OccuTherm: Occupant Thermal Comfort Inference using Body Shape Information

Jonathan Francis*†
School of Computer Science
Carnegie Mellon University
jmf1@cs.cmu.edu

Matias Quintana*
School of Design and Environment
National University of Singapore
matias@u.nus.edu

Nadine von Frankenberg
Department of Computer Science
Technical University of Munich
nadine.frankenberg@in.tum.de

Sirajum Munir
Secure and Intelligent IoT Department
Bosch Research & Technology Center
sirajum.munir@us.bosch.com

Mario Bergés
Civil & Environmental Engineering
Carnegie Mellon University
mberges@andrew.cmu.edu

ABSTRACT

Thermal comfort is a decisive factor for the well-being, productivity, and overall satisfaction of commercial building occupants. Many commercial building automation systems either use a fixed zone-wide temperature set-point for all occupants or they rely on extensive sensor deployments with frequent online interaction with occupants. This results in inadequate comfort levels or significant training effort from users, respectively. However, the increasing ubiquity of cheap, depth-based occupancy tracking systems has enabled an improvement in inferential capabilities. We propose the novel system *OccuTherm* to model thermal comfort of occupants. We conducted a laboratory study with 77 participants to collect data for the implementation of a thermal comfort model that derives thermal comfort using the human body shape. Based on the comparison with model baselines and ablations, we show that our approach infers thermal comfort of individuals with 60 % accuracy when body shape information is taken into account; 6 % more than state-of-the-art approaches. We make our code, mobile app, datasets, and models freely available.

CCS CONCEPTS

• **Computing methodologies** → *Artificial intelligence*; *Model development and analysis*; • **Computer systems organization** → *Embedded and cyber-physical systems*; • **Human-centered computing**;

KEYWORDS

Thermal Comfort, Human Studies, Machine Learning

*Equal contribution.
†Corresponding author.

BuildSys '19, November 13–14, 2019, New York, NY, USA

ACM ISBN 978-1-4503-7005-9/19/11…$15.00
https://doi.org/10.1145/3360322.3360858

ACM Reference Format:
Jonathan Francis, Matias Quintana, Nadine von Frankenberg, Sirajum Munir, and Mario Bergés. 2019. OccuTherm: Occupant Thermal Comfort Inference using Body Shape Information. In *The 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '19), November 13–14, 2019, New York, NY, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3360322.3360858

1 INTRODUCTION

Thermal comfort is an important factor in building control. It drives the operation of heating, ventilation, and air conditioning (HVAC) systems, which are estimated to account for 50 % of the total energy use in the built environment. Moreover, thermal comfort has a significant effect on the physiological and psychological well-being of an individual and affects occupants' health, satisfaction, and performance [4, 9]. Studies have shown that it can lead to either an increase in concentration and productivity in optimal comfort conditions or to lethargy and distraction in poor comfort conditions [5, 11]. Many commercial building control systems in use are based on models that regulate thermal conditions, often by means of pre-defined rules with pre-defined set-points, i.e., the goal temperature in an indoor environment. Temperature set-points are either derived from well-established standards, such as ASHRAE 55 [1], or require continuous feedback from occupants by means of surveys or wearables. Few building control systems prioritize the occupants' inherent physical characteristics, e.g., body shape information (height, weight, shoulder circumference), when making these thermal comfort estimates. The sophistication of non-invasive sensing and privacy preserving occupancy-tracking systems has improved greatly in the last decade, making occupant tracking and occupant parameter estimation more ubiquitous [21, 31]. Thermal comfort prediction, on the other hand, remains a fundamental challenge in this domain, due to the stochasticity of the environment, the non-stationarity of human thermal comfort preferences, and the prohibitive cost of performing large-scale thermal comfort data-collection.

In this work, we propose the *OccuTherm* system which is used for predicting thermal comfort preferences of occupants by leveraging their body shape information. Our approach improves the accuracy of thermal comfort predictions, alleviates the need for frequent occupant comfort feedback during system deployment, and leverages data from existing commercial building sensing infrastructures. The contributions of this work are the following. First, we perform a human study experiment where we collected data from 77 participants during three-hour sessions, each over the course of a year. Second, we develop a model to infer thermal comfort of individuals using body shape information. To the best of our knowledge, no such model exists and hence it is a novel approach to infer thermal comfort of individuals. Third, in order to emphasize the increased inferential power that body shape information offers to comfort modeling, we compare our model with other instances of the same hyperparameter configuration that are trained only on an ablated set of feature inputs. Models trained *with* body shape information perform *better* than their fewer-feature counterparts, by 6%. Finally, we configure our model for temperature set-point prediction and show that our strategy performs proximally to state-of-the-art techniques. Our code, mobile app, datasets, and models are available here: https://github.com/jonfranc/occutherm.

2 RELATED WORK

Thermal comfort has a considerable influence on the overall satisfaction in indoor environments [5, 11]. Many building control systems rely on generic thermal comfort models for temperature regulation that average the air temperature to achieve thermal comfort among building occupants. The most widely used models are the Predicted Mean Vote model (PMV) [8], the Pierce Two-Node Model (PTNM) [12], and the RP-884 model [6]. The PMV and the PTNM models were introduced in the 1970s; the basis of both models are laboratory studies that take physiological parameters as well as environmental data into account [8, 12]. This data includes air temperature, mean radiant temperature, relative humidity, and air velocity, and, as for human factors, clothing insulation and metabolic rate. All three mentioned models consider human factors, rather than using specific set-points, but they average the individual occupants' responses.

Recent literature employs machine learning in order to contextualize environmental data by means of supervised comfort modeling [3, 5]. In a study with 38 participants, Kim et al. [22] show that personalized thermal comfort models perform better than conventional models, such as the PMV, due to the increased model representational capacity. The evaluation in [3] assesses whether or not thermal comfort can be determined by sensor data and environmental variables, and the authors show promising results in personalized models, with an average of 83 % across their 7 participants. Their results are compared against an *always-comfortable* model as well as a linear regression model that only uses `temperature` as input. However, generalizability is hard to conclude given these cohort sizes, whose data may not capture the non-stationary properties of human comfort preference as well as ambient environmental phenomena [7, 26, 32]. Moreover, these approaches do not address the role of *body shape information* (e.g., height, weight, and shoulder circumference) in the thermal comfort predictions.

