

Smart System for Thermal Comfort Prediction on Residential Buildings Using Data-Driven Model with Random Forest Classifier

O. E. Taylor, P. S. Ezekiel, and V. T. Emmah

Abstract — Building area is a vital consumer of all globally produced energy. Structures of buildings absorb about 40 % of the total energy created which transcription about 30 % of the integral worldwide CO₂ radiations. As such, reducing the measure of energy absorbed by the building area would incredibly help the much-crucial depletions in world energy utilization and the related ecological concerns. This paper presents a smart system for thermal comfort prediction on residential buildings using data driven model with Random Forest Classifier. The system starts by acquiring a global thermal comfort data, pre-processed the acquired data, by removing missing values and duplicated values, and also reduced the numbers of features in the dataset by selecting just twelve columns out of 70 columns in total. This process is called feature extraction. After the pre-processing and feature extraction, the dataset was split into a training and testing set. The training set was 70% while the testing set was 30% of the original dataset. The training data was used in training our thermal comfort model with Random Forest Classifier. After training, Random Forest Classifier had an accuracy of 99.99% which is about 100% approximately. We then save our model and deployed to web through python flask, so that users can use it in predicting real time thermal comfort in their various residential buildings.

Index Terms — Data-Driven Model, Residential Buildings, Random Forest Classifier, Thermal Comfort.

I. INTRODUCTION

Building area is a vital consumer of all globally produced energy. Structures of buildings absorb about 40% of the total energy created which transcription about 30% of the integral worldwide CO₂ radiations [1], which transliterate about 30% of the integral worldwide CO₂ radiations. As such, reducing the measure of energy absorbed by the building area would incredibly help the much-crucial depletions in world energy utilization and the related ecological concerns. Nevertheless, the issue of energy utilization in buildings is fairly difficult despite the fact that buildings expect energy to serve in different needs. Despite the fact that there is expanding banter encompassing the chance of zero-energy buildings [2]. The ascent in world air temperatures, essentially because of environmental change, has exacerbated the issue of expanded degrees of inconvenience and warmth stress, which can bring about heat-related mortality, particularly in socioeconomics at the outrageous closures of the populace bend (i.e., both old and younger people) [3]. Pleasant indoor spaces are likewise

considerably connected to improved usefulness [4] and the general well-being of building inhabitants. Thusly, the idea of warm solace is progressively being considered in building administration rehearses. Besides the fact that accomplishing adequate solace levels in living conditions regularly requires the utilization of energy-burning-through mechanical hardware, the idea of warm solace in structures has expansive ramifications corresponding to energy use and the ensuing immediate and backhanded impacts of energy use on the climate [5]. Therefore, a vital objective in the building administration industry is accomplishing adequate warm solace levels while limiting energy utilization.

Thermal solace is the perspective which shows fulfillment with the thermal climate. Specialists have tracked down that thermal inconvenience does not just influence inhabitant productivity, work execution and commitment, yet additionally impacts deep rooted comfort. Henceforth, it is essential to keep a thermal solace climate for the comfort of the inhabitants while limiting the buildings' energy utilization [6]. Thermal solace can likewise be viewed as the state of psyche that communicates fulfillment with the warm climate and is surveyed by abstract assessment. Warm condition in buildings influences inhabitants' productivity and personal satisfaction. It's anything but an immense effect on building energy utilization [7]. Thusly, it is fundamental to assess tenants' warm solace precisely to keep a comfortable warm climate and save energy meanwhile. Absence of thermal solace in a building is a typical issue where studies uncover that up to 43% of building tenants are disappointed with the indoor warm climate, which can prompt debilitated building disorder.

Customarily, thermal solace in buildings has been surveyed and dissected utilizing predicted mean vote (PMV) list. The PMV model depends on the thermodynamic harmony among inhabitants and their prompt warm conditions. It expects that for the human body to be satisfied, there should be a warm balance between the body and its general climate. The essential objective of the PMV list is to decide the mean warm sensation vote in favour of a gathering of inhabitants; it is figured dependent on four actual constants and two individual constants. This paper presents a smart system for predicting thermal comfort on residential buildings using data-driven model with feed forward neural network.

Submitted on July 09, 2021

Published on July 30, 2021.

