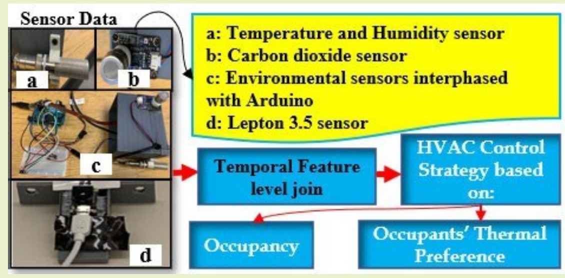
Occupancy and Thermal Preference-Based HVAC ControlStrategyUsingMultisensor Network

Yaa T. Acquaah , Balakrishna Gokaraju, *Senior Member, IEEE*, Raymond C. Tesioro, III, Gregory Monty, and Kaushik Roy

***Abstract*—Human-in-the-loop heating, ventilation, and air conditioning (HVAC) control-based methodologies have gained much attention due to continual discomfort compliance of occupants in residential and commercial buildings; spawning thermal comfort research interest in leveraging emerging advanced technologies to address the prolonged problem of discomfort and energy efficiency. In the past, thermal comfort studies have been conducted to determine the thermal sensation, preference, and comfort based on the American Society of heating, refrigerating, and airconditioning engineers (ASHRAE) Global Thermal Comfort Database II and customized dataset through machine learning. The ASHRAE Database II is an open-source database that includes sets of objective indoor climatic observations with corresponding subjective evaluations by the building occupants who were used as subjects in experiments. Environmental parameters and occupants' skin temperature have been used to develop machine learning algorithms to predict thermal comfort indices in both indoor and outdoor settings. However, none of these studies have investigated merging environmental parameters and thermal images to predict thermal comfort indices of occupants. In this study, the holistic understanding of individuals thermal comfort environment was considered by fusing analog environmental sensors and thermal images captured at the time of the subjective measurement. Wavelet-scattering features were obtained from the occupants' thermal image surroundings and joined to the environmental parameters. This research developed different machine learning models, processing methods and evaluated the results based on the fused dataset. The results show the possibility of real-time prediction of occupancy and thermal preference through classical machine learning, and stacked models with high accuracy. The proposed framework achieved an estimated 45% mean energy savings during a ten-day energy analysis.**



***Index Terms*—Energy efficiency, heating, ventilation and air conditioning (HVAC), machine learning.**

Manuscript received 30 December 2022; revised 11 March 2023; accepted 20 March 2023. Date of publication 10 April 2023; date of current version 31 May 2023. This work was supported in part by the NC Department of Environmental Quality (NCDEQ) through Cares Act, in part by the Center for Energy Research and Technology (CERT), in part by the UNC Policy Collaboratory and Startup Funds by Office of Dean, and in part by the College of Engineering at North Carolina Agricultural and Technical State University. The associate editor coordinating the review of this article and approving it for publication was Dr. Pedro Oliveira Conceicao Junior. *(Corresponding author: Yaa T. Acquaah.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Office of Research Compliance and Ethics (ORCE) under the Division of Research and Economic Development of North Carolina Agricultural and Technical State University.

Yaa T. Acquaah is with the Computational Data Science and Engineering Department, North Carolina A and T State University, Greensboro, NC 27587 USA (e-mail: ytacquaah@aggies.ncat.edu).

Balakrishna Gokaraju is with the Computational Data Science and Engineering Department, North Carolina Agricultural and Technical State University (NCAT), Greensboro, NC 27411 USA.

Raymond C. Tesiero, III is with the Civil, Architectural and Environmental Engineering Department, North Carolina Agricultural and Technical State University (NCAT), Greensboro, NC 27411 USA.

Gregory Monty is with the Center for Energy Research and Technology, North Carolina Agricultural and Technical State University (NCAT), Greensboro, NC 27411 USA.

Kaushik Roy is with the Department of Computer Science, North Carolina Agricultural and Technical State University (NCAT), Greensboro, NC 27411 USA.

Digital Object Identifier 10.1109/JSEN.2023.3264474

I. Introduction

THE importance of the heating, ventilation, and air conditioning (HVAC) system in building thermal regulation cannot be over emphasized. Thermal comfort and energy efficiency remains the top two functions of the HVAC system. Thermal comfort is the state of mind that establishes thermal satisfaction with the thermal environment. Thermal comfort has metrics like the thermal preference, sensation, acceptability, and satisfaction and accessed by subjective measurements [1]. Fanger [2] a pioneer expert in the field of thermal comfort and environment perception, developed an index for measuring indoor thermal comfort. According to Fanger, the requirements for steady-state thermal comfort are when: 1) the body is in heat balance; 2) mean skin temperature and sweat rate, influencing this heat balance, are within certain limits; and 3) no local discomfort exists [3]. A lot of studies composing of real-world and chamber experiments have supported the findings of Fanger. Other Researchers have critiqued the predicted mean vote (PMV) model based on input parameters, building type, and geographic application range [3], [4], [5]. This has led to the exploration of machine learning techniques to predict thermal comfort. The motivation for this work is to find a new method by which the HVAC systems in buildings can be controlled based on occupancy

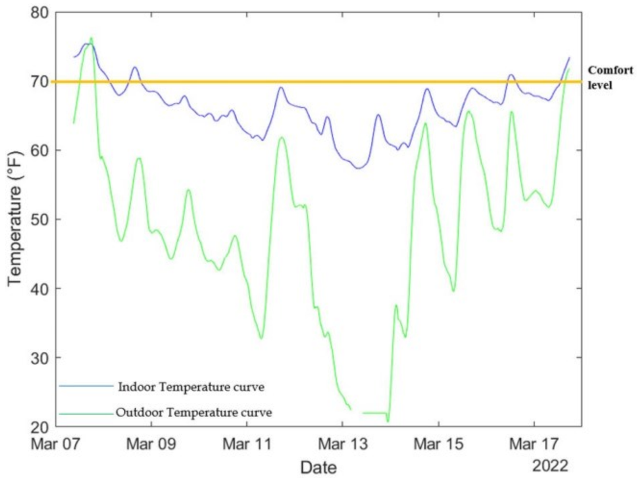


Fig. 1. Indoor and outdoor temperature graph.

and thermal preference. To do this, wavelet-scattering features of time-stamped thermal images captured during survey collection is fused with analog sensors streaming environmental parameters in real time. To achieve this, thermal images of subjects are captured every minute as subjective measurements of thermal comfort indices are collected and at the same time analog sensors stream data. The main contributions of this article are as follows.

1) Collection of accurate multisensor analog readings and infrared images using corresponding timestamp with hardware integration and development.
2) To develop machine learning models for estimating real-time occupancy based on fused dataset of thermal imagery and analog sensors.
3) To develop machine learning models for predicting real-time thermal preference based on fused dataset of thermal imagery and analog sensors.

