ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2021.0359

IOT BASED ENVIRONMENT COMFORT LEVEL PREDICTION MODEL USING ENSEMBLE LEARNING

R. Vijayalakshmi and L. Jayasimman

Department of Computer Science, Bharathidasan University, India

Abstract

Predicting indoor environment comfort through machine learning algorithm is considered as an important research topic nowadays. People spend most of the time inside the building by doing some kitchen work, reading, watching TV, work in office building, learning in classroom, patients in hospital, workers in industry etc. Environment should be comfortable for healthy living. Thus, predicting the comfort level of the environment is necessary for keeping good health and well-being. Machine learning algorithm play an important role in prediction model. This paper focus on predicting the comfort of environment using machine learning classifier model. This model is used to train and improve the robustness of the model. This ensemble model is applied to reduce bias factor to enhance the stability and accuracy of the result.

Keywords:

Ensemble, Confusion Matrix, Boosting, Machine Learning, Comfort Level

1. INTRODUCTION

Environment comfort refers to a state of brain in which human satisfaction is expressed with the conditions of the environment. Internet of things react on the prevailing conditions of the environment based on the intelligent stimulus. The important aspect of the predicting the environment comfort is to ensure how efficient an individual can perform and determine the health aspects with the comfort of the environment. Environment comfort is evaluated from the sensor data collected and used for prediction model and analysis.

Environment comfort is important today in the covid19 situation for healthy and comfortable living. Comfort of the environment is analyzed and investigated from diverse viewpoint. Environment comfort is a multidimensional, complex concept with no precise definition. The preference and satisfaction of each individuals are different and hence there is a need for environment comfort prediction model.

It is very difficult to satisfy all human with same comfort level. Individuals differ in preference and choice of comfort and adaptation. Different bodily processes and mental behaviour influence the health aspect of the individual due to inadequate comfort of the environment. Hence, machine learning model is used with parameter optimization in the proposed methodology. Literature benchmark data set [2] are used to train and test for predicting the machine learning model results. From these results accuracy of the model is evaluated. Ambiance of the environment is a measure of hot or cold in the environment which affect the comfort which is a feeling of brain of the individual satisfaction expressed about the environment. Comfort level of the environment is influenced by several factors like Temperature, Relative Humidity, personal factors like metabolic rate, clothing, height, weight, age, gender etc.. Parameter selection method is used to enhance the accuracy of the machine learning model, increase model performance for high dimensional data and improve model stability. Thus ensemble model is proposed for this work.

2. RELATED WORKS

According to Kavita Srivastava [1], it is evident that temperature and relative humidity are mainly essential features to attain satisfaction of environment comfort level. In this paper various climate zones like Equatorial, Humid sub-tropical, Mediterranean, Marine, tropical savanna, west coast marine and both summer and winter seasons in naturally ventilated buildings have been considered from RP-884 database. The minimum set of essential parameters that influence personal comfort level are identified in this paper. In this work minimum 4 to maximum of 12 features are included for prediction model. Support Vector Machine and Naive Bayes classifier with minimum 65% for both model and a maximum of 80.98% for SVM and 79.75% for NB classifier model accuracy.

Wang et al. [3] proposed Logistic Regression and Support Vector Machine to predict thermal acceptability and thermal preference with thermal sensation and comfort with the prediction accuracy of 87.4% and 63.9% for thermal acceptability and also confirm that thermal sensation and comfort have strong predictive power. Occupants' individual difference in thermal comfort and subjects' behaviors leads to reduction of the performance accuracy. One of the limitations of the study is the selection of metrics but not discussed about the influence of sample size. Another limitation point out in this work is the use of Logistic Regression and SVM algorithm to validate the effectiveness of thermal comfort metrics and suggest advanced algorithms like XGBoost for improved performance.

Luo et al. [4] compared machine learning algorithms to predict thermal sensation and achieved accuracy of 62% using random forest classifier with three input features.

Park et al. [5] implemented thermal model by a temperature and thermal comfort index to control heat in the building. The results showed PMV based control improve occupant satisfaction compared to temperature control.

Sajjadian et al. [6] proposed a thermal comfort prediction with simple questionnaire and sample from 100 users with local assessment perspective. In this study three inputs age, activity time and clothing level are used to predict the lower and upper bound of the comfort level. No environment parameters were used for predicting the comfort.

