./MRT	RH	Vel.	Met.	Clo.
GLOBAL	[0, 30]	[0, 100]	[0, 1.0]	[1, 5]	[0.36, 1.37]
LOC-FULL	[0, 30]	[0, 100]	[0, 1.0]	1.5	0.61
LOC-THV	[22, 28]	[30, 70]	[0, 0.2]	1.5	0.61
LOC-TH	[22, 28]	[30, 70]	0.1	1.5	0.61

the best comfort attains at zero and the large absolute value indicates strong discomfort. Specifically, the seven integers $-3, -2, \dots, 3$ within the range indicate cold, cool, slightly cool, comfortable, slightly warm, warm, and hot, respectively.

Like most nowadays' comfort models, the PMV model is hardly interpretable. In this paper, we introduce interpretability to the PMV model and we use data as a bridge between the comfort model and interpretations.

C. Dataset

In this part, we describe the details of the dataset generation for interpretable thermal comfort. The PMV model is used to compute the comfort index for each feature input and the implementation of the PMV model is based on an open-source Python library called Ladybug Tools [9].

Without loss of generality, let $X = \{x_1, x_2, \dots, x_n\}$ denote a set of n feature input vectors. For any $1 \leq i \leq n$, $x_i = x_{i,1}, x_{i,2}, \dots, x_{i,m}$ where each input vector has m features and $x_{i,j}$ is the value of the j -th feature of x_i . Specifically for the PMV model, we have $m = 6$ and the six features in order are temperature, MRT, RH, velocity, metabolic rate, and cloth insulation. Let $f(x_i)$ be the comfort index corresponding to the feature input x_i where $f(\cdot)$ is the comfort model. The dataset can be defined accordingly as $D = \{d_1, d_2, \dots, d_n\}$ where $d_i = (x_i, f(x_i))$ is the i -th data sample of the dataset.

We generate four datasets with different feature input spaces for our interpretability research and a summary of the datasets is shown in Table I. The first one captures the global space of the comfort model by specifying large value ranges for the feature input for typical buildings. Specifically, we generate random numbers within $[0, 30]$ for temperature and MRT in Celsius degree, $[0, 100]$ for RH in percentage, $[0, 1]$ for velocity in m/s, $[1, 5]$ for metabolic rate in MET, and $[0.36, 1.37]$ for cloth insulation in clo. The ranges above are even beyond the realistic settings, i.e., the metabolic rate is from minimal activity to intense activities like playing tennis [17] and the cloth insulation is from shorts to winter garments [18]. We generate one million (10^6) input vectors and calculate the resulting comfort index of each input. In total, we produce a global dataset with 10^6 samples, simply denoted as GLOBAL.

The rest three datasets capture local subspace and we especially generate the datasets from the perspective of smart building control. While the values in the global space are theoretically valid, the controllable space, in reality, is much smaller. In a typical building setting, the metabolic rate and cloth insulation remain mostly unchanged throughout a day

and are beyond the control of the building operators. Besides, the MRT fluctuates around the temperature and the fluctuation is often assumed to be negligible. Accordingly, we can reduce the six-dimensional global space into a more realistic three-dimensional local subspace, where metabolic rate and cloth insulation are constants and MRT equals the temperature.

Here, we consider a summer office for system demonstration with 1.5MET office work and 0.61 summer cloth insulation and generate three local datasets in different local subspace. The first one denoted as LOC-FULL adopts the same $[0, 30]$, $[0, 100]$ and $[0, 1]$ ranges for temperature, RH, and velocity as for the global dataset. The rest two simulate a use case that the building operator wants to interpret the comfort model in the subspace around the current IoT-monitored ambient situation. Here, we investigate a baseline situation with 25°C temperature, 50% RH, and 0.1m/s velocity for demonstration. Specifically, the local dataset LOC-THV assumes that temperature, RH, and velocity are controllable and explores the subspace of $\pm 3^\circ\text{C}$, $\pm 20\%$, and $\pm 0.1\text{m/s}$, respectively. The last dataset LOC-TH is a variant of LOC-THV assuming that velocity is nonadjustable. Same as the global dataset, each local dataset has 10^6 data samples. In the following sections, we demonstrate and analyze our interpretable thermal comfort system with the four generated datasets.

III. FEATURE IMPACT ON THERMAL COMFORT

In this section, we interpret the model I/O and create a direct link between the features and comfort. We first adopt PDP to evaluate the change of feature impact on comfort with varying feature values. Then, we apply SHAP values to assess the localized contribution of each feature to the comfort.

A. Feature Impact Analysis with PDP

A PDP visualizes the marginal effect of a feature to the comfort index [12], and it is capable to exhibit complex effects. To analyze the marginal effect of a feature, PDP creates a series of dataset variants based on different feature values, i.e., in ascending order. Consider a feature and a specific feature value. The corresponding dataset variant is derived by overriding the feature values of all the data samples in the original dataset with the specific feature value and keeping the other values unchanged. The average comfort index of the dataset variant indicates the marginal effect of the feature at the specific value. For example, the different temperature values can be $0.0, 0.1, \dots, 30.0^\circ\text{C}$, and each value has a corresponding dataset variant. The average comfort index based on the dataset variant of 25.0°C indicates the marginal effect of the temperature feature at 25.0°C . Performing the above operations by iterating different values of a feature, PDP can visualize the change of the thermal comfort when the feature value varies. Specifically, the PDP function $P_j(\cdot)$ of the j -th feature for $1 \leq j \leq m$ at value v can be mathematically formulated as follows.

$$P_j(v) = \frac{1}{n} \sum_{i=1}^n f(x'_i), \quad (1)$$

Fig. 2. The PDPs of the impact of the six features on PMV-based comfort with the dataset GLOBAL. Three features (e.g., temperature, MRT, and RH) show the linear relationship with the comfort index, and the rest three (e.g., velocity, metabolic rate, and cloth insulation) exhibit more complex relationships.

where $x'_i = x_{i,1}, \dots, x_{i,j-1}, v, x_{i,j+1}, \dots, x_{i,m}$ by replacing $x_{i,j}$ for every data sample x_i in the original dataset with the fixed value v and with the comfort model $f(\cdot)$.