We propose occupancy and thermal preference machine learning models based on the combination of thermal image and sensor parameters for improved performance. The fusion of the timestamped thermal images helps to better understand the thermal environment of occupants in a given space and how it affects an individual's thermal comfort at a given time. Ten days temperature data of the HVAC laboratory in North Carolina Agricultural and Technical State University (NCAT) were collected from March 7 to March 17, 2022 and were plotted together with the outdoor temperature during the period as shown in Fig. 1. It was discovered that this room never reaches occupants comfort level because the line for comfort level reaches 70 °F which is the occupants' comfort level only six times throughout the ten days period. It is approximately 6 min that it reaches occupants comfort level in the laboratory, and this proves why this research is significant. The system was on and off all day long whether occupied or not but occupants were comfortable a few times. The thermostat is outside the classroom therefore if the HVAC system understands when there are people around and their comfort preferences, energy can be conserved. It can be observed from the graph on March 13th that the room was below 60 °F. This emphasizes the need to explore a new technique to control HVAC system of the room based on infrared cameras and have it run according to occupancy and occupants' thermal preferences. The upward end of the spike in the graph indicates when the HVAC system is running and thus the spikes in one day represents how much the HVAC system ran. It also emphasizes the need for a variable air volume (VAV) system control, the classrooms that do not have a thermostat, clearly never reach occupant comfort. Also, the laboratories typically have zero occupancies on weekends. During these "no occupancy" periods, the preferred mode of the system should be "drift mode." Drift mode is when a room naturally heats up or cools down to a temperature setpoint that is not in the "comfort range," thus saving energy. Precautions should be taken to not allow the room to get too cold (it could freeze water pipes) or too hot or humid (which could create mold). Ideally, allowing drift between 60 °F and 80 °F saves energy when not occupied. Traditional thermostats do not allow this drift to happen, hence the need for an expensive HVAC control system with programming capabilities for classroom and laboratory temperature scheduling every semester which mostly never happens. This illustrates that the system of using infrared cameras to predict comfort level and occupancy is a better system. This research demonstrates that a smart diffuser system with infrared cameras, allows "pockets" of control (energy savings) within a space or room.

This article is organized as follows: Section II introduces the available techniques for occupancy detection and thermal preference prediction. The technical details of the methodology including data collection, data integration, feature extraction, and modeling techniques used for the study are discussed in Section III. The results and discussion of the proposed methodology are further described in Section IV. Section V presents a summary of conclusions in this work, energy analysis, and future research ideas.

II. Related Work

Pilot research on thermal comfort was conducted by Fanger in the 1970s where the formulation of the PMV model and input parameters were introduced. The PMV was developed using principles of heat balance and experimental data collected in a controlled climate chamber under steady state conditions [1]. Fanger's equations are used to calculate the PMV of a group of subjects for a particular combination of air temperature, mean radiant temperature, relative humidity, air speed, metabolic rate, and clothing insulation (CLO). The adaptive model, on the other hand, was developed based on hundreds of field studies with the idea that occupants dynamically interact with their environment [4]. Outdoor climate influences indoor comfort because humans can adapt to different temperatures during different times of the year is the foundational knowledge that the adaptive model is based on. Occupants control their thermal environment by means of clothing, operable windows, fans, personal heaters, and sunshades [2], [3]. The PMV model can be applied to air-conditioned buildings, while the adaptive model can be applied only to buildings where no mechanical systems have been installed [4]. Over the past decade, researchers in the field of thermal comfort have conducted experiments to investigate machine learning models to predict the metrics of thermal comfort namely thermal preference, sensation, acceptability, satisfaction, and so on. Mostly features utilized in the training of these machine learning models included skin temperature of various parts of the body, parameters from environmental sensors, wearable sensors, and activity meters [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24]. Machine learning models trained on skin temperature features are prone to errors since it becomes quite impossible for occupants in a building to remain in a specific posture for skin temperatures to be extracted. Moreover, distance calculations from thermal cameras needs to be incorporated into these models because the temperature of an object varies depending on the distance from the camera. In a recent study conducted by Zhou et al, machine learning models to predict thermal sensation were developed based on quality-controlled database. Amongst the major findings from this research suggested that the multilayer perceptron (MLP) algorithm achieved higher accuracy in the prediction of thermal sensation [6]. Previous researchers designed experiments or leveraged resources from the ASHRAE database I and II to investigate the thermal sensation of occupants in a building [10], [11], [12]. Narrowing down to the thermal comfort metrics under investigation in this research, there has been numerous works done on thermal preference [26], [27], [28], [29], [30]. Kim et al. [26] predicted individuals' thermal preference using occupant heating and cooling behavior [25]. Cosmo and Simha [27] also designed an experiment to collect thermal preference votes of individuals to train machine learning models. In another study by Aguilera et al. [28], ANN, Naive Bayes (NB), and fuzzy logic machine learning models were trained to predict thermal preference on three categories based on air temperature and relative humidity as features. Martins and coauthors trained a thermal preference artificial neural network (ANN) model for older people based on dry bulb temperature, radiant temperature, relative humidity, air speed, corrected metabolic rate, clothing level, and health status [30]. Currently, to the best of the authors' knowledge, no literature has explored the possibility of joining thermal images and environmental sensor data for occupancy detection and thermal preference prediction for smart HVAC control.

In this study, we are interested in investigating machine learning models trained on combined time stamped thermal imagery and environmental parameters dataset to predict occupancy and individuals' thermal preference in a building. Using various data sources, we aim to leverages the advantages of different environmental sensors and thermal cameras to increase accuracy and mitigate uncertainty. Since each sensor has its own weaknesses, combining the dataset from each sensor creates a complete system improving the reliability of prediction of thermal preferences. This provides top notch data analysis integrating sensory information to make inferences regarding the surrounding thermal environment. This combined information gathered from various sensors have advantage over each sensor being considered individually.

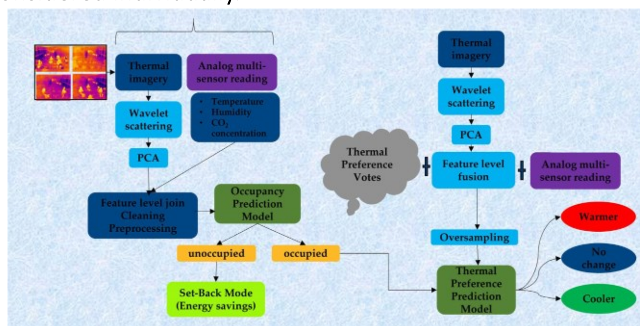


Fig. 2. Methodology diagram of the framework.

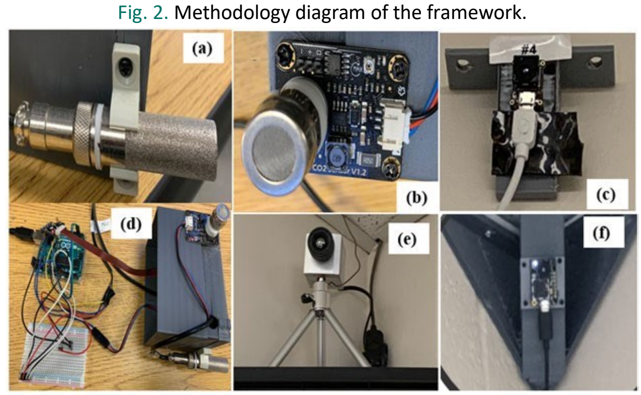


Fig. 3. Experiment setup. (a) Temperature and humidity sensor. (b) Carbon dioxide sensor. (c) Wall mounted lepton 3.5. (d) Arduino Uno interphased with environmental sensor. (e) Optrix PI 160. (f) Corner mounted lepton 3.5.

III. Materials and Methods

The occupancy and thermal preference-based methodology for smart HVAC control is illustrated in Fig. 2. We implemented ten state-of-the-art machine learning algorithms for occupancy detection and thermal preference prediction. This analysis allows us to understand which relevant features and machine learning models perform well with occupancy detection and thermal preference prediction.