Similar attempts that promote personalized comfort models also include occupant feedback, human factors, and bio-signal data (e.g., heart rate, skin temperature, and galvanic skin response) [28, 36]. Another approach proposes to use body features that are identified through video and shows that the human thermoregulation state can be inferred from the human skin [18]. On a similar basis, the FORK system uses a depth sensor to detect, track, and estimate occupancy in buildings [31]. In this paper, we extend the FORK system to include physiological body shape information in order to infer the individual's thermal comfort preferences. Human physiology implies that body shape does play an important part in thermal comfort. An individual with a larger body surface offers a larger area for sensing the temperature outside the body. Additionally, adipose tissue has the effect of trapping heat, meaning that the human core stays warm while the body surface, i.e., the skin, cools down. To our knowledge, no other approach uses body shape information to infer personalized thermal comfort.

Table 1: Thermal comfort index, discretized thermal comfort label on 5-point scale, the number of responses in the dataset for each tier for the 77 participant subset, and the mapping to the ASHRAE thermal comfort scale.

Comfort Index	Label	Count	ASHRAE
Uncomfortably Warm	+2	48	Cooler
Slightly Uncomfortably Warm	+1	198	
Comfortable	0	1152	No change
Slightly Uncomfortably Cold	-1	452	Warmer
Uncomfortably Cold	-2	217	

Other approaches, such as SPOT [15] and SPOT+ [13], describe occupancy sensing systems for thermal control; as a result, the goal of these works is to generate a zone temperature set-point, as opposed to comfort predictions. SPOT uses the Predicted Personal Vote (PPV) model, which takes Fanger's PMV [8] and adds a linear function to include the individual's sensitivity to the variables used by the PMV. Gao and Keshav [13] show reductions in user discomfort, from 0.36 to 0.02, as compared to baselines. We compare our approach to these works, while taking thermal comfort prediction as an additional evaluation criterion for our system.

3 APPROACH

Our goal is to predict a commercial building occupant's thermal comfort, based on body shape information and relevant environmental factors. Our system is composed of three modules, for: body shape inference (Section 3.1), thermal comfort modeling (Section 3.3), and temperature set-point generation (Section 3.3.1). We specifically consider occupant body shape information that can be easily estimated or regressed from depth-camera sensor data: height, weight, and shoulder circumference.

When an occupant enters a room, their `height`, `weight`, and `shoulder circumference` are obtained using a depth-based occupancy tracking assembly, based on FORK [31]. Next, we combine this body shape information with environmental sensor data from the commercial *Building Automation and Control Network* (BACNet) infrastructure on our campus. We then classify the occupant's thermal comfort preference, conditioned on the body shape information and the environmental factors. Finally, from the set of comfort predictions, we infer the optimal zone temperature set-point range.

(a) (b) (c) (d)

Figure 1: A sample depth frame (a) and corresponding RGB image (b) of when someone going in and a sample depth frame (c) and corresponding RGB image (d) of when the same person is coming out.

3.1 Determining Body Shape Information

The *Fine-grained Occupancy estimatoR using Kinect* (FORK) system [31] uses a ceiling-mounted depth sensor to estimate the number of occupants in a room. In order to identify and track humans in the sensor's field-of-view, FORK uses a model-based approach which relies on the anthropometric properties of human heads and shoulders. We use FORK's human detection algorithm to determine body shape information in OccuTherm. The reason we choose a depth sensor for estimating body shape over an RGB camera is that it is considerably less privacy-invasive: a depth sensor cannot sense skin color, hair, cloth, and – since it is mounted overhead – it cannot see facial features. Hence, it is difficult to identify individuals using depth frames, even if the sensor is compromised. OccuTherm performs all computation at the edge and does not upload any image to any remote server. OccuTherm uses the Microsoft Kinect V2, which provides depth frames at a 512x424 resolution (see Figure 1). In this section, we describe how OccuTherm estimates height and shoulder circumference of occupants.

3.1.1 Determining Height. After FORK has detected a human head, it fits a contour and a minimum enclosing circle around the head, as shown in Figure 1(a). Among all the pixels within the circle, OccuTherm finds the pixel $P_{min} = (px, py)$ that has the minimum depth value D_{min}. Note that, since a depth sensor provides distance to its nearest object in millimeters, P_{min} is the pixel representing the highest point on the head of the person. In order to estimate the participant's height, OccuTherm estimates floor height F_{max} by building a histogram of *number* of depth pixels at different distances from the sensor, as in [31]. The bin with the highest number of pixels is considered the floor. The height of the person is then computed as $F_{max} - D_{min}$. Since a person is captured in multiple frames while they are entering and exiting, we estimate the height only when the person is directly underneath the sensor and ignore the height estimation at the edges of the frame.

3.1.2 Determining Shoulder Circumference. Inspired by FORK [31], OccuTherm determines shoulder circumference using anthropometric properties of human bodies. FORK itself does not estimate shoulder circumference, it merely detects the presence of a shoulder. OccuTherm estimates shoulder circumference using the following steps. First, we obtain the center of the head, using FORK. Next, given that the end-to-end distance between two shoulders of a person is approximately three times the diameter of head, OccuTherm fits a region-of-interest (ROI) that includes the head and shoulder and discards all pixels below a threshold $D_{min} + H + S$, in order to discard depth values below the shoulder level (see Figure 3(a)). Next, OccuTherm captures the head by discarding all the depth pixels below a threshold H, which is a bit less than the length of an average human head (see Figure 3(b)). We choose H, S to 150 and 300 millimeters, respectively. Third, it subtracts the second image from the first image to capture the shoulder (Figure 3(c)). Fourth, it detects contours using the third image and fits an ellipse to determine the circumference of the shoulder (Figure 3(d)). The fitted ellipse is shown blue in the Figure. Figure 3 shows how it captures shoulder circumference when someone is going in (Figures (a), (b), (c) (d)), and when someone is coming out (Figures (e), (f), (g), (h)). The circumference of the ellipse is used to estimate the circumference of the shoulder. Note that the circumference of the ellipse is in the pixel coordinate. In order to map it to real-world shoulder circumference, the ellipse builds a linear-regression model, using elliptical circumference as predictor to fit the training data.

This approach suffers when two shoulders get separated or when one shoulder gets occluded, as shown in Figure 4. In order to address these corner-cases, OccuTherm uses the following approach: when both shoulders get separated, it reports the elliptical circumference as *(total sum of circumference of both ellipses)*3/2.* When it detects only one shoulder, it reports the elliptical circumference as *(ellipse circumference)*3/2.* Then it uses the aforementioned linear-regression model to estimate the real-world shoulder circumference using elliptical circumference as predictor variable.

3.2 Data Collection

We mounted a Microsoft Kinect V2 depth sensor above the door, inside a large conference room (Figure 5); next, we gained real-time access to the conference room's HVAC actuator state information via BACNet; finally, we developed a mobile application that collects occupant comfort surveys (Figure 6).