Quintana et al. [7] proposed modified conditional Generative Adversarial Network (GAN), comfortGAN, to address the data imbalance. No attention was given to Subjective Comfort feedback.

Gao et al. [8] proposed Transfer Learning Multi-Layer Perceptron Neural Network classifier thermal comfort prediction using limited sample data with 55% accuracy. Random forest Algorithm with 51.41% accuracy and F1-score with 52.93% due to small sample taken for source and target data. The proposed work considers more data than the existing research work. Thus, the proposed work achieved accuracy of 99 % and F1-score 99%. In the existing work same climate zones are considered needs addition of other climate zones but proposed work considered different climate zones.

Kim et al. [9], proposed Personal Comfort System (PCS) model that use individual behaviour with PCS chairs for thermal comfort prediction of individual preference using machine learning algorithm. This model predicted the individual thermal comfort by comparing the temperature and humidity environmental measurements and occupant feedback obtained through survey using machine learning algorithm. Around six machine learning algorithms performance were compared using personal comfort and adaptive method and Adaboost gradient boosting classifier produced 95% prediction accuracy. The limitations of this model is the size of data set which depends upon receipt of survey feedback from each individuals reduce the model performance. Existing work is based on one time batch learning which is difficult when the data size increase which require dynamic adaptation of the model.

Thus, the proposed model with large sample from benchmark dataset and also model suitable for dynamic adaptations.

According to Qianto et al. [10], among six algorithms used and high precision models the prediction accuracy of Decision Tree algorithm model is more than 90%. Many research work done on certain environment like sleeping and outdoor environment for thermal comfort prediction but proposed work done on different environments and different age groups for thermal comfort prediction.

Li et al. [11], proposed a non-intrusive infrared thermography framework to estimate an occupant’s thermal comfort level by measuring skin temperature collected from different facial regions. From the experimental results it is evident that facial skin temperature varies with respect to change of environment temperature. Some of the limitations of the study includes, interpretation of thermal conditions with minimal interruption of the building occupants, the experiments were conducted in the heating seasons and the performance of comfort prediction models may not hold for the cooling seasons and also require advanced investigation on other seasonal climates to evaluate the skin temperature. If camera not working or blocked data collection failed. More detailed study is needed for evaluating other factors.

Table.1. Summary of Related works for Environment Comfort Modelling

Existing Work	Technique Used	Key Findings
Prediction Model for Personal Thermal Comfort for Naturally Ventilated Smart Buildings [1].	Personal comfort levels using different climate zones, summer and winter season in naturally ventilated buildings have been considered from RP-884 database.	Used minimum set of features.
Dimension analysis of subjective thermal comfort metrics based on ASHRAE Global Thermal Comfort Database using machine learning [3].	Proposed Logistic Regression and Support Vector Machine to predict thermal acceptability and thermal preference with thermal sensation and comfort with the prediction accuracy.	Limitation of the study is the selection of metrics but not discussed about the influence of sample size. Advanced algorithms are required for improved performance and also need enough samples.
Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II [4].	Applied nine ML algorithms and three data sampling methods to predict the 3-point and 7-point thermal sensation vote (TSV)	RF algorithm can achieve 63.6% overall accuracy in TSV prediction with the top three features, which is only 2.6% lower than involving 12 input features.
Using machine learning algorithms to predict occupants’ thermal comfort in naturally ventilated residential buildings [6].	Established Artificial Neural Networks model using naturally ventilated residential buildings based on input-output relationship among related factors.	More data during summer rather than other season, which cause bias in the model. And work considers naturally ventilated residential building with limited parameters.
Balancing thermal comfort datasets: We GAN, but should we? [7].	Modified conditional Generative AdversarialNetwork (GAN), comfortGAN, to address the data imbalance. The environmentaland physiological parameters are measured at a much higher frequency.	Comfort feedback is essential.
Transfer Learning for Thermal Comfort Prediction in Multiple Cities [8].	Proposed Transfer learning based multilayer perceptron model from different cities.	Considered same climate zones. Needs addition of other climate zones.
Personal comfort models: predicting individuals’ thermal preference using occupant heating and cooling behavior and machine learning [9].	Personal Comfort System (PCS) model using Adaboost Gradient Boosting Classifier	The size of data set which depends upon receipt of survey feedback from each individuals reduce the model performance and also require dynamic adaptation of the model.