A. Data Collection

To explore the possibility of fusing environmental parameters with thermal imagery and survey of subjective thermal preference votes, an experiment was designed. In the experimentation, thermal images were captured with the Lepton 3.5 infrared camera and at the same time environmental parameters stream and surveys were completed online by subjects. Surveys were conducted each time a thermal image was captured. The data gathered included time-stamped thermal images with corresponding environmental parameters and thermal preference votes. The laboratory for this experimentation was an air-conditioned room. The data collection hardware consisted of carbon dioxide concentration sensor, temperature, and humidity sensor interfaced with an Arduino Uno microcontroller connected to a desktop computer as shown in Fig. 3. The data collection software was a CoolTerm software and a Google form survey, which resided on the desktop computer. The data collection is followed with data preparation before the data is fed into classification models. The thermal preference experiment was carried out in the Artificial intelligence and Visualization laboratory at NCAT from February to September of 2021. Fifteen subjects in the age range of 19–34 years, comprising nine males and six females were recruited to aid in the data collection. Environmental data (within the laboratory) were collected using carbon dioxide, temperature, and humidity sensors. The temperature and humidity sensors are calibrated with the Optrix PI 160 camera which interrogated the same area these sensors had been placed. Data were collected from these sensors interfaced with an Arduino Uno microcontroller and connected to a desktop computer. Environmental data were streamed every minute with the help of the CoolTerm software at the same time as subjects completed a Google form survey about their thermal sensation, comfort, acceptability, and preferences. The environmental data and the survey data were combined based on the timestamp for further analysis. The 222 thermal images, environmental studies and target data of thermal preference votes were utilized in this study. Additionally, the experiment for the occupancy dataset was carried out in the Graham Hall HVAC laboratory at NCAT from February to March of 2022.

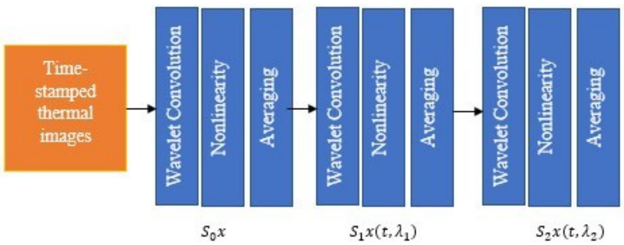


Fig. 4. Wavelet-scattering feature extractor.

B. Feature Extraction

Feature extraction is an important step in the workflow of classification machine learning model development. This is because the performance of the machine learning model is widely dependent on the extraction of relevant discriminating features for training. For each $x \in I_N$, (where x represents an image in a set of images I_N. N is the total number of samples contained in the set.) wavelet-scattering features up to the third order are computed. In the wavelet-scattering framework, which was first introduced by Bruna and Mallat [31], a succession of operations involving convolutions, nonlinearity, and averaging are performed to extract features which are invariant to translation, rotation, scaling, and deformation. The modulus computes a lower frequency envelope, thus squeezing the energy from the images. Integrating the modulus as presented in (1) of the convolution results in $L1$ norm which is invariant [31]. This is achieved by leveraging wavelets as filters to convolve thermal image, applying the modulus and then averaging to obtain invariant features as shown in Fig. 4

$$\int |x * \psi_\lambda(t)|\, dt = \|x * \lambda\|_1 \tag{1}$$

where x is a time stamped image and ψ_λ is a wavelet filter. A three-layer wavelet scattering with a 100-by-100 invariance scale is constructed as the feature extractor.

$S_0x(t,\lambda_0)$, $S_1x(t,\lambda_1)$, $S_2x(t,\lambda_2)$ represent the zeroth-order, first-order, and second-order scattering coefficients. λ_0,λ_1, and λ_2 are the center frequency of zeroth-order, first-order, and second-order wavelets, respectively. At the beginning of the wavelet-scattering feature extraction, the original time-stamped image is convolved with the low-pass wavelet Φ to obtain the average of the input image as shown in the following equation:

$$S_0x = x * \Phi. \tag{2}$$

The application of the modulus shifts the energy in the higher frequency bands toward the lower frequency bands in the next decomposition. First-order scattering coefficients are obtained by convolving modulus with lowpass as shown below in the following equation:

$$S_{1x}(t,\lambda_1) = \left|x * \Psi_{(\lambda_1)}\right| * \Phi. \tag{3}$$

Similarly, following equation presents second-order scattering coefficients:

$$S_{2x}(t,\lambda_1,\lambda_2) = ||x * \Psi_{(\lambda_1)}| * \Psi_{(\lambda_2)}| * \Phi. \tag{4}$$

The MATLAB wavelet toolbox was used to compute the wavelet-scattering features [32]. The coefficients which were the scattering features are down sampled to reduce computational complexity.

C. Feature Reduction

Principal component analysis is used as a dimension reduction technique to find a few orthogonal linear combinations of the scattering features called principal components. Most of the variance in the dataset are captured by the first and second principal components.

D. Training Data and Data Preprocessing

Data integration from multisensor sources was possible by combining wavelet-scattering features of thermal images from thermal camera, environmental sensors, and survey forms based on timestamp leveraging a python script. A feature matrix $A_{N,M}$ (N is the number of observations and M is the number of features for each observation) is constructed for training of machine learning thermal preference prediction model. A feature row, $A_{\{1,2,\ldots,N\},\{1,2,\ldots,M\}} \in A_{N,M}$ for each observation $a_{n,m}$ therefore composes of detailed relevant scenery information from wavelet scattering obtained from thermal images, physiological information on activity and clothing level, and environmental parameters from sensors. The multiclass labels for occupant thermal preference are represented as $y_i \in Y$. Furthermore, a feature matrix $B_{N,M}$ (M is the number of features for each observation) is also constructed for occupied and not-occupied class training for observations for the occupancy dataset. A feature vector, $B_{\{1,2,\ldots,N\},\{1,2,\ldots,M\}} \in B_{N,M}$ for each observation $b_{n,m}$ therefore composes of detailed relevant scenery information from wavelet scattering obtained from thermal images and environmental parameters from sensors. The binary class labels for occupancy are represented as $y_i \in Y$. The two datasets were cleaned to remove missing values. Duplicated observations were identified and removed from the datasets. Observations that were significantly different from other data points were removed to avoid distortion of the analysis. The CLO and metabolic equivalent of task (MET) values from ASHRAE of clothing insulation, and metabolic activity levels in the thermal preference dataset were substituted for the variables. Data transformation was necessary for both occupancy and thermal preference dataset due to the different units and scales of the features, the dataset is transformed using the standard scaler to ensure mean of the data is zero and the standard deviation is one.

E. Feature Classification Using Machine Learning Models

In this section, various state-of-the-art supervised classification machine learning models were explored to classify both the occupancy dataset and the thermal preference dataset. Description of each of this machine learning models is presented in the subtitles below:

1) K-Nearest Neighbors: The K-nearest neighbors is a nonparametric classification algorithm that utilizes neighbor point information to predict target class. A sample is classified by the popularity vote of its nearest neighbors. The distance between and training samples and new samples to be classified is computed and the class is predicted based on the most common nearest neighbors' class labels. The distance computation can be a Euclidean and Manhattan distance. The KNN classification algorithm has been successfully applied in the classification of features in thermal comfort analysis.

2) Classification and Regression Tree: Classification and regression tree (CART) is also a nonparametric algorithm that seeks to find the best split to subset the data, and they are also known as decision trees. Evaluation of the quality of splits is based on metrics such as Gini impurity, information gain, or mean square error (MSE). Decision tree algorithm has the tendency to overfit since all samples are tightly fit within the training data.