In this fully-controlled thermal chamber, we performed 77 individual comfort experiments, approved by our Institutional Review Board (IRB) and in satisfaction of participant consent guidelines. Our goal was to generate a dataset than enables comprehensive study of human thermal comfort preferences, in a commercial building environment, across a wide range of indoor environmental conditions. Each comfort experiment lasted for 3.5 hours and began by manually measuring the participant's ground-truth body shape information; next, the participant was asked to pace in and out of the room, beneath the depth camera, so that we could obtain accurate body shape predictions. For the remaining 3 hours, the participant was equipped with a wearable biometrics device[1] and was provided with a smartphone that had our thermal comfort mobile application

[1]While the Microsoft Band II wearable device was used in our experiments, any wearable fitness tracker that produces such biometrics as Skin Temperature, Heart Rate, and Galvanic Skin Response may be used to replicate our results.

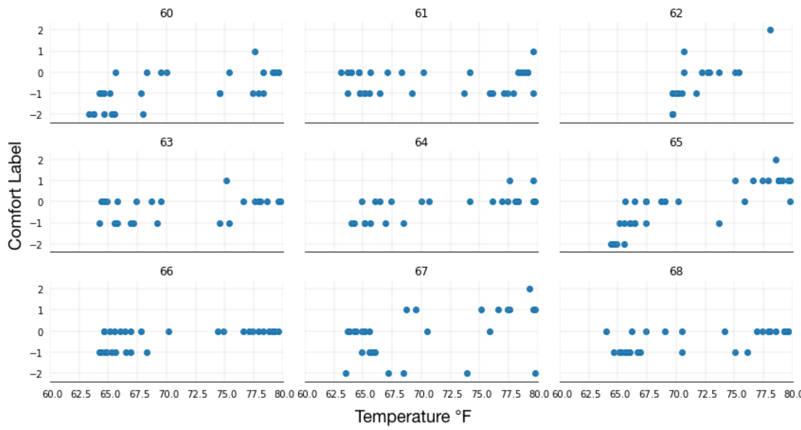
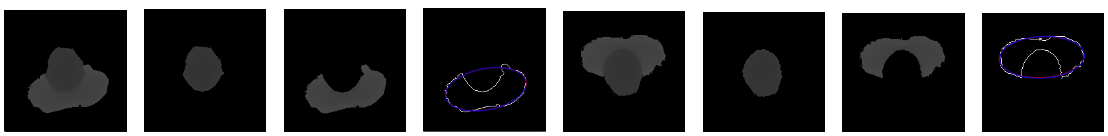
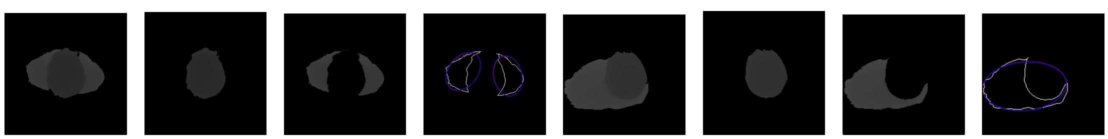


Figure 2: Plots of *Comfort versus Temperature*, for a subset of the thermal comfort human study participants (#60-#68).



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 3: Shoulder circumference estimation, when: user is entering, (a)-(d); user is exiting, (e)-(h).



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 4: Shoulder circumference estimation cases: separated shoulders, (a)-(d); one shoulder is occluded, (e)-(h).

Table 2: Participant Population Statistics for the 77 filtered participants in the OccuTherm dataset.

Feature	Min	Max	Mean	Standard Deviation
Zone Temperature	60.1°F (15.6°C)	85.0°F (29.4°C)	71.4°F (21.9°C)	6.22°F (3.5 °C)
Outdoor Temperature	6.8°F (-14.0°C)	91.4°F (33.0°C)	49.6°F (9.8°C)	20.9°F (9.8°C)
Skin Temperature	71.96°F (22.2°C)	95.0°F (35.0°C)	85.1°F (29.5°C)	4.0°F (2.2°C)
Outdoor Rel. Humidity	33.5%	100%	69.5%	13.2%
Shoulder Circumference	89.5 cm	133 cm	109.3 cm	10.9 cm
Height	151.0 cm	191.2 cm	170.1 cm	9.7 cm
Weight	90 lbs (40.82 kgs)	236.6 lbs (107.32 kgs)	153.0 lbs (69.4 kgs)	30.8 lbs (13.98 kgs)
Clothing Insulation (clo)	0.25	1.15	0.57	0.19

* Self-reported participant **Gender** was obtained as an additional feature: 34 males and 43 females.

pre-installed: the participant was instructed to engage in a low-intensity activity of their choice (e.g., reading), while completing quick thermal comfort surveys in the mobile application (Figure 6). Concurrently, we fixed the airflow rate in the room and varied the zone temperature via BACNet, between approximately 60°F to 80°F (16°C to 27°C), according to a *cold-hot-cold-hot* control schedule.

The participant completed a thermal comfort survey every five minutes *or* whenever they initiated a change in their clothing level (e.g., adding or removing a sweater) or activity type. Participants provided their comfort votes on the basis of a reduced 5-point ASHRAE 55 scale [1], see Table 1, which we used in order to reduce the complexity of voting. The use of seven-point scales generally improves reliability, however, in a setting where participants are polled in a frequent interval, less steps to perform the task increases the efficacy of the responses [2, 34]. In our study, the main objective was to determine the participant's thermal comfort, which is, according to ASHRAE, mapped to "warmer", "cooler", or "no change". However, we were also interested in whether the participant felt "uncomfortably" or "slightly" warm or cold, as this gives important meta-information for the relevance of a change in temperature for the specific individual. Our scale can be mapped to ASHRAE's thermal comfort scale as follows: "slightly uncomfortably cold" and

Figure 5: Deployment of a Microsoft Kinect for depth data-collection highlighted by a blue square.

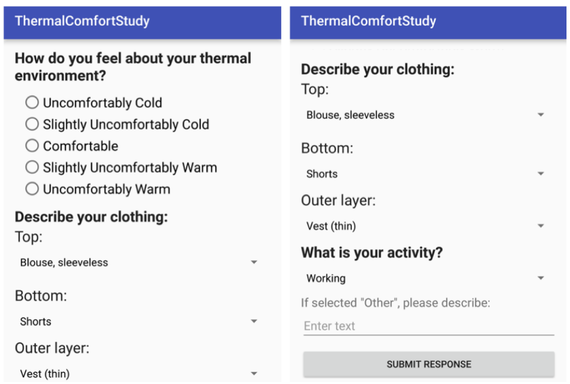


Figure 6: User interface of the mobile application that was used during the thermal comfort experiment.