In this part, we use the dataset GLOBAL to generate the PDPs for the six features of the PMV model. The plots are shown in Fig. 2 and let us describe them one by one.

1) *Air Temperature*: As seen from the figure, the increase of the comfort index is approximately linear in temperature, where the slope of the approximated line is 0.118. In general, temperature shows a huge impact on the comfort index, which varies significantly from cool to warm for the temperature from 0°C to 30°C . This justifies the fact that many comfort control systems nowadays are temperature-based.

2) *MRT*: The radiant heat transfer from walls and windows to the occupants is modeled by MRT. Generally, the relationship between MRT and the comfort index is linear as well and they are positively correlated. However, within the same range $[0, 30]^\circ\text{C}$ as temperature, the comfort change is much less dramatic. Specifically, the slope of the approximated line is 0.039, which is only around one-third of the slope for temperature. The interpretation here is that for every 1°C temperature increase (decrease), the impact on thermal comfort is equivalent to $\sim 3^\circ\text{C}$ MRT increase (decrease). As MRT more or less equals to temperature and shows a relatively low impact on comfort, it has not been one of the key considerations of most building systems.

Should MRT be controlled to improve comfort, one option is to implement an intelligent curtain system. MRT is largely related to solar irradiance, and it is why we often feel warm

beside a window and cold in an enclosed concrete space. So, given the warm thermal sensation, we may close the curtains to block the solar irradiance. When the thermal sensation is cool, the curtains can be opened up a bit to allow some sunshine to go through the windows, so that we can feel warmer. The optimal control scheme is beyond the scope of this research, but it is interesting to receive further attention in future works.

3) *RH*: Another feature with a linear impact on comfort is RH, but its degree of impact is even lower than that of MRT. Even for the full possible range of RH from 0% to 100%, the change of the comfort index is less than 0.5, i.e., half of the thermal sensation change from comfortable to slightly warm. The slope of the approximated line is 0.005, which is only 12% and 4% of the slopes for MRT and temperature, respectively. Even the RH is changed unrealistically from the absolutely dry (0%) to absolutely humid (100%), the comfort effect can be offset by 4.1°C temperature change. In a typical building, the variation of RH shall be within 40 – 60%. The maximum comfort effect with such a range is equivalent to the change of temperature by merely 0.8°C . Despite the little impact on comfort, RH is often a variable to optimize in the building systems, mostly because it is more controllable than MRT and the potential win-win situation with both comfortableness and energy efficiency.

4) *Air Velocity*: While linear effect applies to the first three features, it does not hold for the rest ones and velocity is one of them. Until 0.1m/s , the change of velocity's impact on comfort is negligible, with ~ 0.005 comfort index difference. Increasing the air movement beyond 0.1m/s , the thermal sensation quickly turns to be cooler, and the comfort index keeps decreasing until 1.0m/s . For example, the average comfort index is 0.85, -0.10 and -0.50 for 0.1m/s , 0.5m/s and 1.0m/s velocity, respectively. Note that such a decrease is convex and the gain of cooling diminishes with the increase of air movement. Accelerating the airflow from 0.1m/s to 0.5m/s , we observe 0.95 comfort index change, but doubling the acceleration to 0.9m/s , only 0.34 additional comfort change can be achieved. Overall, velocity control is relatively easy and energy-efficient (by using fans), so it has been a popular control variable in the building systems.

However, air movement too intense can be disturbing for occupants, and velocity shall be maintained below a certain threshold, i.e., 0.5m/s , with satisfied comfortableness. So, the control range is quite narrow, say 0 to 0.5m/s and as a result, the impact on comfort is bounded. Also the attenuation of the airflow intensity is significant, and the velocity decreases fast heterogeneously from the source fan to occupants. As often only the source or fan is controllable, uncertainty does exist for the airflow distribution to different occupants and advanced control methods shall be developed.

5) *Metabolic Rate*: Same as the other features except for velocity, metabolic rate shows a positive correlation with comfort, where higher metabolic rate results in warmer sensation. A low rate like 1MET can even lead to a cold sensation in our dataset. For a typical office setting with 1.5MET, the comfort improves but is still in the middle between slightly cool and cool. Besides, the effect is concave in metabolic rate till 2MET, corresponding to the light intensity activities like

walking slowly. Beyond 2MET, the comfort index changes linearly. As described earlier, it is not realistic to control the metabolic rate of the occupants by the building operators. But the effect can guide the building control in a region-based fashion, i.e., high metabolic rate in a gym and thus low setpoint temperature.