Thermal Comfort models and their developments: A review [10].	Review of models applied in different environments like sleeping environment and outdoor environment. Reviewed the modelsused for different groups of people, such as elderly and different races.	Efficient predictionalgorithm combined with data is needed
Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography [11].	Random Forest Model trained using python. Hyper parameters are tuned using grid search to evaluate the accuracy for performance optimization.	Investigations on different seasonal climates are further needed to evaluate the proposed approach. Human factors such as clothing insulation and activity level were controlled and assumed constant in the experiments. Performance in such situations still requires more detailed investigations on occupants' activity level and adaptive behaviors.

3. PROPOSED SYSTEM DESIGN

The Fig.1 illustrates the workflow of the proposed Environment Comfort Level Prediction (ECLP) algorithm. Machine learning algorithm is applied to develop a mathematical model based on training data (learning) that predict results for new data (Prediction) and acclimatize the model based on new environment context. Thermal comfort is the class attribute and the classifier is correspond to predict the environment comfort on the basis of rules defined. Test data is used to predict the accuracy of the rules framed. Rules for future records are formed based on the acceptable values. Temperature and Humidity sensor values of the environment is used for comparing the independent features for making prediction of environment comfort.

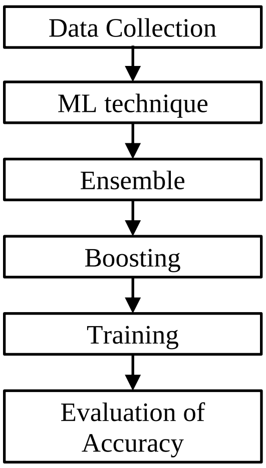

Fig.1. Environment Comfort Level Prediction

Large number of parameters are scrutinized and evaluated for splitting each node in the proposed work. Max feature size regularization parameter is used to restrict the bias in tree generation in the proposed environment comfort prediction model.

Pre-processing the data to remove the string values and keep the categorical values. Remove the column values with null values. Data cleaning is handled with removing null values and it is replaced with average values. Data is encoded after data cleaning.

Environment comfort prediction model is created with encoded data. Features are divided into target variables and feature variables. Environment comfort model is proposed by splitting the training data 80% and testing data 20 % for training the data.

This work uses boosting ensemble classifier. Gradient Boosting is a type of ensemble algorithm designed to improve the stability and accuracy of machine learning algorithms. It also reduces bias and helps to reduce variance. Multiple data subsets are created for the boosting process. This work uses Decision Tree as the base learner.

Gradient boosting is an ensemble algorithm that add models to the ensemble sequentially where in subsequent added models correct the performance of the previous models.

Gradient boosting is fairly robust to over-fitting so a large number usually results in better performance and build the model in a forward stage wise fashion. Creates a model to predict the value of target variable by learning simple decision rules from data features. Boosting decreases the bias error and builds strong predictive models and also significantly reduce variance.

3.1 STEPS FOR ENVIRONMENT COMFORT LEVEL PREDICTION (ECLP) ALGORITHM

Step 1: Import the required libraries

Step 2: Import the dataset

Step 3: Specify the parameters for data processing

Step 4: Eliminating the string columns

Step 5: Eliminating Null columns

Step 6: Encoding the cleaned Data

Step 7: Train the data using Boosting classifier

Step 8: Print the tested data

Step 9: Predict the performance of the classifier using confusion matrix, precision, recall and F1-score.

Algorithm 1: Environment Comfort Level Prediction (ECLP)

Initialize the base class libraries

Initialize the dataset $s = \{x_i, y_i\}_{i=1}^{m}$

Input Parameters: $x, x_{train}, x_{test}, y_{train}, y_{test}$

Output Parameters: *CM, ACC, RC, PRE*

Call Data Preprocessing (*del_x_rws*, *del_xobj_data*, *del_xuni*, *sel_x_rws*)