"uncomfortably cold" to "warmer", "comfortable" to "no change", and "slightly uncomfortably warm" and "uncomfortably warm" to "cooler". We did not include ASHRAE's thermal sensation scale as it merely indicates the subject's current sensation, but not comfort, which is the most important factor in our case. Our thermal comfort index information is summarized in Table 1, human subjects population statistics are summarized in Table 2, and sample plots of participant comfort versus temperature are shown in Figure 2.

During each experiment, we sampled Zone Air Temperature in 30-second intervals, and we sampled Set Point Temperature and Air Flow Rate upon change. Additionally, we collected outside temperature and relative humidity (60-second granularity) from the nearest campus weather station, located a quarter mile (half kilometer) from the experiment location.

3.2.1 Dataset Curation. The data is comprised of the following *feature groups*: biometrics sensor data (`Band`), body shape information (`Body`), subjective comfort data from the mobile device application (`Survey`), environmental sensor data from the HVAC system (`HVAC`), and outdoor weather station data (`Weather`). The dataset modalities themselves are summarized in Table 3. We temporally align the samples from `Band`, `HVAC`, and `Weather` data to the nearest comfort labels specified by the `Survey` data. We observed that `Band` values exhibit little volatility in the space of 1 minute, which is the sampling rate of the wearable device and also the maximum temporal difference between `Survey` and a `Band` sample timestamps.

Next, we generated five datasets for evaluation (Table 3), based on feature subsets of the full data. This will allow us to compare ablations of our thermal comfort models that are trained with and without, e.g., body shape information, biometrics features, or external weather information (Section 4.2). *Featureset-1* (FS1) consists of 9 features. Although we collected other features, such as `Activity` and `Galvanic Skin Response (GSR)`, participants did not report many different classes for the former feature: many selected 'OTHER' and proceeded to describe their activities in their own words. For the latter, after fitting a linear regression model with all the features, we noticed that `GSR` contributed the least when compared to the remaining features. Using *Featureset-2* (FS2), we examine the effect of omitting body shape characteristics from trained models, through direct comparison with FS1. *Featureset-3* (FS3) consisted of a more limited set of features. We test the inferential value of just these modalities, since the first two are easily obtained from BACNet and local weather stations, respectively (see Tab 3, and the last three are easily regressed or inferred from depth-camera sensor data [31]. *Featureset-4* (FS4) tests the inferential value of environmental features alone. Finally, *Featureset-5* (FS5) only considers `Zone Temperature` and serves as a baseline featureset for which only a room thermostat is needed. Additionally, some participants exhibit missing *Skin Temperature* measurements due to faulty connections between the smartband and the mobile application. To address this, we proceeded to augment the missing measurements by implementing the heuristic $SkinTemp = RoomTemp + k$ where k was drawn from a normal distribution with mean and standard deviation calculated, using the heuristic, on the instances where Skin Temperature was successfully recorded. All previous tables do not consider these new value in their calculations.

We hypothesize that participants with similar physical characteristics will have similar comfort preferences, as a *basis*; confounding factors may be satisfied by, e.g., online adaptation or reinforcement of the model, over time [29]. We use *K-means* [27] to discover clusters in the OccuTherm dataset. Clusters were generated according to the set of modalities that we regard as body shape information: `Height`, `Shoulder Circumference`, and `Weight`; this information can be easily estimated or regressed using OccuTherm's depth-sensing module (see Section 3.1). We determine the number of clusters needed and regulate the quality of the clusters, by empirically minimizing the mean-squared Euclidean distances, between cluster centers and members, resulting in $K = 10$. Figure 7 shows a visualization of the 10 clusters in two-dimensional, t-distributed stochastic neighbor embedding space (2D t-SNE). We generally observe cohesion in the distribution of participant body shape information, encouraging our approach.

3.3 Thermal Comfort Modeling

We pose the thermal comfort modeling task as a supervised multi-class classification problem, wherein our model estimates the likelihood of having accurately predicted a specific comfort label for an occupant, $C = y$, conditioned on some context. With the "full" data featureset in Table 3, FS1, the model's context consists of `Band` (Ba), `Body` (Bo), `Survey` (S), `HVAC` (H), and `Weather` (W) data:

$$P(C^t = y | Ba, Bo, S, H, W)$$

$$where,$$

$$y \in \mathcal{Y} = \{-2, -1, 0, 1, 2\}$$

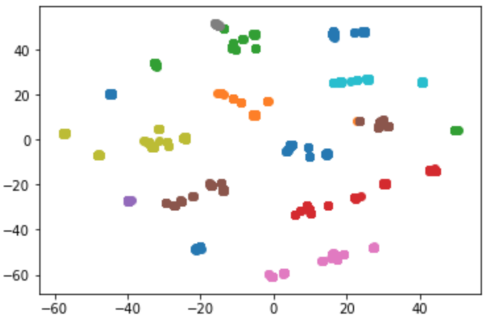


Figure 7: 2D t-SNE visualization of thermal comfort participant cluster membership, with K=10 clusters, based on participant body shape information.

Thus, our training objective is to minimize the aggregate negative log-likelihood of these predictions, over an arbitrary time horizon, with respect to the corresponding ground-truth comfort labels:[2]

$$\min \sum -log(P(C^t = y|Ba, Bo, S, H, W))$$

For the model architectural class, we select a *multi-layer perceptron* (MLP), as this has been the standard for flexibly representing and mapping diverse multimodal input distributions, even within the thermal comfort literature [7, 20, 23, 24]. At each timestep, the model makes a comfort prediction as a *distribution* over all the comfort class labels, given in Table 1b.[3] We select the label with the largest probability mass as the predicted occupant's comfort label, given the input context. Our model configuration includes 4 hidden layers (`in`, 250, 100, 25, 5, `out`), *tanh* activations, an adaptive learning rate with an initialization of $1e^{-3}$, batch size of 5, one-hot label-vector representations, *Adaptive moment estimation* (Adam) as the optimization function [25], and an 80%/20% dataset split with 10-fold cross-validation in the training split.

3.3.1 Temperature Set-Point Generation. The OccuTherm thermal comfort model takes as input body shape and environmental information and outputs comfort class labels in the set -2, -1, 0, 1, 2. From these labels, we additionally infer the zone temperature set-point that maximizes the number of participants in the dataset test split that would report "0" or *Comfortable* as their subjective response. We take our trained thermal comfort model and the same test split that was used in the previous section. The test split, stratified according to participant, is comprised of the subjective comfort responses (and associated environmental and body shape information) for 20% (16) of the total 77 participants. For each participant in the test set, we perform a forward-pass through the trained comfort model in order to infer participant comfort preferences. This yields a distribution over zone temperatures, conditioned on comfort label, from which we extract the temperature range that maximized the number of "0" votes across the test set. In Section 4.3, we compare the resultant temperature set-points with the set-point generated by baseline control strategies.