6) *Cloth Insulation*: The last investigated feature is cloth insulation, which exhibits a positive comfort impact as well and the effect is nonlinear. Insulation plays a more important role in comfort effect when it is small, but as it becomes large the effect change diminishes. For example, the comfort index changes by 1.36 for the first half of the tested insulation range, and the change narrows down to 0.70 for the second half. Normally, cloth insulation is assumed to be unchanged throughout a day, i.e., in the office, so its comfort impact is uncontrollable within a short duration like a few hours. But low insulation does help achieve a cool sensation, and the impact is more remarkable for lower insulation. Indeed, cooling energy can be saved if cloth insulation can be adjusted. One option is to advocate the long-term behavior change from a social science perspective. For example, the Cool UN initiative encourages the UN employees to adopt a casual dress code [19]. As a result, the energy usage for cooling the UN Headquarters is reduced with a bit higher but acceptable room temperature.

In summary, the marginal effect of the features on comfort is diverse. Most features show a positive effect where the comfort index grows in the feature value, but there exists a negative effect as well like high-velocity results in low comfort index. The effect can be either linear or nonlinear with each feature. The linear effect applies to temperature, MRT, and RH; and for the rest features like velocity, the effects are more complex. The interpretations enable the building operator to obtain an intuitive and comprehensive understanding of the system and accordingly perform effective system optimization.

B. Feature Impact Analysis with SHAP Values

In this part, we use SHAP values [14] to measure the contribution of each feature to a specific comfort index output. SHAP values emerge from the game-theory based Shapley values [15], where the players of the game are the features in our study. A key assumption of the SHAP values is that any model output is the consequence of the collaborative effort of a group of features. To quantify the contribution of each feature, the comparative analysis is conducted between the cases with and without the feature in the modeling process. Specifically, given a feature input x and the j -th feature, the corresponding SHAP value $S_j(x)$ can be computed as follows.

$$S_j(x) = \sum_{z' \subseteq x'} \frac{|z'|!(m - |z'| - 1)!}{m!} (f_x(z') - f_x(z' \setminus j)), \quad (2)$$

where x' is a simplified input of x such that $x = h_x(x')$ for a mapping function $h_x(\cdot)$, $z' \subseteq x'$ includes all the z' vectors where the non-zero features are a subset of the non-zero features in x' and $|z'|$ is the number of non-zero features in z' , $f_x(z') = f(h_x(z')) = \mathbb{E}[f(z)|z_Z]$ for a thermal model $f(\cdot)$ where Z is the set of non-zero features in z' , and $z' \setminus j$ means

Fig. 3. The SHAP values with our baseline feature setting using the global dataset as well as the LOC-THV and LOC-TH datasets. Negative (positive) value means that the feature contributes to achieve higher (lower) comfort index compared to the average comfort.

setting the j -th feature to zero. As shown above, computing the SHAP values can be computationally intensive. Some approximation solutions to accelerate the computation can be found in [14].

Fig. 3 shows the SHAP values for a baseline setting of 25°C temperature and MRT, 50% RH, 0.1m/s velocity, 1.5MET metabolic rate, and 0.61 cloth insulation on the global dataset and two local datasets LOC-THV and LOC-TH. We describe the results as follows.

1) *Global Dataset*: Overall, the baseline setting leads to a slightly warm sensation. When the global dataset is considered, the baseline values of four features (e.g., temperature, MRT, RH, and velocity) contribute to pushing up the comfort index by 0.15 from the average index of the dataset. On the contrary, the baseline values of the rest two (e.g., metabolic rate and cloth insulation) offset part of the increase. Specifically, the expected comfort index with the global dataset would be 0.08 lower should the 25°C temperature be not configured. Similarly, the 25°C MRT and 0.1m/s velocity contribute to driving up the index by 0.04 and 0.03, respectively. RH also plays a part in increasing the index, but its contribution is relatively negligible (< 0.003). In a reverse direction, the 1.5 metabolic rate and 0.61 cloth insulation pull the comfort index towards the cool direction by 0.09 and 0.03, respectively, together contributing to ease of the uncomfortably warm sensation by 0.12.

2) *Local Datasets*: The two local datasets focus on the controllable features, where the MRT equals to temperature and metabolic rate together with cloth insulation do not change much throughout a day. Thereby, the metabolic rate and insulation do not contribute to differentiate the comfort index and the MRT's impact is transferred to the temperature's SHAP values. For LOC-THV, the comfort index of the baseline setting is 0.38 higher than the average value derived by the dataset. Temperature is the more significant contributor, showing 85% of the impact. Air velocity shows a less significant impact, contributing 15%, and RH shows an even lower impact ($< 0.2\%$).

LOC-TH considers an even smaller subspace than LOC-THV by assuming a nonadjustable velocity. In such a subspace, the impact of temperature is more dominant than that of LOC-THV, with a SHAP value of 0.36 contributing to a warmer environment. RH also contributes in the same direction, but

the impact is minimal, much below 0.001.

The results of the local datasets may imply that there is not much the building operators can do for adjusting comfort except controlling temperature for the given baseline setting. But note that SHAP values only reflect the contribution for a specific output corresponding to the baseline setting and the values are not necessarily the same for different settings. In the following sections, we will show that the features like RH and velocity can also be utilized to manage the thermal comfortableness with fine-grained building control under certain conditions.