For *i*:1to *n*

del_x_rows=NULL

del_xobj_data='OBJECT'
del_xuni=*del_x_columns*('UNIQUE')
end
sel_x_rows=∑(!*del_x_rows*+*del_xobj_data*+!*del_xuni*)
Encode(*fil_x_col*)

$$(fil\_x\_col)^n = \sum_{i=0}^{n}\left(sel\_x\_rows\right)x^i a^{n-1}$$

$Train_x$[*fil_x_col*] = Encode[*fil_x_col*] as type[*int*]
#Split the dataset for training and testing
$Test_x$=*x*(*frac* = 0.2)
In the concept of Gradient Boosting we need two variables *x* and *y* where *x* is predicted variable and *y* is a response variable
$Train_x$ = ⌐(*TC*,*x*=1).values
$Test_x$ = Train(∑[*TC*].values])
$Train_y$ = ⌐(*TC*,*x*=1).values
$Test_y$=Train(∑[*TC*].values])
For *i*:1 to *n*
GB=∑($DT(Train_x)+BOO(Train_x)+SEQ(Train_x)$)
End
gb.fit(X_{train}, y_{train})
pred=gb.predict(X_{test})
Compute CM($Test_y$,*pred*)
Compute ACC($Test_y$,*pred*)
Compute RC($Test_y$,*pred*, *avg* = '*w*')
Compute PRE($Test_y$,*pred*, *avg* = '*w*')
Return CM

Algorithm 1 explains the base class libraries that are included in the program. It is followed by importing the bench mark dataset collected from [2] which consists of 1 lakh values for the thermal comfort assessment with various parameters. Machine Learning process is started to clean the data followed by training and testing the data. Gradient Boosting technique is used to obtain the predicted data results. The performance of the classification technique is tested using the confusion matrix metrics.

Table.2 Measurement of comfort Level

Sensor	**<20**		**≥20 &≤29**		**≥30 &<35**		**≥35 &<40**		**≥40**		**>50**	
	E	**P**	**E**	**P**	**E**	**P**	**E**	**P**	**E**	**P**	**E**	**P**
Temperature	-	DC	-	NDC	DC	MDC	-	SDC	-	VSDC	-	VSDC
Humidity	-	DC	-	NDC	-	MDC	-	SDC	-	VSDC	DC	VVSDC

E-Existing, P-Proposed, DC-Discomfort, NDC-No Discomfort, MDC-Moderate Discomfort, SDC-Strong Discomfort, VSDC-Very Strong Discomfort, VVSDC - Very Very Strong Discomfort

4. PERFORMANCE EVALUATION

The proposed algorithm is implemented using Intel Pentium CPU Processor with installed memory of 6 GB RAM using 64 bit Windows 7 Operating System as hardware. Python software is used for evaluation of this proposed approach.

4.1 CONFUSION MATRIX

Confusion matrix is a technique to outline the performance of the classification algorithm. Each row of the confusion matrix corresponds to a predicted class. Each column of the confusion matrix corresponds to an actual target class. Diagonal cells show the true classes of observations estimated after training the data. It depicts the match between the actual and predicted class. It also shows the difference between the actual and predicted class. It is used to check the performance per class. It also helps to identify the poor performance of the classifier. In Table.3 confusion matrix below shows the contingency table for the environment comfort of the proposed work. The diagonal element shows the correct classification for the respective class of comfort. Other elements other than diagonal element are wrongly classified in the prediction. From this the performance accuracy of the prediction can be evaluated.

Table.4. Confusion Matrix

		Actual					
		1	**2**	**3**	**4**	**5**	**6**
Predict	**1**	49	0	0	0	1	0
	2	0	172	0	2	0	0
	3	0	0	409	3	2	0
	4	0	0	0	742	29	0
	5	0	0	0	8	1397	0
	6	0	0	0	0	0	847

In the proposed environment comfort prediction model the comfort is measured on 6 point scale From 1 (very uncomfortable) to 6 (very comfortable) using Decision Tree Classifier. Thus the above confusion matrix in Table.2 shows the predicted and actual value of the comfort. In the Table.above first row and first column value 49 indicates the predicted value of the comfort scale 1 and actual value of the thermal comfort scale 1 is 49. Number of correctly classified instance is the sum of the diagonals in the confusion matrix all the others are incorrectly classified. Based on the above confusion matrix values the accuracy of the model is 98.7%. Thus out of 100 records 98 records are correctly predicted by the proposed model.

4.2 PERFORMANCE OF EVALUATION METRICS

Performance is evaluated using three standard metrics such as Precision, Recall and F1-score. These performance evaluation parameters are defined below.