[2] In the machine learning literature, this formulation is also referred to as cross-entropy.
[3] Time-recurrent neural encoding structures (e.g., LSTMs, GRUs) lend themselves well to these sequential data inputs and may be placed in front of the MLP classifier. However, recurrent models have significantly higher training complexity and would provide best results, only after using various data-augmentation techniques, e.g., weakly-supervised generative modeling. We plan to explore this in future work.

4 EVALUATION

We have generated a paired dataset of comfort profiles and physical characteristics, from 77 participants in a commercial building environment. Using this data, we now perform an evaluation of common modeling strategies [7, 20] for empirical thermal comfort prediction, then we study how powerful physical characteristics are for estimating the thermal comfort preferences of occupants. We compare our models to baselines and ablations.

4.1 Body Shape Inference Performance

In this section, we evaluate the performance of OccuTherm in terms of its ability to estimate human height and shoulder circumference. Despite having collected ground-truth height, shoulder circumference, and depth frames for 77 subjects, we only use data from 72 subjects for evaluating body shape inference, due to logging issues.

4.1.1 Height Estimation Performance. Table 4 shows the performance of OccuTherm for estimating human height: the average and median error is 3.28 cm and 3.0 cm, respectively, when someone is entering. The average and median error is respectively 2.99 cm and 2.55 cm, when a person is exiting. Considering the mean and median height of our subjects are 171.25 cm and 171 cm, respectively, our height estimation has an accuracy of 98%.

4.1.2 Shoulder Circumference Estimation Performance. Table 4 also shows OccuTherm performance for estimating human shoulder circumference. We use 40% of the data to fit the linear regression model (see Section 3.1.2) and the remaining 60% as test data. The average and median errors for a person entering is 9.96 cm and 8.19 cm; the average and median errors for a person exiting is 10.03 cm and 9.82 cm. Considering the mean and median shoulder circumference of our subjects are 109.44 cm and 107.15 cm, respectively, our shoulder circumference estimation is over 90% accurate.

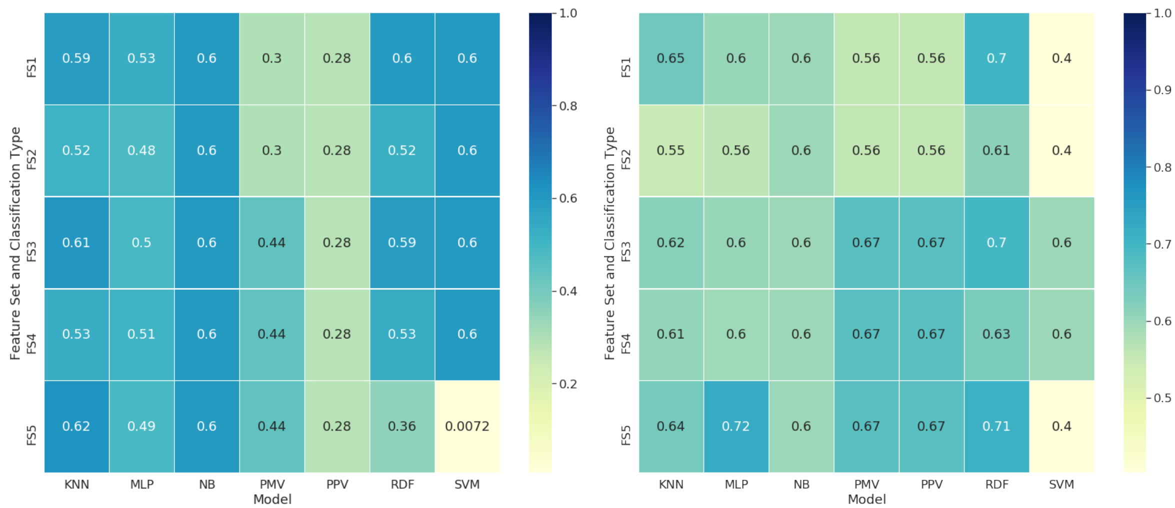
4.2 Thermal Comfort Modeling Performance

In this section, we evaluate OccuTherm in terms of its thermal comfort inference capability. To remain grounded in the related literature [7, 17, 23, 24], we evaluate our model across three dimensions: (i) holistic versus personalized comfort models; (ii) binary versus multi-class classification; and (iii) model ablations using different modality subsets. Throughout each of these experiments, we consider the evaluative datasets that we generated from our human comfort experiment (Section 3.2.1, Table 3): FS1 (all features), FS2 (all features, minus body shape information), FS3 (environmental features and body shape information), FS4 (environmental features only), and FS5 (zone temperature only). To close this section, we discuss the effect of specific feature groups, e.g., body shape information, for providing models with improved inferential capability.

4.2.1 Baselines. We use datasets FS1-FS5 to compare our chosen model configuration (Section 3.3) with discriminative classifiers, such as Random Decision Forest (RDF) and Support Vector Machines (SVM). We also include the non-parametric K-Nearest Neighbors (k-NN) classifier, the naive Bayes (NB) classifier, the *predicted personal vote* (PPV) model proposed by [16], and the *predicted mean vote* (PMV) model [8], which remains the baseline for comfort-aware commercial building control [1]. For all classifier baselines,

Table 3: We generated five evaluative data subsets: *Featureset-1* (FS1) includes environmental sensor information, occupant physical characteristics, occupant biometrics, and mobile app survey information; FS2 includes all the feature from FS1, except the body shape information; FS3 includes environmental sensor information and occupant physical characteristics; FS4 includes only environmental sensor information; and FS5 includes only zone temperature information.

Collected Features	Lin. Reg. Coeff. x 10^3	Feature Sets FS1	FS2	FS3	FS4	FS5
Zone Temperature (°F)	85.07	✓	✓	✓	✓	✓
Outdoor Temperature (°F)	0.23	✓	✓	✓	✓	×
Outdoor Relative Humidity (%)	1.86	✓	✓	✓	✓	×
Shoulder Circumference (cm)	12.77	✓	×	✓	×	×
Height (cm)	-0.47	✓	×	✓	×	×
Weight (lbs)	-2.11	✓	×	✓	×	×
Skin Temperature (°F)	-2.84	✓	✓	×	×	×
Clothing Insulation (clo)	-596	✓	✓	×	×	×
Gender	-52.59	✓	✓	×	×	×
Activity	11.96	×	×	×	×	×
GSR	0.00	×	×	×	×	×



(a) (b)

Figure 8: Holistic approach `f1-micro` results on the test set for the combination of different models with features sets for both multi-class (a) and binary target featuresets (b).