IV. INTERPRETABLE LR-BASED COMFORT SURROGATE

We dive deeper and unveil the unknown and potentially complex internal mechanisms. Our solution is to build up a surrogate of the comfort model, such that the surrogate itself is intuitive and interpretable, and can approximate the model with good fidelity. LR and DT are possibly the most commonly used interpretable ML models, so we would like to interpret the comfort models using both LR and DT. Specifically, we consider thermal comfort modeling as a regression problem and also consider it as a classification problem. For the former, we aim to output a continuous value, or float value, as the comfort index. In this case, it is essentially a regression problem and we use LR, a popular and interpretable regression algorithm, to model thermal comfort. For the latter, we want to map the comfort into some discrete categories, or integers. In this case, it is a classification problem, where the data samples are grouped into different disjoint sets, each with a unique class label. And we use DT, a well-known classification algorithm to model and interpret thermal comfort.

In this section, we first adopt LR to produce the comfort model surrogates. Thereafter, we interpret the surrogates to deliver interpretations about the model's internal mechanisms.

An LR surrogate is mathematically formulated as follows. Given a set of n feature inputs X and a comfort model $f(\cdot)$. Let $f(x_i)$ be the comfort index (continuous) of the feature input $x_i \in X$ for $1 \leq i \leq n$. The surrogate aims to approximate $f(\cdot)$ with a linear function $\hat{f}(\cdot)$ as follows.

$$\hat{f}(x_i) = \omega_0 + \sum_{j=1}^m \omega_j x_{i,j}, \quad (3)$$

where $\omega = \omega_0, \omega_1, \dots, \omega_m$ is the coefficients of $\hat{f}(\cdot)$. Specifically, ω_0 is the intercept and ω_j is the coefficient of feature j . The optimal ω is derived with the least-squares as follows.

$$\omega = \arg \min_{\omega'} \sum_{i=1}^n \left(\omega'_0 + \sum_{j=1}^m \omega'_j x_{i,j} - f(x_i) \right)^2. \quad (4)$$

We first investigate the fidelity of the LR-based surrogate in terms of accuracy and stability to establish confidence in our surrogate-based interpretations.

A. Surrogate Fidelity

The four datasets are generated in different feature spaces of the comfort model. Our intuition is that it is challenging to

Fig. 4. The fidelity of the LR-based surrogates in the global feature space and different local subspace with the four datasets. Fig. 4(a) and Fig. 4(b) show the statistical results of MAE and R^2 , respectively. Local subspace learning helps improve the surrogate fidelity.

approximate the mechanisms in the global space. This is especially true when we plan to approximate with the interpretable ML algorithms, which often suffer from subpar approximation capability compared to the black-box ones like deep learning. Thus, our basic idea is to target a more manageable, or smaller, feature space, similar to the idea of calculating the area of a circle. While the mechanism in the global space can be complicated, it is expected to be less complex in the subspace where we tend to achieve good fidelity.

In this part, we build four surrogates for the four datasets to verify our idea on local subspace learning, and the surrogate performance with 10 independent runs is shown in Fig. 4.

1) *Accuracy*: As seen in Fig. 4, surrogate fidelity improves significantly by the local subspace learning. The mean absolute error (MAE) is 0.28 for the global surrogate and it is reduced significantly by 64% below 0.10 for the local surrogates. Among the three local ones, the MAE improves when local subspace shrinks, i.e., 0.04 for LOC-THV's 3-dimensional subspace and 0.01 for LOC-TH's 2-dimensional subspace. The R^2 metric shows similar patterns, where local surrogates are more accurate than the global one. Especially for the smallest local subspace with LOC-TH, the average R^2 hits 0.99, with near-optimal fidelity.

2) *Stability*: Generally, the four surrogates perform good on stability and the standard deviations (SDs) of the MAE and R^2 are all below one-thousandth. Nevertheless, the local surrogates show better stability with even smaller SDs. For LOC-FULL, LOC-THV, and LOC-TH, the MAE SDs are only 25%, 10%, and 1% of the global surrogate's SD, and the R^2 SDs are 16%, 19%, and $< 1\%$, respectively. Same as the conclusions for accuracy, a small subspace helps achieve good fidelity as well from the perspective of stability.

3) *Data Size*: We would like to discuss a bit about the impact of data size. In our tests, the performance variation becomes minimal as long as a few percent of the data is used. Performance gain can be observed with more data, but the gain is almost negligible. For time usage, it is manageable as our system does not have high requirements on the data amount and meanwhile the data generation can be highly paralleled.

4) *Discussion*: Overall, we demonstrate the advantage of local subspace learning and we achieve remarkable fidelity in small subspace like for the LOC-TH surrogate. From a building control point of view, we often want to predict the

TABLE II
PARAMETERS AND PERFORMANCE OF THE LR-BASED SURROGATES.

GLOBAL	Temp.	RH	Vel.	MRT	Met.	Clo.	
Mean	0.109	0.470	−1.415	0.038	0.858	1.819	
SD	3e−5	7e−4	9e−4	3e−5	1e−4	6e−4	
GLOBAL				Intercept	MAE	R^2	
Mean				−5.911	0.281	0.927	
SD				0.001	4e−4	3e−4	
LOC-FULL	Temp.	RH	Vel.	Intercept	MAE	R^2	
Mean	0.226	0.507	−1.237	−5.198	0.099	0.990	
SD	1e−5	2e−4	2e−4	4e−4	1e−4	3e−5	
LOC-THV	Temp.	RH	Vel.	Intercept	MAE	R^2	
Mean	0.221	0.600	−1.199	−5.045	0.036	0.989	
SD	1e−5	2e−4	4e−4	4e−4	3e−5	3e−5	
LOC-TH	Temp.	RH			Intercept	MAE	R^2
Mean	0.219	0.593			−5.094	0.006	1.000
SD	2e−6	5e−5			5e−5	1e−5	1e−6

outcome of making certain adjustments around the current configuration. Such adjustment shall be gradual to avoid the abrupt ambient environment change, which in turn leads to occupant discomfort. To this end, local subspace learning not only offers good surrogate fidelity but also align with the building control concept with gradual control change. In the following part, we interpret the surrogates to offer interpretations of the comfort model.