3) Support Vector Machine: A support vector machine (SVM) is a supervised learning algorithm useful for classification and regression problems, including signal-processing image classification, natural language processing, and speech. The SVM algorithm is generally suited for binary problems so multiclassification problems are reduced to series of binary problems. The objective of the SVM algorithm is to find a hyperplane that, to the best degree possible, separates data points of one class from those of another class. The support vectors are the extreme points in the dataset. The distance between a support vector and a hyperplane should be as far as possible. Thus, the hyperplane has the maximum distance to the support vectors of any class. The sum of the shortest distance to the closest positive point to the hyperplane and shortest distance from the hyperplane to the closest negative point is called the distance margin. This is also the distance between the two support vectors. Datasets are transformed to a higher dimension in problems where a hyperplane cannot be used to separate easily. Kernel functions map the data to a different, often higher dimensional space with the expectation that the classes are easier to separate after this transformation.

4) Extra Tree: Extra tree is an ensemble machine learning technique that combines predictions from decision tree.

5) Extreme Gradient Boosting: Extreme gradient boosting also known as XGBoost it is a specific implementation of the gradient boosting method which uses more accurate approximations to find the best tree model. XGBoost is exceptionally successful particularly with structured data. The most important are computation of second-order gradients that is second partial derivatives of loss function provides more information about the direction of gradients and minimization of the loss function. Advanced regularization improves model generalization. Amongst the algorithms, advantages over other machine learning models are faster training and the ability to be parallelized or distributed across clusters.

6) Gradient Boosting: Gradient boosting is a boosting algorithm. The principle behind boosting algorithms is first we built a model on the training dataset, then a second model is built to rectify the errors present in the first model. The goal is to build models sequentially to reduce the errors of the previous model by building a new model on the errors or residuals of the previous model. Gradient boosting uses the loss function of a base model like the decision tree as a proxy for minimizing the error of the overall model.

7) Random Forest: Random forest is also an ensemble machine learning technique but differ from extra tree algorithm by choosing the optimum splits. Random forests grow many classification trees. A new object is classified from an input vector by putting the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest). The random forest is unexcelled in accuracy among current algorithms, runs efficiently on large datasets, and handles thousands of variables. Estimates of important variables in the classification model can be provided by this algorithm. Random forest algorithms have three main hyperparameters, which need to be set before training. The hyperparameters node size, the number of trees, and the number of features sampled are set before training. A random forest classifier is not likely to overfit the model because the uncorrelated trees lower the overall variance and prediction error.

8) Adaptive Gradient Boosting: Adaptive boosting also known as AdaBoost is an ensemble learning method that uses an iterative approach to learn from the mistakes of weak classifiers and turn them into strong ones. Boosting basically reduce the bias error which arises when models are not able to identify relevant trends in the data.

9) Multilayer Perceptron Neural Networks: A multilayer perceptron is called Neural Networks. A perceptron, a neuron's computational model, is graded as the simplest form of a neural network. Frank Rosenblatt invented the perceptron at the Cornell Aeronautical Laboratory in 1957. The aim of this learning problem is to use data with right labels for making more accurate predictions on future data and then helps for training a model. The perceptron mainly consists of four parts that is the input values or layer, weights and bias, net sum, and activation function.

10) Gaussian Naïve Bayes: NB are a group of supervised machine learning classification algorithms based on the Bayes theorem for calculating probabilities and conditional probabilities with an assumption of independence among predictors. NB classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature and are particularly useful for very large datasets. NB algorithm is suitable for real-time prediction, multiclass prediction, recommendation system, sentiment analysis, and spam filtering.

11) Stacking Ensemble: Stacking is an ensemble learning technique where the predictions of multiple base or level

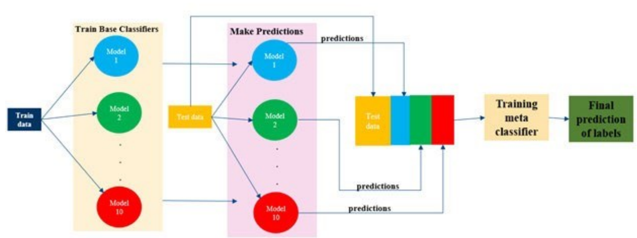


Fig. 5. Stacking ensemble model.

one classifiers multiple classifiers are used as new features to train a meta-classifier for higher predictive performance. The algorithm for the stacking ensemble is presented in Algorithm 1. The visual representation of stacking procedure is presented in Fig. 5.

Algorithm 1 Stacking Ensemble

Input: Training data D $=\{A_i, y_i\}_i^n{}_{=1}$

$(a_{n,m} \in A_{N,M} \dashrightarrow A_i \in A,\ y_i \in Y)$

Output: Ensemble Classifier G for

Baseclassi f iers = 1,*T* do.

Store predictions as $D' = \{P(D_1'), P(D_2')\ldots P(D_T')\}$

Append Training data to new dataset of classifier predictions $(D \cup D')$.

end for

Learn a meta classifier G' based on D'. Introduce a permutation scheme to find metaclassifier and combinations of base classifiers for optimal accuracy.

F. Experiments

Experiments were conducted at two different locations at North Carolina A and T to evaluate the performance of the proposed methodology in the prediction of occupancy and thermal preference. Two kinds of data were collected in different indoor environments.

1) Occupancy Prediction: The occupancy dataset was collected in the Graham Hall HVAC laboratory between February and March 2022. This dataset consisted of time-stamped thermal images captured every minute when the laboratory was occupied or not. At the same time, environmental parameters corresponding to the time stamped thermal images were streamed every minute. The 2419 images with corresponding environmental parameters were collected. The detail description of the occupancy prediction dataset is presented in Table I. Fig. 7 presents thermal images captured in the laboratory. The method of cross-validation was implemented by dividing training dataset into ten stratified folds. The first nine folds were used to train a model, and the holdout tenth fold is used as the test set. This process is repeated and each of the folds is given an opportunity to be used as the holdout test set. A total of ten models were fit and evaluated, and the performance of the model was calculated as the mean of these runs. Stratified *k*-fold sampling is a variation of *k*-fold which places approximately the same percentage of samples of each target class as in the complete dataset in each of the training and testing sets. Furthermore, the method of hold-out validation was implemented by splitting up the occupancy and thermal preference dataset into a train and test set. The model was trained on the training set and the test set was used to

TABLE I

Total Number of Samples Obtained for Each Sensor for Occupancy Dataset

sensor	Occupied	Not occupied	Total
Temperature	915	1504	2419
Humidity	915	1504	2419
Carbon dioxide	915	1504	2419
Lepton 3.5	915	1504	2419

TABLE II

OCCUPANCY TRAINING AND TESTING DATASET

Category name	Index	Training samples	Testing samples
Occupied	0	1548	162
Not occupied	1	387	322

TABLE III

TOTAL NUMBER OF SAMPLES OBTAINED FOR EACH SENSOR FOR THERMAL PREFERENCE DATASET

Sensor	Cooler	No change	Warmer	Total
Temperature	59	67	96	222
Humidity	59	67	96	222
Carbon dioxide	59	67	96	222
Lepton 3.5	59	67	96	222

see how well the model performed on the unseen data. In this study, 80% of data was used for training and the remaining 20% of the data was used for testing for the hold-out method as presented in Table II.