Table 4: Body shape inference performance

	Direction	Average Error	Median Error
Height	Entering	3.28 cm	3.0 cm
	Exiting	2.99 cm	2.55 cm
Shoulder Circumference	Entering	9.96 cm	8.19 cm
	Exiting	10.03 cm	9.82 cm

we conduct a hyperparameter grid search with respect to the training set, and we choose the parameters for each baseline as those of the model that performed with the highest average 10-fold cross-validation `f1-micro` score.

4.2.2 Holistic vs. Personalized. We refer to models that are trained on the entire population's thermal comfort data as *Holistic* comfort models; here, we do not distinguish the comfort responses of one participant from the responses of another participant. Instead, we stratify across all participant data within the train and validation split, in this way samples from the same participant will not exist across the train and validation split. This holistic model configuration illustrates a crowd-level thermal comfort prediction strategy, where individual biases are disregarded and optimization is instead performed across the entire population. Figure 8 shows the holistic modeling results for both multi-class and binary approaches; the value in each tile represents the `f1-micro` score of a given model (X-axis), using a specific featureset (Y-axis). For instance, in the case of thermal comfort as a multi-class problem (Figure 8(a)), we see an 6% increase in accuracy (f1-micro score) from using only environmental features (*FS3*) and environmental and physiological features (*FS1*).

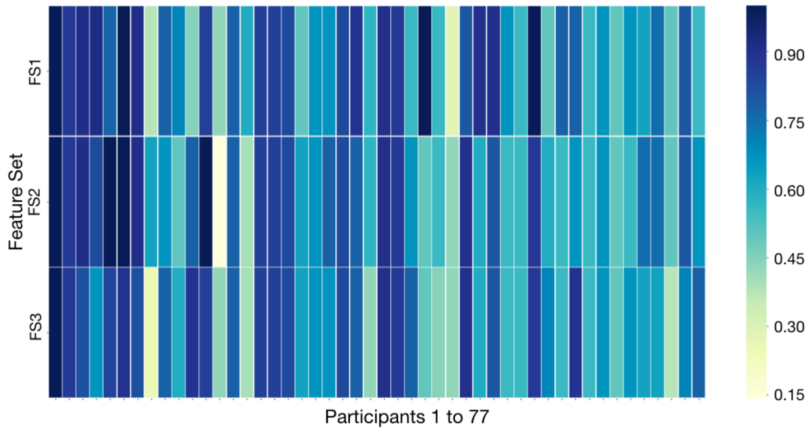


Figure 9: Personalized approach `f1-micro` results on the test set of Random Forest for a multi-class target feature on only the first three FS. The model parameters were optimized for each subject based on train/test split, resulting in better performance for a subset of subjects. This is reflected in the different colors (variance in the performance metrics) across the X-axis.

Conversely, we refer to models trained only on individual participants' thermal comfort data as *Personalized* models. Figure 9 shows the result of this personalized approach. Through this evaluation, we are able to observe how the same model performs differently for each participant. In particular, we see that the tiles can completely change their color over the horizontal axis. However, even as we increase the number of features used (Y-axis), the performance for each subset is generally consistent. This implies that the same personalized model *configuration* is able to capture each participant's unique set of preferences. Moreover, we see that the highest performance achieved in the holistic approach is surpassed by around 20% of the participants that use the same model in a personalized fashion. This seems consistent with the results obtained by, e.g., Barrios and Kleiminger [3] who were able to achieve similar performance for their personalized models on a smaller cohort.

4.2.3 Binary vs. Multi-class. In order to provide binary classifiers for baseline comparison, we re-map the target labels in each featureset from the 5-class categorical distribution to a binary one, with the label "0" representing *Comfortable* and any label in {-2,-1, 1, 1} representing "1" or *Uncomfortable*.

Figure 8(b) shows the binary prediction `f1-micro` scores for the various classifiers. Naturally, binary characterization reduces the representational burden on the models, as they only have to learn to distinguish between two effective distributions. However, such coarse-grained predictions may not be immediately suitable for temperature set-point inference, online (and reinforcement) learning, comfort-aware control, or other downstream tasks.

Figure 8(a) shows model results for multi-class classification. For the multi-class classification problem, RDF models had for *FS1*: Balanced class weights, Gini Index criterion, 2 minimum sample split, 100 estimators, and tree depth of 10; *FS2*: changed to 1000 estimators; *FS3*: changed to entropy criterion, and 100 estimators; *FS4*: changed to balanced subsamples, 100 estimators; abd *FS5*: changed to 1000 estimators, Gini criterion, and depth of 12. k-NN models had for *FS1*: brute-force search as algorithm, standard Euclidean distance as metric and K = 14; for *FS2*: K changed to 5; for *FS3*: K changed to 13; for *FS4*: K changed to 4; and for *FS5* K changed to 15. SVM models had for all first four FS: C = 1000, balanced class weight, gamma of 0.1, radial basis function kernel, and one-versus-all decision function shape, with the exception that C = 1 and gamma of 0.001 for *FS5*. Naive Bayes models were initialized without priors with a varitional smoothing of $10e^{-9}$. The MLP architecture has been discussed in Section 3.3. We see that SVMs and NB have the highest accuracies followed closely by k-NN.

4.2.4 Ablations. We performed our model ablation experiment by first generating several instances of the OccuTherm comfort model, then feeding each instance with a unique featureset (Table 3), during training and evaluation. From figures 8 and 9, we observe the effect of our ablation experiments, where supervised classification models show improvements when adding features related to body shape information, i.e., the tile value increases over the Y-axis. FS1 (all features) improves over FS2 (all features, minus body shape information) by 8%, illustrating the importance of conditioning our model's thermal comfort predictions on body shape information. We also notice RDF drops significantly with *F5*. We attribute this underperformance to the overlapping of Zone Temperature, only feature in *F5*, for all comfort labels. This low-dimensional input with significant temporal interdependencies that the rest of models are flexible enough to capture, unlike RDF.