B. Interpretations for the Surrogates

The LR parameters and performance of the four surrogates are shown in Table II. The table offers clear interpretations about the interaction between the features and comfort and the interaction is assumed to be linear by LR. Such a linear relationship is vital for practical deployment of the interpretable systems, as the interpretations consist of only constant coefficients that can be visualized to and easily interpreted by the building operators.

1) *Coefficient Stability*: Table II confirms the fidelity advantage when a small learning space is targeted. Here, we further demonstrate that the advantage also extends to the coefficients of the LR surrogates. Compared to LOC-FULL, LOC-TH with a smaller subspace is 5x, 4x, and 8x more stable in terms of the coefficients of temperature, RH, and intercept, respectively. The improvements are more significant when comparing with GLOBAL. Stability essentially measures the robustness of a system. It is critical for building control, as operators can gain confidence about the system adjustments to be made and the corresponding outcome to avoid unexpected consequences.

In the following part, we interpret the local surrogates of LOC-THV and LOC-TH from the building control perspective.

2) *LOC-THV*: Compared to the global surrogate, temperature and RH become more important for the LOC-THV surrogate with 103% and 28% higher coefficients, respectively. Meanwhile, the impact of air movement diminishes by 14% when we limit the velocity change within ± 0.1 m/s. Note that

the LOC-THV surrogate has better fidelity than the global one, so the system optimization can be misled by the global surrogate. For example, while the less accurate global surrogate shows that 1°C temperature difference can be compensated with 0.07m/s velocity change, a much larger velocity of 0.2m/s is needed in reality based on the more accurate surrogate for LOC-THV. Besides, RH does not impact comfort much, i.e., the comfort index only varies by 0.12 even by changing the humidity substantially from 50% to 30% or 70%. This could be a helpful insight as humidity has a significant impact on energy usage [20], i.e., humidity change for big energy saving with negligible discomfort incurred.

3) *LOC-TH*: Only temperature and RH are controllable in the LOC-TH surrogate, so we derive the coefficient for both features. Seen from Table II, LOC-TH surrogate shares similar coefficients as LOC-THV, where the differences are only 0.002 and 0.007, respectively, for temperature and RH. While the interpretations of LOC-THV surrogate mostly hold for LOC-TH, the latter does have better fidelity in its subspace than the former. As long as the building operator does not plan to change the velocity, LOC-TH surrogate with higher fidelity shall be the right choice.

4) *Discussion*: In summary, LR-based surrogates offer intuitive interpretations such that every unit change of a feature j results in the comfort index change by the amount of the feature's coefficient ω_j . With the coefficients, the building operators can adjust and reconfigure the building settings with accurate estimation about the thermal comfort outcome before the actual deployment. Building operators can even virtually test different possible system reconfigurations based on the coefficients and pick out the most suitable one. In the following section, we introduce the model mechanism interpretations based on another popular interpretable ML algorithm, DT.

V. INTERPRETABLE DT-BASED COMFORT SURROGATE

We investigate the LR-based surrogates from a regression perspective, where a surrogate generates a continuous comfort index. In this section, we consider another popular type of building applications and switch to a classification setting, where the data samples are classified into different clusters. Here, we round the continuous comfort index of a data sample to its nearest integer, which is deemed as the cluster label of the data sample. Note that in the PMV model, there are seven possible integers $-3, -2, \dots, 3$. We further name the integers as cold, cool, slightly cool, comfortable, slightly warm, warm, and hot, respectively, to facilitate the user-friendly description.

DT, possibly the most well-known classification algorithm, is adopted in this section. DT classifies a data sample using a tree. A data sample starts its journey at the root of the tree, and it goes to either the left or the right child node in the next level based on a threshold of a feature. Eventually, the sample reaches a leaf node, which assigns its associated cluster label to the data sample. The popularity of DT is large because of its simplicity and intuitive process, where a decision rule is just a path of the tree. Indeed, trees are suitable to offer human-friendly interpretations verified by social science [21].

Fig. 5. The fidelity of the DT-based surrogates in the global space and different local subspace with the four datasets. Fig. 5(a) and Fig. 5(b) show the results of MAE and R^2 , respectively. Local subspace learning helps improve the surrogate fidelity.

In the following part, we analyze DT's performance in different feature spaces and provide interpretations with the DT-based comfort model surrogates.

A. Fidelity vs. Interpretability

DT-based comfort modeling is not a new topic and there exist several related works [22], [23]. However, the emphasis of those works is the modeling fidelity instead of interpretability. To drive the fidelity, a tree shall be deep with many leaves, so that more information and model details can be contained within the tree. Take our global dataset for example. DT can achieve high fidelity with the R^2 metric of over 0.9, but the performance is at the cost of creating a tree of around 30 levels deep and maintaining over 80k leaves. Clearly, the complexity of the tree is beyond interpretable, at least for a human.