4.2.1 Precision:

Precision is the ratio of the correctly predicted true classes to all the predicted true class i.e. correct and incorrect true class which is shown in the Table.4. Formula for Precision is given in Eq.(1).

$$Precision = TP/(TP+FP) \quad (1)$$

Table.4 Precision Measure

Comfort Level	**Precision**
1	1.00

2	1.00
3	1.00
4	0.98
5	0.98
6	1.00

The Fig.2 below show the class based analysis for precision measure of performance metrics for the proposed model. It range between 1.00 for level 1 to 3, 6 and 0.98 for other class level.

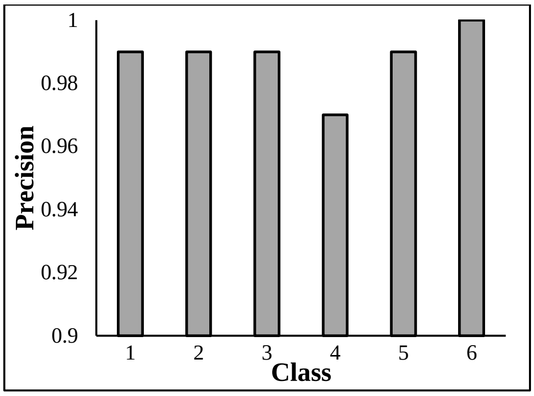

Fig.2. Class Based Analysis - Precision

4.2.2 *Recall:*

Recall is measurement is the True Positive divided by true positive plus false negative.Recall measure is shown in Table.5 below. Formula for recall measure is shown in Eq.(2) below.

$$Recall = TP/(TP+FN) \quad (2)$$

Table.5. Recall Measure

Comfort Level	Recall
1	0.98
2	0.99
3	0.99
4	0.96
5	0.99
6	1.00

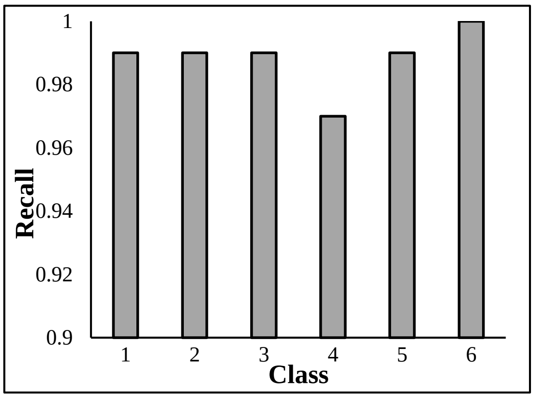

Fig.3. Class Based Analysis – Recall

The Fig.3 shows the recall measure of performance of the proposed model. It ranges between 0.96 to 1.00 for the classes which is shown graphically.

4.2.3 *F1-score:*

F1-score is a combined measure for precision and recall. It is a measure that takes both false positives and false negatives into account to strike a balance between precision and recall. F1-score measure is depicted in Table.6. Formula for F1-score is given in Eq.(3).

$$F1\text{-}score=2*Precision*Recall/(Precision+Recall) \quad (3)$$

From the Eq.(3), the F1-score is computed using Twice Precision and Recall value divided by precision+recall. From the precision table column1 row 1 1.00 and recall table column 1 and row 1 0.98 F1-score is computed as 2*1*0.98/(1+0.98) which is 0.99. Thus F1-score measure for all the classes is computed using Eq.(3). F1-score measure is depicted in Table.6.

Table.6. F1-score Measure

Comfort Level	F1-score
1	0.99
2	0.99
3	0.99
4	0.97
5	0.99
6	1.00

The Fig.4 show the class based analysis for F1-score which is the combined measure of performance metric for the proposed prediction model. It range between 0.97 to 1 for the respective classes which is shown below.

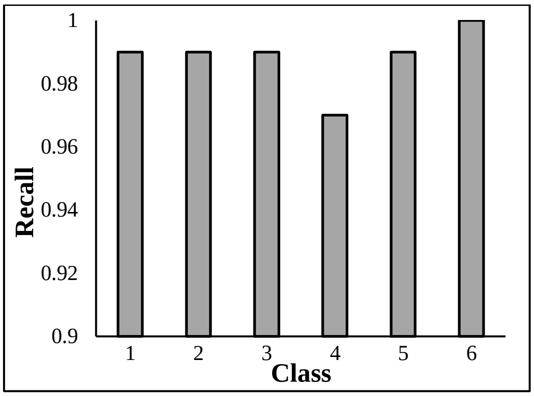

Fig.4. Class Based Analysis F1-Score

These measures described above are used for comparing the classifiers. Performance Accuracy measure is defined as the quotient of correct predictions (both True positives (TP) and True negatives (TN) predicted by a classifier model divided by the total of all predictions predicted by the classifier, including False positives (FP) and False negatives (FN). Therefore, the formula for quantifying accuracy is given below in Eq.(4).