4.3 Finding Optimal Temperature Set-Point

In order to validate OccuTherm's accuracy in temperature set-point prediction, we compared its comfort prediction capability with five common fixed temperature control strategies used in practice and in existing literature. These strategies include a fixed temperature set-point range that mimics the current control strategy commercial buildings use, a fixed temperature set-point baseline used in [33] and [14], a reactive set-point model PPV from [16], and two fixed temperature models based on the mean and median temperatures of the validation split. For these baselines, models such as OccuTherm and PPV that require parameter tuning based on existent data were trained on a 40/60 train-validation split based on the number of participants for both FS1 and FS3. This means that all five models were evaluated in 44 participants. In order to create

Table 5: Baseline comparison

Models	RMSE FS1	RMSE FS3
OccuThermMLP	**0.56**	0.73
OccuThermRDF	**0.65**	**0.65**
SPOT/SPOT+	0.66	0.66
OccuThermSVM	0.68	0.68
OccuThermNB	0.68	0.68
PPV [-0.5 | 0.5]	0.73	0.73
Median Set Point	0.79	0.79
OccuThermKNN	0.80	**0.64**
Mean Set Point	0.81	0.81
Measured Building Set Point	0.82	0.82

a range of temperature that each models perceives as a range were comfortable labels are always produced, the fixed set-point model were treated as their set-point ± 2°F, whereas in the other models this range was obtained from the training split. The PPV used the mininum and maximum temperature at which the training samples predicted [-0.5, 0.5]. On the other hand, OccuTherm's comfortable temperature range was calculated from the temperatures at which the 'Comfortable' label was 0. For each model we calculated the RMSE across all participants' responses in the validation split. Only responses at which the *indoor temperature* lied within the model's 'Comfortable' temperature range were used:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(0-y)^2}$$

The equation above shows the respective calculation where the 'predicted' label is treated as 0 for all models since we are only considering instances in their respective 'Comfortable' temperature range. y is the ground truth label from the participant. These results, in terms of RMSE, are summarized in Table 5. Here, we can see that, when using a feature set that include body shape information, such as FS1 and FS3, *OccuThermMLP*, *OccuThermRDF*, and *OccuthermKNN* are able to surpassed existing control strategies by 0.26 and 0.18, in FS1 and FS3, respectively.

5 DISCUSSION

Though the sample size of our human subject study is significantly larger than many other thermal comfort studies in the literature, it is nevertheless small for making claims about the population (e.g., commercial building occupants in the US). Despite this, our results indicate that OccuTherm can estimate body shape information with high accuracy *and*, more importantly, can leverage this information to significantly improve thermal comfort preference predictions when compared to baselines and feature ablations. Though this improvement may seem modest, as highlighted in section 4.2.2, it is worth remembering that OccuTherm works without the need for frequent user comfort feedback reports and that it leverages data from depth-imaging sensors, which are quickly becoming commonplace in indoor environments.Furthermore, to the best of our knowledge, this is the first demonstration of the predictive power of body shape information for inferring thermal comfort.

OccuTherm uses a 5-point comfort scale, rather than the 3-point comfort and 7-point sensation scale proposed by Fanger [1, 8]. Though a systematic analysis of this decision is outside of the scope of this paper, we note that there is a robust literature dating back several decades regarding the trade-offs made when using any particular scale, due to numerous issues including effects on participant's behavior and responses, ability to differentiate results, etc. Ultimately, the number of choices presented to study participants (which resembles issues in discrete-choice experiments[4]) needs to be chosen wisely and could be studied with more care in future work. Finally, the train-validation split by participant had an impact on the model's performance. When a complete stratification of the dataset was done first and then split into train and validation, the OccuTherm k-NN was able to achieve 79% and 72% accuracy in the binary and multi-class approach. However, this approach allows the co-existance of samples from the same participant in both splits, exposing the model to a portion of the participant's responses distribution; such a setting could be preferrable for re-occuring participants. Thus, we opted for the split based on participants.

We used a scale in order to measure weight of the subjects, which was used as a feature to infer their thermal comfort; in the future, a model can be built using body shape to regress weights of individuals. We also noted clothing insulation in the dataset: in the future, depth frames can be directly used to infer level of clothing insulation. Note that inaccuracies in the estimation of shoulder circumference could affect the performance of thermal comfort inference. People may carry objects, e.g., backpacks, laptops, helments hanging in the shoulder that could affect the estimation of shoulder circmference. It may require detection of such objects as in [30] and refine the shoulder estimate. We leave a thorough sensitivity analysis on this issue as future work.

We realize that 77 subjects is not large enough to capture all possible factors (e.g., social, environmental) that could impact the thermal comfort preference of individuals and the resultant commercial building control strategies [10, 19]. However, it has been shown that heat dissipation rate of individuals depends on the body surface area. As a result, a tall and skinny person can tolerate higher room temperature compared to a person having a rounded body shape since the tall person has a larger surface to volume ratio [35]. So, it is intuitive to assume that body shape can be useful to infer thermal comfort preference of individuals to some extent. We leave a long term, large scale thermal comfort study to analyze the impact of other factors as future work.

6 CONCLUSION

In this work, we present OccuTherm, a novel thermal comfort prediction system based on occupant body shape information. We conducted a human thermal comfort study in a fully-controlled and fully-sensed smart environment, where biometrics, physical measurements (height, shoulder circumference), and subjective comfort responses were recorded and integrated. With this dataset, we compared *holistic comfort models* with *personalized comfort models* and showed the significance of physical characteristics across a sample population for thermal comfort modeling. While holistic approaches can achieve `f1-micro` scores as high as 0.8, personalized models can surpass this value. Nevertheless, we saw in Figure 9 that, even if the models are trained for a particular user, it may not

[4]https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.1587

perform as well for others. Further exploration of the sensitivity of our model to the input features is warranted in order to fully understand the impact of these on the thermal comfort prediction in both the holistic and personalized settings, especially in light of possible privacy implications of the data that OccuTherm collects. Finally, we make our code, datasets, models, and mobile application available to the community.

Our results open many promising avenues of future research. Given the need for additional human subject samples, data augmentation could be investigated as well as novel semi-supervised ways to take advantage of the much larger collection of depth-imaging data that is available without thermal comfort labels. If more data becomes available, learning embeddings for more generalizable comfort models could be fruitful. Finally, though OccuTherm was described here as an inference system, there is significant potential for including it in a closed-loop control scenario, where we may perform online learning and elicit thermal comfort responses opportunistically in order to improve the models.

ACKNOWLEDGMENTS

We thank Michael Frenak, Nicole Ho, Oluwatobi Adekanye, Alexander Davis, and Roykrong Sukkerd for their support with the human studies experiment. This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Building Technologies Office Award Number DE-EE0007682. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

[1] American Society of Heating Refrigerating and Air-Conditioning Engineers. Standards Committee. 2013. Thermal environmental conditions for human occupancy. *ASHRAE standard ; 55-2013* 2013, STANDARD 55 (2013), 1–44.

[2] Emin Babakus and W Glynn Mangold. 1992. Adapting the SERVQUAL scale to hospital services: an empirical investigation. *Health services research* 26, 6 (1992).

[3] L. Barrios and W. Kleiminger. 2017. The Comfstat - automatically sensing thermal comfort for smart thermostats. In *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 257–266.

[4] Tom Y. Chang and Agne Kajackaite. 2019. Battle for the thermostat: Gender and the effect of temperature on cognitive performance. *Plos One* 14, 5 (2019).