Our objective is to find comfort surrogates with both satisfied fidelity and interpretability. Therefore, we try to limit the tree depth to 3 and offer the users an intuitive illustration of the comfort model mechanisms. Here, we investigate the tradeoff of fidelity and depth for the DT-based surrogates and we report the results in Fig. 5.

As seen from Fig. 5, DT, which achieves remarkable fidelity on the global dataset with no limit on tree topology, suffers from the tree depth limitation here. With five levels of the tree, the best DT surrogate can achieve is 0.56 MAE and 0.65 R^2 . The performance is far from applicable, i.e., many buildings have the comfort index requirement of within $[-0.5, 0.5]$, and this tree may raise no alert even the actual comfort is over 1.0. When the tree depth is further limited to 3, the performance even drops to 0.73 MAE and 0.47 R^2 , which provides barely any approximation value.

The local surrogates show much better fidelity than the global one given the same tree depth. With the depth of 3, even the worst performed local surrogate LOC-FULL shows a 2.8x improvement on MAE and 1.8x on R^2 compared with the global surrogate. The local ones also all achieve over 0.8 R^2 with LOC-TH surrogate even beyond 0.9, which cannot be achieved by the global one even with ten more tree levels. The results justify the necessity of performing local subspace learning, which helps approximate the local dynamics of a complex model with simple and interpretable ML surrogates.

The comparison among the three local surrogates offers the same implication described previously, where the shrinking

subspace contributes to the improved fidelity. LOC-FULL covers certain space that may not be realistic for buildings, i.e., temperature as low as 0 and humidity as high as 100%. With a more reasonable setting, i.e., temperature within $25 \pm 3^\circ\text{C}$ and humidity within $50 \pm 20\%$, LOC-THV shows 6.3x better MAE than LOC-FULL for depth 3. Further reducing the subspace, LOC-TH surrogate outperforms LOC-THV by 1.8x on MAE and 14% on R^2 , with an average MAE of only 0.01 and average R^2 of 0.94. Thus, it is recommended to minimize the learning space of the features if possible to maximize both surrogate fidelity and interpretability. In the following part, we present our demonstration of tree visualization and interpretation.

B. Tree Visualization and Interpretation

We analyze the derived trees of the local surrogates and demonstrate how the visualization facilitates the building system investigation and configuration. Specifically, we consider LOC-THV with three controllable features, including temperature, RH, and velocity, and LOC-TH with only the first two features controllable.

1) LOC-THV: Fig. 6(a) shows the visualized tree of the LOC-THV surrogate. The dataset contains three comfort clusters, including comfortable, slightly warm, and warm. The portion of the warm data samples is too small (below one thousandth) to make a sizable impact on the tree decisions.

Temperature plays a key role in the tree. The tree splits in the first two levels are all based on temperature. Only after the first two levels, the tree is able to classify part of the data, i.e., comfortable as long as the temperature is below 23.8°C and slightly warm for temperature $\geq 24.8^\circ\text{C}$.

RH and velocity come to affect the decision at the third level when the temperature is neither too low nor too high. For temperature between 23.8°C and 24.3°C , the tree checks the RH and indicates comfortable if RH is below 52% and slightly warm otherwise. When the temperature is higher but below 24.8°C , velocity serves as the reference and the comfort index is slightly warm for low velocity below 0.15m/s and comfortable otherwise.

Given the current configuration of 25.0°C temperature, the tree immediately decides slightly warm as shown in the right-most leaf in Fig. 6(a). The RH and velocity do not affect such a decision when the temperature is sufficiently high. When a building operator wants to drive the classification to comfortable, two options are available in the visualized tree. First, decrease the temperature below 24.8°C but above 24.3°C and speed up the air movement above 0.15m/s as shown in the second right-most leaf. Or, further decrease the temperature below 23.8°C without changing RH and velocity as shown in the first two left-most leaves.

To summarize, the decision rules of the DT-based LOC-THV surrogate within its feature subspace are as follows.

- Temperature $[22.0, 23.8)^\circ\text{C} \rightarrow$ Comfortable.
- Temperature in $[23.8, 24.3)^\circ\text{C}$,
 - RH $[30, 52)\% \rightarrow$ Comfortable.
 - RH $[52, 70]\% \rightarrow$ Slightly warm.
- Temperature in $[24.3, 24.8)^\circ\text{C}$,

Fig. 6. The visualization of the tree of depth 3 for the DT-based local surrogates of LOC-THV in Fig. 6(a) and LOC-TH in Fig. 6(b). The controllable features are temperature, RH, and velocity for LOC-THV, and only the first two are controllable for LOC-TH. The data distribution in each internal node is shown as the stacked histogram chart and the final classification is shown in the leaves as a pie chart based on the majority of data samples in the leaf. The decision rules for comfort classification can be easily interpreted and the current configuration is highlighted in the dashed rectangular. 80% of the data samples (800k) are used for training the surrogates.

- Velocity $[0.0, 0.15)\text{m/s} \rightarrow$ Slightly warm.
- Velocity $[0.15, 0.2]\text{m/s} \rightarrow$ Comfortable.
- Temperature $[24.8, 28.0]^\circ\text{C} \rightarrow$ Slightly warm.

Note that the current setting is evaluated to be slightly warm based on the last rule. In the following part, let us interpret the DT-based LOC-TH surrogate as well.