2) Thermal Preference Prediction: Thermal preference dataset was collected in the Artificial Intelligence Laboratory in Hall between February 2021 and September 2021. Fig. 6 presents Hines Hall Artificial Intelligence Laboratory Set Up, 3-D Laboratory layout, analog sensors, Flir Lepton 3.5, and thermal images captured in the Hines Hall laboratory. The 222 timestamped thermal images with corresponding environmental parameters and survey response on the thermal preference of subjects recruited for the experiment was collected. The detail description of the occupancy prediction dataset is presented in Table III. The thermal preference dataset is divided into 80% for model training and 20% for testing. Stratified sampling and hold out validation are used separately when splitting the data into the training and test sets for the thermal preference dataset too. Table IV presents the details for the training and testing samples.

IV. RESULTS

A. Performance Analysis

Performance of the classification algorithms for occupancy and thermal preference prediction has been measured and compared in terms of recall, precision, f1-scores, accuracy, and area under the curve (AUC) for both stratified cross validation and hold out validation methods. Recall is the proportion of the actual positive class members that are correctly identified in a classification machine learning problem as presented in the following equation:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \tag{5}$$

Precision is the proportion of positive identifications that are correct and is defined as follows:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}. \tag{6}$$

TABLE IV

THERMAL PREFERENCE TRAINING AND TESTING DATASET

Category name	Index	Training samples	Testing samples
Warmer	0	77	19
No change	1	54	13
Cooler	2	47	12

The f1-score is the harmonic mean of precision and recall, which is defined in the following equation:

$$F_1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{7}$$

The AUC is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve. Cohen's kappa is a metric often used to assess the agreement between two raters. It can also be used to assess the performance of a classification model. Cohen's kappa takes imbalance in class distribution into account and provides a more objective description of the model performance in problems where accuracy fails because of severe class imbalance. The kappa coefficient gives an indication if a classifier is performing over the performance of a classifier that simply guesses at random according to the frequency of each class. Landis and Koch (1977) provided a way to characterize the kappa coefficient values [33]. According to the authors' scheme a valueless than zero is indicating no agreement, 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement [33]. The Cohen's kappa coefficient is expressed in the relation in the following equation:

$$\kappa = \frac{\rho_0 - \rho_e}{1 - \rho_e} \tag{8}$$

where ρ_0 is the relative observed agreement among raters and ρ_e is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly seeing each category.

B. Occupancy Prediction

The results show that with an accuracy of 99.2% for tenfold cross validation, extra trees algorithm has the best performance among the tested algorithms for occupancy prediction as presented in Table V. Also, AdaBoost Algorithm has the best accuracy for the hold-out method. The ten-fold cross validation method is a computationally intensive process; it took a fraction of a second to implement a hold-out AdaBoost occupancy model while the ten-fold cross validation of the same algorithm was executed in more than 78 s. The first principal component, second principal component, temperature, humidity, third principal components, and minute are the top six important features for occupancy prediction based on the AdaBoost feature importance. Table VI presents the hyperparameter tuned for all occupancy prediction models. Table VII presents the results from the implementation of the stacking methodology discussed in Section III. The highest accuracy was achieved when *K*-nearest neighbor, extra tree, and CART were utilized as base learners and SVM as the meta classifier for the occupancy prediction.

C. Thermal Preference Prediction

XGBoost algorithm achieved the best performance for thermal preference prediction with mean accuracy and F1 score of 80.5% and 79.90% for ten-fold cross validation, respectively as presented in Table VIII. Furthermore, XGBoost achieved the best performance accuracy and F1 score of 86.70% and 86.10%, respectively for hold-out validation of thermal preference prediction. The temperature, minute, humidity, first, second, and forth principal components are the top six important features for thermal preference prediction based on the XGBoost feature importance. Table IX presents the hyperparameter tuned for all thermal preference prediction models.

The SVM classifier was again utilized as the meta classifier in the stacking methodology in all combinations for thermal preference prediction. Table X presents the results. The highest accuracy was achieved when AdaBoost, extra tree, and random forest were utilized as base learners and SVM as the metaclassifier for the thermal preference prediction. It was observed that the predicted probabilities were more efficient as compared to the predicted class labels for the stacking ensembles. An improvement of about 9% more when compared to the highest accuracy of the individual models.

V. Discussions

This work introduced a novel methodology for occupancy and thermal preference prediction based on fusion of time stamped thermal images and environmental parameters for automated HVAC control. Wavelet features obtained from scenery thermal images were fused with temperature, humidity, and carbon dioxide concentration based on timestamps to ascertain the thermal environment of building occupants. The occupancy of a building was classified into one of two classes (that is whether the building is occupied or not). Thermal preference of individuals is classified into one of three classes: should the thermal environment be cooler, no change, or warmer. Through the experiments, we have shown the possibility of real-time prediction of occupancy and thermal preference through classical machine learning, ensemble, and stacked models with high accuracy.

The performance analysis reiterates that the waveletscattering coefficients for different classes contain detailed information. Wavelet scattering performs well even with smaller dataset, unlike deep learning that requires huge dataset and a lot of time to train. The architecture of the waveletscattering network was optimally designed by selecting an ideal invariance scale for improved results in occupancy and thermal preference prediction. Additionally, hyperparameters for all machine learning models were tuned for best results.

Previous work done by other authors extracted temperature values from the thermal images for modeling through sensor equations provided by manufacturers of the thermal cameras which may be prone to errors. The wavelet-scattering feature extraction methodology investigated in this study does not require the utilization of temperature values of occupants which alleviates the errors likely to be introduced.

In the area of thermal preference research, a recent work by Martins et al. [30] trained an ANN model based on dry bulb

TABLE V

EVALUATION METRICS OF TEN DIFFERENT OCCUPANCY MODELS WITH TEN-FOLD CROSS VALIDATION AND HOLD OUT METHODS

Model	Accuracy	F1	Precision	Recall	AUC	Kappa Coefficient	Time (s)
KNN	0.991	0.991	0.991	0.991	0.989	0.980	8.496
CART	0.985	0.985	0.985	0.984	0.983	0.969	9.143
SVM	0.985	0.985	0.985	0.985	0.992	0.969	15.445
ET	0.992	0.992	0.992	0.992	0.996	0.982	16.552
XGB	0.991	0.990	0.991	0.991	0.996	0.981	39.008
GB	0.990	0.990	0.989	0.989	0.996	0.976	58.592
RF	0.991	0.991	0.990	0.991	0.998	0.981	63.518
ADABOOST	0.985	0.985	0.985	0.985	0.996	0.967	78.449
MLP	0.990	0.990	0.990	0.989	0.998	0.977	184.662
GNB	0.971	0.925	0.926	0.926	0.963	0.842	189.029
* KNN	0.986	0.986	0.986	0.986	0.982	0.967	0.016
*CART	0.971	0.971	0.971	0.971	0.969	0.935	0.016
*SVM	0.979	0.979	0.979	0.979	0.975	0.953	0.037
*ET	0.970	0.970	0.970	0.970	0.978	0.977	0.220
*XGB	0.983	0.983	0.983	0.983	0.981	0.962	0.134
*GB	0.981	0.981	0.981	0.981	0.977	0.958	0.409
*RF	0.983	0.983	0.983	0.983	0.981	0.963	1.972
*ADABOOST	0.988	0.988	0.988	0.988	0.986	0.972	0.138
*MLP	0.979	0.979	0.979	0.979	0.976	0.953	2.019
*GNB	0.975	0.975	0.975	0.975	0.974	0.944	0.001

'*' Evaluation for Hold Out method.