$$\text{Accuracy} = (TP+TN)/(TP+FP+FN+TN) \quad (4)$$

where TP = True positive; FP = False positive; TN = True negative; FN = False negative

Comparison of Environment Comfort Model Accuracy with existing work is shown in Table.7. From the results it is clear that the model accuracy of the proposed model is around 12% more than the existing work.

Table.7 Comparison of Proposed Model Accuracy with Existing work

Measures	Accuracy
comfortGAN [7]	51 %
Random Forest [8]	51.41 %
TL-MLP-C [8]	54.5 %
PCS [9]	95 %
ECLP (proposed)	99 %

5. CONCLUSION

The proposed prediction algorithm supports the environment comfort level for any type of environment. The comfort level may vary for different environment but the performance and accuracy remains the same. Large samples are taken from bench mark dataset with more parameters with one lakh samples. Prediction results are achieved using machine learning technique. Proposed model performance accuracy is justified by the experimental results using confusion matrix metrics.

REFERENCES

[1] Kavita Srivastava, "Prediction Model for Personal Thermal Comfort for Naturally Ventilated Smart Buildings", *Proceedings of International Conference on Emerging Trends in Information Technology*, pp. 117-127, 2020.

[2] M. Quintana, S. Schiavon and K. Tham, "Balancing Thermal Comfort Datasets", Proceedings of ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, pp. 1-9, 2020.

[3] Zhe Wang, Jingyi Wang, Yueer He, Yanchen Liu, Borong Lin and Tianzhen Hong, "Dimension Analysis of Subjective Thermal Comfort Metrics based on ASHRAE Global Thermal Comfort Database using Machine Learning", *Journal of Building Engineering*, Vol. 29, pp. 1-19, 2020.

[4] Maohui Luo, Jiaqing Xie, Yichen Yan, Zhihao Ke, Peiran Yu, Zi Wang and Jingsi Zhang, "Comparing Machine Learning Algorithms in Predicting Thermal Sensation using ASHRAE Comfort Database II", *Energy and Buildings*, Vol. 210, pp. 1-17, 2020.

[5] Herie Park and Sang-Bong Rhee, "IoT-Based Smart Building Environment Service for Occupants' Thermal Comfort", *Journal of Sensors* Vol. 2018, pp. 1-10, 2018.

[6] Seyed Masoud Sajjadian, Mina Jafari and Direnc Pekaslan, "An Expandable, Contextualized and Data-Driven Indoor Thermal Comfort Model", *Energy and Built Environment*, Vol. 1, No. 2, pp. 385-392, 2020.

[7] Matias Quintana, Stefano Schiavon, Kwok WaiTham and Clayton Miller, "*Balancing Thermal Comfort Datasets: We GAN, But Should We?, BuildSys'20*", ACM Publisher, 2020.

[8] Nan Gao, Wei Shao, Mohammad SaiedurRahaman, Jun Zhai, Klaus David, and Flora D. Salim, "Transfer Learning for Thermal Comfort Prediction in Multiple Cities", *Proceedings of ACM/IEEE Conference on Internet of Things Design and Implementation*, pp. 1-14, 2020.

[9] Joyce Kim, Yuxun Zhou, Stefano Schiavon, Paul Raftery and Gail Brager, "Personal Comfort Models: Predicting Individuals' Thermal Preference using Occupant Heating and Cooling Behavior and Machine Learning", *Building and Environment*, Vol. 129, pp. 96-106, 2018.

[10] Qiantao Zhao, Zhiwei Lian and Dayi Lai, "Thermal Comfort Models and Their Developments: A Review", *Energy and Built Environment*, Vol. 2, No. 2, pp. 1-16, 2020.

[11] Da Li, Carol C. Menassa and Vineet R. Kamat, "Non-Intrusive Interpretation of Human Thermal Comfort Through Analysis of Facial Infrared Thermography", *Energy and Buildings*, Vol. 176, pp. 246-261, 2018.