2) LOC-TH: The visualized tree of the LOC-TH surrogate is shown in Fig. 6(b). The dataset covers two clusters, comfortable and slightly warm. Same as the LOC-THV surrogate, temperature dominates the decision rules of LOC-TH surrogate and the nodes in the first two levels all split based on temperature. Based on the derived tree, it suffices to indicate comfortable if the temperature is below 23.8°C and slightly warm if $\geq 24.5^\circ\text{C}$. Note that the latter value is different from that of the LOC-THV surrogate by 0.3°C . Although the LOC-TH subspace is contained within the LOC-THV subspace, the former reveals different mechanisms from the latter. Indeed, the small subspace helps achieve better fidelity compared to the larger subspace. Thus, the LOC-TH surrogate capturing smaller subspace shall be analyzed when adjustable velocity is not part of the consideration.

When the temperature is neither too small ($< 23.8^\circ\text{C}$) nor too large ($\geq 24.5^\circ\text{C}$), RH becomes the criteria to produce the final classifications. In the third level of the tree, when the temperature is below 24.2°C but above 23.8°C , the RH of 56% is the cut point, where lower RH indicates comfortable and larger RH leads to slightly warm. The feature subspace with the temperature between 24.2°C and 24.5°C shares the same mechanism except for the RH cut point, which is 44%. Generally, in LOC-TH's subspace, dryer air helps achieve cooler thermal sensation and improve comfort. With low temperatures, the requirement of less humid air is not that

strict, and vice versa.

The LOC-TH surrogate gives the same indication of slightly warm for the current configuration. Should the building operator plans to achieve a comfortable thermal sensation with controllable temperature and humidity, the visualized tree offers three options intuitively. One option is to decrease the temperature below 23.8°C , and RH does not affect the result within the change of $\pm 20\%$. Should the temperature is higher but below 24.2°C , the RH must be lower than 55% to achieve a comfortable sensation. The comfortable sensation can still be achieved with an even higher temperature below 24.5°C , but RH must be further reduced to 44%.

In summary, the DT-based LOC-TH surrogate has the following decision rules.

- Temperature $[22.0, 23.8)^\circ\text{C} \rightarrow$ Comfortable.
- Temperature in $[23.8, 24.2)^\circ\text{C}$,
 - RH $[30, 56)\%$ $\rightarrow$ Comfortable.
 - RH $[56, 70]\%$ $\rightarrow$ Slightly warm.
- Temperature in $[24.2, 24.5)^\circ\text{C}$,
 - RH $[30, 44)\%$ $\rightarrow$ Comfortable.
 - RH $[44, 70]\%$ $\rightarrow$ Slightly warm.
- Temperature $[24.5, 28]^\circ\text{C} \rightarrow$ Slightly warm.

Again, the current setting is classified as slightly warm based on the last rule. Overall, we show in this section that the DT-based surrogates are able to deliver both fidelity and interpretability with manageable tree depth, i.e., 3, and local feature subspace learning. The trees are also intuitive and allow the building operators to understand the current comfort state as well as make decisions on system adjustment to improve thermal comfort.

Fig. 7. The comparison of LR, DT-3 with a maximum tree depth limit of 3, and DT-N with no depth limit. Fig. 5(a) and Fig. 5(b) show the results of MAE and R^2 , respectively. Low MAE and high R^2 mean good performance, and vice versa. DT fidelity suffers from a depth limit in the trade of interpretability.

C. Comparison between LR and DT

The key difference between LR and DT is about whether thermal comfort is modeled as a regression or classification problem as aforementioned. Here, we also would like to investigate the fidelity comparison between LR and DT. Note that we do not impose any constraints on LR as described in Section IV, but constraint does exist for DT. Specifically, we impose a hard limit of the tree depth, so as to guarantee the interpretability of the tree. Therefore, we investigate the fidelity of LR, DT with a maximum depth of 3 (denoted as DT-3), and DT with no depth limit (denoted as DT-N). We present the results in Fig. 7 in terms of MAE and R^2 .

As seen from the figure, we have the same observation as discussed above that subspace learning contributes to the high fidelity. For example, the MAE is lower and the R^2 is higher for small space modeling compared to large space modeling. The interesting part comes from the two DT implementations, one with a tree depth limit of 3 and the other with no limit. It is not fully rigorous to compare the performance for regression and classification together. But still, we can observe that DT-N does exhibit excellent modeling performance, by achieving comparable or even better performance than LR in the tests. However, this comes at a price of poor interpretability, with tree depth even exceeding 30 for GLOBAL dataset and over 10 for the other datasets. With the number of leaf nodes exponentially increasing with depth, the derived trees are hardly interpretable. Then, we limit the tree depth to 3 (as DT-3) for sake of interpretability. As a result, the fidelity drops compared to DT-N. For most test cases, DT-3 is outperformed by LR and DT-N for both MAE and R^2 .

Overall, a tradeoff exists between fidelity and interpretability for DT, while LR does not have such an issue. Given that interpretability must be guaranteed, a reasonable tree depth is 3, for which the fidelity is acceptable with a satisfying fidelity. Besides the tree depth, subspace learning with high fidelity shall also be adopted to update the system settings gradually instead of dramatically.