TABLE VI

HYPERPARAMETER TUNING FOR OCCUPANCY MODELS

Models	Hyper Parameters Tuned
KNN	$n\ neighbors = 1$
CART	$criterion = entropy, max\ depth = 7, min\ samples\ leaf = 1, min\ samples\ split = 5$
SVM	$gamma = 1, C = 1, kernel = rbf, probability = True$
ET	$criterion = gini, max\ depth = 15, max\ features = sqrt, min\ samples_{s}plit = 2, n\ estimators = 50.$
XGB	$max\ depth = 12, learning\ rate = 0.1, colsample\ bytree = 0.5, min\ chid\ weight = 1, n\ estimators = 100$
GB	$learning\ rate = 1, max_{d}epth = 5, n\ estimators = 250$
RF	$n\ estimators = 110, max\ features = 1$
ADABOOST	$algorithm = SAMME.R, learning\ rate = 1.0, n\ estimators = 1000$
MLP	$alpha = 0.05, max\ iter = 200, activation = tanh, learning\ rate = constant, solver = adam.$

TABLE VII

STACKING MODELS FOR OCCUPANCY PREDICTION

Stacked Models	Base Classifiers	Accuracy
Occupancy-Stack-1	KNN and CART	0.971
Occupancy-Stack-2	KNN and AdaBoost	0.980
Occupancy-Stack-3	KNN and ET	0.971
Occupancy-Stack-4	KNN and RF	0.986
Occupancy-Stack-5	CART and AdaBoost	0.986
Occupancy-Stack-6	CART and ET	0.971
Occupancy-Stack-7	CART and RF	0.986
Occupancy-Stack-8	AdaBoost and ET	0.971
Occupancy-Stack-9	AdaBoost and RF	0.986
Occupancy-Stack-10	ET and RF	0.985
Occupancy-Stack-11	KNN, DT and AdaBoost	0.994

Occupancy-Stack-12	KNN, CART and ET	0.994

temperature, radiant temperature, relative humidity, air speed, corrected metabolic rate, clothing level, and health status for older people's thermal preference prediction. The average accuracy of the individualized models was reported to be 74%, Cohen's kappa coefficient of 0.61 and AUC of 0.83 [30]. Another recent work published by Liu et al. [28], four machine learning models namely, SVM, CART, RF, and KNN were investigated. The highest performing model was random forest with accuracy performance of 85% with self-adaptive behavior, indoor temperature, and relative humidity [28]. Previous study by Aguilera et al. [29] trained ANN, NB and Fuzzy logic

Fig. 6. Hines Hall Artificial Intelligence Laboratory Set Up. (a) 3-D Laboratory layout, analog sensors, and Flir Lepton 3.5. (b) and (c) Thermal images.

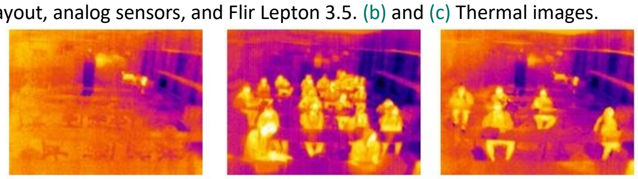


Fig. 7. Graham Hall HVAC laboratory.

machine learning models to predict thermal preference on three categories based on air temperature and relative humidity as features, the reported best AUC performance of 73% during evaluation was from NB. If the proposed approach is integrated in the HVAC control of the laboratory mentioned earlier during the problem statement, 45% more energy would be

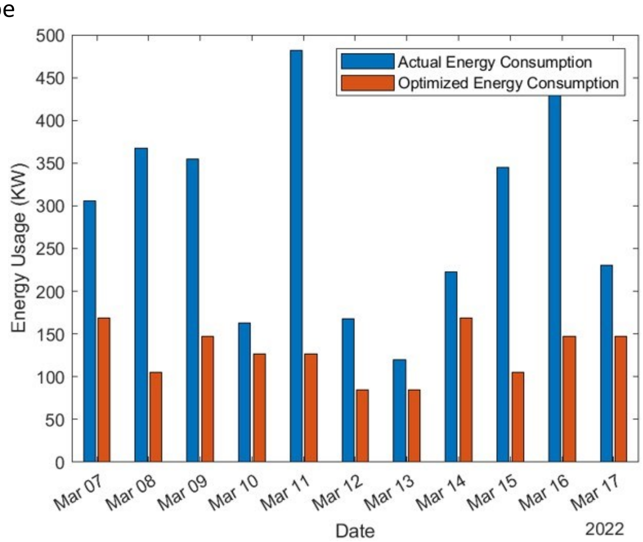


TABLE VIII
EVALUATION METRICS OF THERMAL PREFERENCE MODELS WITH TEN-FOLD CROSS VALIDATION AND HOLD OUT METHODS

Model	Accuracy	F1	Precision	Recall	AUC	Kappa Coefficient	Time (s)
KNN	0.767	0.761	0.778	0.767	0.825	0639	3.329
CART	0.722	0.707	0.720	0.706	0.780	0.547	3.715
SVM	0.712	0.708	0.723	0.712	0.856	0.557	4.441
ET	0.776	0.772	0.782	0.783	0.909	0.662	5.259
XGB	0.805	0.799	0.818	0.805	0.920	0.696	22.284
GB	0.770	0.779	0.798	0.779	0.907	0.664	36.251
RF	0.793	0.790	0.803	0.785	0.910	0.667	39.460
ADABOOST	0.695	0.698	0.739	0.695	0.870	0.537	47.547
MLP	0.767	0.748	0.763	0.762	0.855	0.604	74.716
GNB	0.663	0.651	0.681	0.663	0.846	0.472	75.096
* KNN	0.733	0.730	0.761	0.733	0.788	0.591	0.010
*CART	0.689	0.683	0.680	0.689	0.784	0.529	0.010
*SVM	0.622	0.611	0.617	0.622	0.830	0.427	0.494
*ET	0.756	0.742	0.785	0.756	0.952	0.626	0.655
*XGB	0.867	0.861	0.901	0.867	0.959	0.800	0.186
*GB	0.822	0.818	0.823	0.822	0.937	0.732	0.154
*RF	0.622	0.611	0.617	0.622	0.830	0.427	0.154
*ADABOOST	0.800	0.794	0.815	0.800	0.882	0.696	0.078
*MLP	0.756	0.742	0.784	0.756	0.910	0.626	0.655
*GNB	0.778	0.775	0.803	0.776	0.886	0.660	0.010

'*' Evaluation for Hold Out method.

TABLE IX
HYPERPARAMETER TUNNING FOR THERMAL PREFERENCE MODELS

Models	Hyper Parameters Tunned
KNN	$n\ neighbors$ = 1
CART	$criterion = entropy, max\ depth = 8, min_samples\ leaf = 2, min\ samples_split = 5$
SVM	$gamma = 1, C = 1, kernel = poly, probability = True$
ET	$min\ samples_split = 2, n\ estimators = 10.$
XGB	$max\ depth = 12, learning\ rate = 0.1, colsample\ bytree = 0.5, min\ chid\ weight = 1, n\ estimators = 100, subsample = 0.75$
GB	$learning_rate = 1, max_depth = 5, n\ estimators = 250$
RF	$n\ estimators = 40, max\ features = 2$
ADABOOST	$algorithm = SAMME.R, learning\ rate = 1.0, n\ estimators = 100$
MLP	$alpha = 0.0001, max\ iter = 200, activation = relu, learning\ rate = adaptive, solver = adam.$

Fig. 8. Energy Savings for ten days period.

saved. Instead of conditioning the laboratory at all times, full conditioning would thus be based on the presence of occupants and thermal preference. There are two limitations to this study: first, the work presented did not take into consideration the thermal preference of occupants who are closer to ceiling diffusers. There exists a possible relationship between thermal preference of occupants and distance from ceiling diffusers.