[5] Weilin Cui, Guoguang Cao, Jung Ho Park, Qin Ouyang, and Yingxin Zhu. 2013. Influence of indoor air temperature on human thermal comfort, motivation and performance. *Building and Environment* 68 (2013), 114–122.

[6] Richard J. de Dear, Gail Schiller Brager, James Reardon, Fergus Nicol, et al. 1998. Developing an adaptive model of thermal comfort and preference/discussion. *ASHRAE transactions* 104 (1998), 145.

[7] Diana Enescu. 2017. A review of thermal comfort models and indicators for indoor environments. *Renewable and Sustainable Energy Reviews* 79, Supplement C (2017), 1353 – 1379.

[8] Poul O. Fanger. 1967. Calculation of thermal comfort, Introduction of a basic comfort equation. *ASHRAE transactions* 73, 2 (1967), III–4.

[9] William J Fisk. 2002. How IEQ affects health, productivity. *ASHRAE journal* 44 (2002).

[10] J. Francis, A. Oltramari, S. Munir, C. Shelton, and A. Rowe. 2017. Poster Abstract: Context Intelligence in Pervasive Environments. In *2017 IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI)*. 315–316.

[11] Monika Frontczak and Pawel Wargocki. 2011. Literature survey on how different factors influence human comfort in indoor environments. *Building and Environment* 46, 4 (2011), 922–937.

[12] Adolf P. Gagge. 1971. An effective temperature scale based on a simple model of human physiological regulatory response. *Ashrae Trans.* 77 (1971), 247–262.

[13] Peter Xiang Gao and Srinivasan Keshav. 2013. Optimal personal comfort management using SPOT+. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 1–8.

[14] Peter Xiang Gao and S. Keshav. 2013. Optimal Personal Comfort Management Using SPOT+. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings (BuildSys'13)*. ACM, New York, NY, USA, Article 22, 8 pages.

[15] Peter Xiang Gao and Srinivasan Keshav. 2013. SPOT: a smart personalized office thermal control system. In *Proceedings of the fourth international conference on Future energy systems*. ACM, 237–246.

[16] Peter Xiang Gao and S. Keshav. 2013. SPOT: A Smart Personalized Office Thermal Control System. In *Proceedings of the Fourth International Conference on Future Energy Systems (e-Energy '13)*. ACM, New York, NY, USA, 237–246.

[17] Ali Ghahramani, Chao Tang, and Burcin Becerik-Gerber. 2015. An online learning approach for quantifying personalized thermal comfort via adaptive stochastic modeling. *Building and Environment* 92 (2015), 86–96.

[18] Farrokh Jazizadeh and S Pradeep. 2016. Can computers visually quantify human thermal comfort?: Short paper. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, 95–98.

[19] Z. Jiang, J. Francis, A. K. Sahu, S. Munir, C. Shelton, A. Rowe, and M. Bergés. 2018. Data-driven Thermal Model Inference with ARMAX, in Smart Environments, based on Normalized Mutual Information. In *2018 Annual American Control Conference (ACC)*. 4634–4639. https://doi.org/10.23919/ACC.2018.8431085

[20] Soteris A. Kalogirou. 2000. Applications of artificial neural-networks for energy systems. *Applied Energy* 67, 1 (2000), 17 – 35.

[21] Nacer Khalil, Driss Benhaddou, Omprakash Gnawali, and Jaspal Subhlok. 2017. Sonicdoor: scaling person identification with ultrasonic sensors by novel modeling of shape, behavior and walking patterns. (2017), 3.

[22] Joyce Kim, Stefano Schiavon, and Gail Brager. 2018. Personal comfort models–A new paradigm in thermal comfort for occupant-centric environmental control. *Building and Environment* 132 (2018), 114–124.

[23] Joyce Kim, Stefano Schiavon, and Gail Brager. 2018. Personal comfort models–A new paradigm in thermal comfort for occupant-centric environmental control. *Building and Environment* 132 (2018), 114–124.

[24] Joyce Kim, Yuxun Zhou, Stefano Schiavon, Paul Raftery, and Gail Brager. 2018. Personal comfort models: predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning. *Building and Environment* 129 (2018), 96–106.

[25] Diederik P Kingma and Jimmy Lei Ba. 2015. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations (ICLR)* 2015 (2015).

[26] Laura Klein, Jun Young Kwak, Geoffrey Kavulya, Farrokh Jazizadeh, Burcin Becerik-Gerber, Pradeep Varakantham, and Milind Tambe. 2012. Coordinating occupant behavior for building energy and comfort management using multi-agent systems. *Automation in Construction* 22 (2012), 525–536.

[27] S. Lloyd. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory* 28, 2 (March 1982), 129–137.

[28] Parisa Mansourifard, Farrokh Jazizadeh, Bhaskar Krishnamachari, and Burcin Becerik-Gerber. 2013. Online learning for personalized room-level thermal control: A multi-armed bandit framework. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 1–8.

[29] Parisa Mansourifard, Farrokh Jazizadeh, Bhaskar Krishnamachari, and Burcin Becerik-Gerber. 2013. Online Learning for Personalized Room-Level Thermal Control: A Multi-Armed Bandit Framework. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings (BuildSys'13)*. ACM, New York, NY, USA, Article 20, 8 pages.

[30] Niluthpol Chowdhury Mithun, Sirajum Munir, Karen Guo, and Charles Shelton. 2018. ODDS: real-time object detection using depth sensors on embedded GPUs. In *Proceedings of the 17th ACM/IEEE International Conference on Information Processing in Sensor Networks*. IEEE Press, 230–241.

[31] S. Munir, R. S. Arora, C. Hesling, J. Li, J. Francis, C. Shelton, C. Martin, A. Rowe, and M. Berges. 2017. Real-Time Fine Grained Occupancy Estimation Using Depth Sensors on ARM Embedded Platforms. In *2017 IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*. 295–306.

[32] V. Putta, G. Zhu, D. Kim, J. Hu, and J. Braun. 2012. A Distributed Approach to Efficient Model Predictive Control of Building HVAC Systems. *International High Performance Buildings Conference* (2012).

[33] Alimohammad Rabbani and S. Keshav. 2016. The SPOT* Personal Thermal Comfort System. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments (BuildSys '16)*. ACM, New York, NY, USA, 75–84.

[34] Sheetal B Sachdev and Harsh V Verma. 2004. Relative importance of service quality dimensions: A multisectoral study. *Journal of services research* 4, 1 (2004).

[35] S.V. Szokolay. 2008. *Introduction to Architectural Science*. Taylor & Francis. https://books.google.com/books?id=s79BFBYgJYUC

[36] Liang Zhang, Abraham Hang-yat Lam, and Dan Wang. 2014. Strategy-proof thermal comfort voting in buildings. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 160–163.