VI. CONCLUSION

In this paper, we propose an interpretable thermal comfort system to offer interpretations of the existing thermal comfort models. We interpret the comfort models for the model I/O

and internal mechanisms. The model I/O study targets the impact of each feature on the resulting comfort index. We adopt the PDPs to analyze each feature's comfort impact in a global feature space, where the impact is modeled as a function of the feature value. We observe that the impact can be diverse. The features like temperature and humidity have a positive impact on the comfort and the impact is negative for the air velocity feature. The impact function can also be linear or nonlinear. Afterward, we apply the game-theory backed SHAP values to quantify each feature's contribution to a specific comfort output. For the internal mechanisms, we use the interpretable ML algorithms including LR and DT to produce the surrogates of the comfort model, and thereafter we interpret the surrogates to offer interpretations about the comfort model. We demonstrate that maintaining a manageable small feature space is critical to achieving outstanding modeling fidelity. Besides, such subspace learning well aligns with the concept of smart building control. The feasible control adjustments can be derived from the surrogates and presented to the model developers and building operators intuitively and informatively.

In the future, we would like to extend the interpretability concept to other building applications like smart energy management. We plan to integrate the interpretable comfort module into the existing building management systems to materialize the impact. We also track the latest progress in interpretable ML and try to explore the further enhancements of the proposed system.

REFERENCES

- [1] P. O. Fanger, "Assessment of man's thermal comfort in practice," *Occupational and Environmental Medicine*, vol. 30, no. 4, pp. 313–324, 1973.
- [2] W. Liu, Z. Lian, and B. Zhao, "A neural network evaluation model for individual thermal comfort," *Energy and Buildings*, vol. 39, no. 10, pp. 1115–1122, 2007.
- [3] R. J. de Dear, T. Akimoto, E. A. Arens, G. Brager, C. Candido, K. Cheong, B. Li, N. Nishihara, S. Sekhar, S. Tanabe *et al.*, "Progress in thermal comfort research over the last twenty years," *Indoor air*, vol. 23, no. 6, pp. 442–461, 2013.
- [4] K. B. Janda, "Buildings don't use energy: people do," *Architectural science review*, vol. 54, no. 1, pp. 15–22, 2011.
- [5] W. Hu, Y. Wen, K. Guan, G. Jin, and K. J. Tseng, "itcm: Toward learning-based thermal comfort modeling via pervasive sensing for smart buildings," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 4164–4177, 2018.
- [6] W. Zhang, W. Hu, and Y. Wen, "Thermal comfort modeling for smart buildings: A fine-grained deep learning approach," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2540–2549, 2018.
- [7] G. Gao, J. Li, and Y. Wen, "Deepcomfort: Energy-efficient thermal comfort control in buildings via reinforcement learning," *IEEE Internet of Things Journal*, 2020.
- [8] B. W. Olesen and G. S. Brager, "A better way to predict comfort: The new ashrae standard 55-2004," 2004.
- [9] M. S. Roudsari, M. Pak, A. Smith *et al.*, "Ladybug: a parametric environmental plugin for grasshopper to help designers create an environmentally-conscious design," in *Proceedings of the 13th international IBPSA conference held in Lyon, France Aug*, 2013.
- [10] T. Hoyt, S. Schiavon, A. Piccoli, T. Cheung, D. Moon, and K. Steinfeld, "Cbe thermal comfort tool center for the built environment," *University of California Berkeley*. Available online: <http://comfort.cbe.berkeley.edu> (accessed on 28 September 2019), 2017.
- [11] C. Molnar, *Interpretable machine learning*. Lulu.com, 2019.
- [12] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

- [13] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘’ why should i trust you?’’ explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.
- [14] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in neural information processing systems*, 2017, pp. 4765–4774.
- [15] L. S. Shapley, “A value for n-person games,” *Contributions to the Theory of Games*, vol. 2, no. 28, pp. 307–317, 1953.
- [16] P. O. Fanger *et al.*, “Thermal comfort. analysis and applications in environmental engineering,” *Thermal comfort. Analysis and applications in environmental engineering.*, 1970.
- [17] W. L. Haskell, I.-M. Lee, R. R. Pate, K. E. Powell, S. N. Blair, B. A. Franklin, C. A. Macera, G. W. Heath, P. D. Thompson, and A. Bauman, “Physical activity and public health: updated recommendation for adults from the american college of sports medicine and the american heart association,” *Medicine & science in sports & exercise*, vol. 39, no. 8, pp. 1423–1434, 2007.
- [18] T. C. Chapter, “Fundamentals volume of the ashrae handbook. ashrae,” *Inc., Atlanta, GA*, 2005.
- [19] “Cool united nations,” <https://ourworld.unu.edu/en/cool-united-nations>, accessed: 2020-05-05.
- [20] N. S. Raman, K. Devaprasad, B. Chen, H. A. Ingley, and P. Barooah, “Mpc for energy efficient hvac control with humidity and latent heat considerations,” *arXiv preprint arXiv:1903.04652*, 2019.
- [21] T. Miller, “Explanation in artificial intelligence: Insights from the social sciences,” *Artificial Intelligence*, vol. 267, pp. 1–38, 2019.
- [22] A. A. Farhan, K. Pattipati, B. Wang, and P. Luh, “Predicting individual thermal comfort using machine learning algorithms,” in *2015 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE, 2015, pp. 708–713.
- [23] Y. Gao, E. Tumwesigye, B. Cahill, and K. Menzel, “Using data mining in optimisation of building energy consumption and thermal comfort management,” in *The 2nd International Conference on Software Engineering and Data Mining*. IEEE, 2010, pp. 434–439.