TABLE X
STACKING MODELS FOR THERMAL PREFERENCE PREDICTION

Stacked Models	Base Classifiers	Accuracy
Preference-Stack-1	KNN and AdaBoost	0.899
Preference-Stack-2	KNN and ET	0.899

Preferencey-Stack-3	KNN and RF	0.899
Preference-Stack-4	AdaBoost and ET	0.899
Preference-Stack-5	AdaBoost and RF	0.899
Preference-Stack-6	ET and RF	0.899
Preference-Stack-7	KNN, ET and AdaBoost	0.932
Preference-Stack-8	KNN, AdaBoost and RF	0.930
Preference-Stack-9	KNN, ET and RF	0.936
Preference-Stack-10	AdaBoost, ET and RF	0.946

Proximity to ceiling diffusers will therefore be incorporated as a parameter in future work. Second, thermal preferences presented in this study are subjective votes collected from individual occupants in a room. The subjective measurements of a group of occupants are therefore not accounted for in this work. Also, some of the subjective measurements may be prone to errors which the authors do not have control over.

VI. Conclusion

This study proposed an HVAC control methodology based on occupancy and thermal preference. Preliminary experiments and performance evaluations have been conducted and it can be concluded that fusion of multisensor readings provides a holistic understanding of an individual's thermal environment. Individualized thermal preference prediction is necessary for improved thermal comfort and energy efficiency. The experiments illustrate that the occupancy and thermal preference models can provide accurate predictions for smart HVAC control. In so doing, the thermal environment will be conditioned only when is occupied and the thermal preference of occupants demands attention for comfort. Future work will expand thermal preference dataset to train deep learning models for comparative analysis. Additionally, we plan to include features like proximity to vents and adaptive human behaviors to the proposed framework. Furthermore, the subjective measurements of group of occupants would be incorporated in the framework as mean votes to ascertain how that affects the evaluation metrics. We analyzed the energy savings to be achieved with the occupancy and thermal preference prediction framework in place. Data collected from the HVAC system were leveraged to estimate energy usage. The size of the laboratory is approximately 1500 square footage and according to the rule of thumb the range is between 150–250 sf/ton for elementary, high school, and college building usage type. The extreme number in the range is used for computation because it is an old building with no insulation. There is no heat transfer through the corridor because is conditioned and the only heat transfer is on the exterior wall. The estimated maximum overall tonnage and electrical usage is six tons which is equivalent to 72000 British thermal unit per hour (Btu/h) equivalent to 21.096 KW. Similarly, this means estimated 0.3516 and 0.00586 KW energy is consumed every minute and second, respectively, in the laboratory. Computation of energy usage during the ten days period is achieved by the aggregation of the widths of the peaks in Fig. 1 daily. The daily energy savings achieved by the study approach is presented in Fig. 8. An estimated 45% energy savings was realized during the ten days of analysis.

Acknowledgment

The authors would like to thank the partial funding from the sponsoring agency, United States Department of Commerce (USDOC), economic development administration good jobs challenge Awardee, STEPS4GROWTH. Weblink : https://eda.gov/arpa/good-jobs-challenge/awardees/North-Carolina-Agric ultural-and-Technical-State-University.htm. This graduate student research was funded by a National Centers of Academic Excellence in Cybersecurity Grant (H98230-21-1-0326), which is part of the National Security Agency. The authors would also like to thank the Sponsoring Agency National Science Foundation's Engineering Research Center (NSF-ERC) "The Engineering Research Center, Hybrid Autonomous Manufacturing, Moving from Evolution to Revolution (HAMMER).

References

[1] R. J. De Dear and G. S. Brager, "Developing an adaptive model of thermal comfort and preference," *ASHRAE Trans.*, vol. 104, pp. 145–167, Jan. 1998.

[2] P. O. Fanger, *Thermal Comfort: Analysis and Applications in Environmental Engineering*. Copenhagen, Denmark: Danish Technical Press, 1970.

[3] J. Van Hoof, "Forty years of Fanger's model of thermal comfort: Comfort for all?" *Indoor Air*, vol. 18, no. 3, pp. 182–201, 2008.

[4] J. F. Nicol and M. A. Humphreys, "Adaptive thermal comfort and sustainable thermal standards for buildings," *Energy Buildings*, vol. 34, no. 6, pp. 563–572, 2002.

[5] *Thermal Environmental Conditions for Human Occupancy*, ANSI/ASHRAE Standard 55-2017, 2017.

[6] X. Zhou, L. Xu, J. Zhang, L. Ma, M. Zhang, and M. Luo, "Development of data-driven thermal sensation prediction model using qualitycontrolled databases," *Building Simul.*, vol. 15, no. 12, pp. 2111–2125, Dec. 2022.

[7] Y. T. Acquaah, B. Gokaraju, R. C. Tesiero, and G. H. Monty, "Thermal imagery feature extraction techniques and the effects on machine learning models for smart HVAC efficiency in building energy," *Remote Sens.*, vol. 13, no. 19, p. 3847, Sep. 2021.

[8] Y. Acquaah, J. B. Steele, B. Gokaraju, R. Tesiero, and G. H. Monty, "Occupancy detection for smart HVAC efficiency in building energy: A deep learning neural network framework using thermal imagery," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2020, pp. 1–6.

[9] M. Luo et al., "Comparing machine learning algorithms in predicting thermal sensation using ASHRAE comfort database II," *Energy Buildings*, vol. 210, Mar. 2020, Art. no. 109776.

[10] Z. Wang, J. Wang, Y. He, Y. Liu, B. Lin, and T. Hong, "Dimension analysis of subjective thermal comfort metrics based on ASHRAE global thermal comfort database using machine learning," *J. Building Eng.*, vol. 29, May 2020, Art. no. 101120.

[11] X. Zhou et al., "Data-driven thermal comfort model via support vector machine algorithms: Insights from ASHRAE RP-884 database," *Energy Buildings*, vol. 211, Mar. 2020, Art. no. 109795.

[12] H. Park and D. Y. Park, "Prediction of individual thermal comfort based on ensemble transfer learning method using wearable and environmental sensors," *Building Environ.*, vol. 207, Jan. 2022, Art. no. 108492.

[13] H. Mao et al., "A thermal comfort estimation method by wearable sensors," in *Proc. 36th Annu. ACM Symp. Appl. Comput.*, Mar. 2021, pp. 603–610. [Online]. Available: http://press-pubs.uchicago.edu/ founders/

[14] N. Morresi et al., "Sensing physiological and environmental quantities to measure human thermal comfort through machine learning techniques," *IEEE Sensors J.*, vol. 21, no. 10, pp. 12322–12337, May 2021.

[15] Z. Wu et al., "Using an ensemble machine learning methodologybagging to predict occupants' thermal comfort in buildings," *Energy Buildings*, vol. 173, pp. 117–127, Aug. 2018.

[16] Y. Wu et al., "Individual thermal comfort prediction using classification tree model based on physiological parameters and thermal history in winter," *Building Simul.*, vol. 14, no. 6, pp. 1651–1665, Dec. 2021.

[17] T. Chaudhuri, D. Zhai, Y. C. Soh, H. Li, and L. Xie, "Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology," *Energy Buildings*, vol. 166, pp. 391–406, May 2018.

[18] M. Abdulgader and F. Lashhab, "Energy-efficient thermal comfort control in smart buildings," in *Proc. IEEE 11th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2021, pp. 22–26.

[19] M. Fayyaz, A. A. Farhan, and A. R. Javed, "Thermal comfort model for HVAC buildings using machine learning," *Arabian J. Sci. Eng.*, vol. 47, no. 2, pp. 2045–2060, Feb. 2022.

[20] K. Katic, R. Li, and W. Zeiler, "Machine learning algorithms applied to´ a prediction of personal overall thermal comfort using skin temperatures and occupants' heating behavior," *Appl. Ergonom.*, vol. 85, May 2020, Art. no. 103078.

[21] D. Li, C. C. Menassa, and V. R. Kamat, "Feasibility of low-cost infrared thermal imaging to assess occupants' thermal comfort," in *Proc. ASCE Int. Conf. Comput. Civil Eng.*, 2019, pp. 58–65.

[22] R. Ashrafi, M. Azarbayjani, and H. Tabkhi, "Machine learning-based automated thermal comfort prediction: Integration of low-cost thermal and visual cameras for higher accuracy," 2022, *arXiv:2204.08463*.

[23] B. Sirmaçek and M. Riveiro, "Occupancy prediction using low-cost and low-resolution heat sensors for smart offices," *Sensors*, vol. 20, no. 19, p. 5497, Sep. 2020.

[24] Y. Wu, H. Liu, B. Li, and R. Kosonen, "Prediction of thermal sensation using low-cost infrared array sensors monitoring system," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 609, no. 3, Sep. 2019, Art. no. 032002.

[25] Y. T. Acquaah, B. Gokaraju, R. C. Tesiero, III, and K. Roy, "Machine learning techniques to predict real time thermal comfort, preference, acceptability, and sensation for automation of HVAC temperature," in *Proc. 35th Int. Conf. Ind., Eng. Other Appl. Appl. Intell. Syst. (IEA/AIE)*. Kitakyushu, Japan: Springer, Jul. 2022, pp. 659–665.

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

[27] A. C. Cosma and R. Simha, "Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions," *Building Environ.*, vol. 148, pp. 372–383, Jan. 2019, doi: 10.1016/j.buildenv.2018.11.017.

[28] S. Liu, S. Schiavon, H. P. Das, M. Jin, and C. J. Spanos, "Personal thermal comfort models with wearable sensors," *Building Environ.*, vol. 162, Sep. 2019, Art. no. 106281.

[29] J. J. Aguilera, J. Toftum, and O. B. Kazanci, "Predicting personal thermal preferences based on data-driven methods," in *Proc. E3S Web Conf.*, 2019, p. 7.

[30] L. A. Martins, V. Soebarto, T. Williamson, and D. Pisaniello, "Personal thermal comfort models: A deep learning approach for predicting older people's thermal preference," *Smart Sustain. Built Environ.*, vol. 11, no. 2, pp. 245–270, Jul. 2022, doi: 10.1108/SASBE-08-2021-0144.

[31] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.

[32] V. Lostanlen. *Scattering.m—A MATLAB Toolbox for Wavelet Scattering*. Accessed: Sep. 24, 2022. [Online]. Available: https://github.com/lostanlen/scattering.m

[33] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, pp. 159–174, 1977.

Yaa T. Acquaah received the B.Sc. degree in mathematics from the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, in 2009, and the M.Phil. degree in computational nuclear science and engineering from the University of Ghana, Accra, Ghana, in 2013. She is pursuing the Ph.D. degree with the Department of Computational Data Science and Engineering (CDSE), North Carolina Agriculture and Technical State University (NCAT), Greensboro, NC, USA.

Currently, a Graduate Research Assistant, she has worked/still working on major projects around transportation, object detection, text data mining, and smart sensor network for occupancy and machine learning techniques for thermal preference-based HVAC control automation.

Balakrishna Gokaraju (Senior Member, IEEE) received the B.S. degree in electronics and communication engineering from JNTU, Hyderabad, India, in 2003, the M.S. degree in satellite remote sensing from BIT, Ranchi, India, in 2006, and the Ph.D. degree in computer engineering from Mississippi State University, Starkville, MS, USA, in 2010.

He recently visited NASA Goddard Space and Flight Center as an AI Research Faculty Member for seven months in 2019. He has extensive research background in satellite remote sensing applications and HPC, with eight years of service in various national laboratories, including High Performance Computing Collaboratory (HPCC), the Center for Advanced Vehicular Systems, the Geosystems Research Institute, MSU, and the Indian Space Research Organization (ISRO). He is an Associate Professor with the Department of Computational Science and Engineering and the Associate Director of Visualization and Computing Advanced Research Center (ViCAR), North Carolina A and T State University, Greensboro, NC, USA. His current research interests include spatiotemporal scientific visualizations, HPC, fully autonomous ground rover maneuvering, and cyber security.

Dr. Gokaraju is an Active Reviewer for various journals and conferences, including *Journal of Applied Geography* by Elsevier Publications, *International Journal of Image and Data Fusion* by Taylor and Francis, Oxford, U.K., and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS) by IEEE Geoscience and Remote Sensing Society, IEEE TRANSACTIONS IN GEOSCIENCE AND REMOTE SENSING, and IEEE LETTERS IN GEOSCIENCE AND REMOTE SENSING.

Raymond C. Tesioro, III received the B.S. degree in mechanical engineering from Rochester Institute of Technology, Rochester, NY, USA, in 1995, the M.S. degree in civil engineering, and the Ph.D. degree in computational science and engineering from North Carolina Agricultural and Technical State University, Greensboro, NC, USA, in 2010 and 2014, respectively.

He is currently an Assistant Professor with the Civil, Architectural, and Environmental Engineering Department, North Carolina Agricultural and Technical State University. He has extensive experience in executive level roles in the private sector holding titles as the President of a Systems Integration Company, the VP and the General Manager of a Global Foam Manufacturing Corporation, and other senior-level managerial positions. His research interests include HVAC systems, heat-mass transport and energy systems, modeling, analysis, optimization and control of HVAC Systems, sustainable built environment and smart capabilities in building energy systems, energy efficiency and technologies in buildings, diagnosis of cooling and heating energy systems, energy auditing of buildings, energy conversion and utilization, energy technology modeling, greenhouse gas analysis, building thermal simulation, and reliability engineering and engineering management.

Gregory Monty received the B.S.E.E. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1975, and the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1982.

He is the Director at the Center For Energy Research and Technology, North Carolina A and T State University, Greensboro, NC, USA. He is an analytical and results-oriented engineering professional with extensive experience in all aspects of research and development, startup companies, semiconductors and nanotechnology, electrical engineering, product development, project management, standards development, patents, quality, sustainability, and business development. Uses keen eye for detail to solve problems quickly and implement innovations to improve processes, ensuring continuous achievement for an organization. A creative thinker that brings visions to reality, drives a company to the forefront of the industry, and makes substantial contributions to the success of a corporation.

Kaushik Roy is a Professor and the Interim Chair with the Department of Computer Science, North Carolina A and T State University (NCAT), Greensboro, NC, USA. He has been the PI/Co-PI in research grants funded by the National Science Foundation (NSF), Department of Defense (DoD), National Security Agency (NSA), Department of Energy (DoE), and CISCO Systems. His current research is heavily focused on cybersecurity, cyber identity, biometrics, machine learning (deep learning), data science, and big data analytics. He is the Director of the Center for Cyber Defense (CCD). The CCD is designated by the National Security Agency and the Department of Homeland Security as a National Center for Academic Excellence in both Cyber Defense Education and Cyber Defense Research. He also directs the Cyber Defense and AI Laboratory.