O. E. Taylor, Rivers State University, Port Harcourt, Nigeria.
(corresponding e-mail: taylor.onate@ust.edu.ng)

P. S. Ezekiel, Rivers State University, Port Harcourt, Nigeria
(e-mail: ezekielpromise27@gmail.com)

V. T. Emma, Rivers State University, Port Harcourt, Nigeria
(e-mail: Vikae2solo@yahoo.com).

II. RELATED WORKS

Cosma and Simha [8] presents a non-intrusive methodology for automatic forecast of individual thermal solace and interim to warm uneasiness utilizing machine learning. The expectation structure depicted utilizations of temperature data removed from various nearby body parts to show a person's warm inclination, with detecting estimations that catch neighborhood body part fluctuation just as contrasts between body parts. They compared the adequacy of utilizing machine-learning with elegant estimations, for example, skin temperature alongside their methodology of utilizing multi-part estimations and inferred information. An exploration of the presentation of machine-learning shows that their strategy improved the precision of individual warm solace expectation by a normal of 60%, and the exactness of interim to warm uneasiness forecast by a normal of 40%. The proposed warm models were tried on subjects' information separated from an office arrangement with room temperature changing from low (21.11 °C) to high (27.78 °C) [8].

Brik *et.al.* [9] present a novel techniques (Machine Learning) to foresee and control residents' warm solace by through predicted mean vote model, progressively. Their framework utilizes different linear regression algorithms and depends on discoveries from a one-year longitudinal contextual investigation of residents' thermal solace in place of business (office). They additionally propose an advanced genetic algorithm based strategy to enhance constant values of warm solace, while noticing residents' thermal inconvenience, and thus to improve the inward warm solace. The test results show the productivity of ThermCont in terms of accuracy and time intricacy when contrasted with other machine learning algorithms, notwithstanding its capacity to control and improve residents' thermal solace progressively [9].

Chai *et al.* [10] anticipated residents' thermal sensation votes and warm comfort votes on 5512 database of thermal solace information gathered in a normally ventilated personal buildings of fourteen cities in China using machine learning. Environmental values, individual values, climatic sorts, and versatile control measures were thought of and utilized as input values for the machine learning model. It was tracked down that Environmental values (both outside and inside), individual values (metabolic rate and dress protection), and climatic sorts all essentially influenced by both TCV and TSV. Contrasting and set up models (PMV, ePMV and aPMV) [10].

Chaudhuri *et al.* [11] proposed a data-driven technique to forecast discrete warm solace level (cool-inconvenience, solace, warm uneasiness) utilizing ecological and human elements as input values. Six kinds of classifiers have been executed which are logistic regression, support vector machine, Linear Discriminant analysis, k-nearest neighbors, artificial neural network, and classification trees, on an openly accessible information base of 817 residents for cooled and free-running structures independently. Results show that their methodology accomplishes forecast exactness of 73.14-81.2%, outflanking the traditional fanger's predicted mean vote model, which has an accuracy of about 41.68-65.5% [11].

Li *et al.* [12] examines the minimal expense warm camera as a non-intruding technique to deal with evaluate thermal

solace continuously, utilizing facial skin temperature. The system created can naturally identify residents, extricate facial areas, measure skin temperature, and decipher warm solace with insignificant interference or support of residents. The system is authenticated utilizing the facial skin temperature gathered from twelve residents. They trained individual solace models from various machine learning techniques are compared. Their experimental results show that random forest model can attain an accuracy of 85% and furthermore propound that the skin temperature of ears, nose, and cheeks are generally characteristic of warm solace [12].

Ngarambea *et al.* [13] present a survey of the current AI-based systems being utilized to upgrade warm solace in indoor spaces. They center on warm solace prescient models utilizing different AI (ML) techniques and their placement in building control frameworks for energy saving purposes. They also examine the gaps in the current literature and feature potential future exploration directions [13].

Salamone *et al.* [14] portrays the aftereffects of an in-field examination of warm conditions using admissible and wearable devices, and parametric models and machine learning techniques. They additionally investigated the dependability of IoT-based techniques joined with modern technique, to make a replicable structure for the appraisal and improvement of client warm fulfillment. For this reason, an experimental test in real resident buildings was done including eight laborers. Parametric models are applied for the assessment of warm solace; IoT arrangements are utilized to screen the environmental factors and the users' values; the AI CART strategy authorize to foresee the users' profile and the warm solace insight regard to the inward climate [14].

Lou *et al.* [15] developed a machine learning models of inward humidity and temperature, in view of a random autoregressive exogenous model (NARX). The developed models were utilized to figure the temperature and humidity set-focuses expected to accomplish least warm solace consistently. The outcomes showed cooling energy savings in surplus of 83% and 95%, separately, for high-and low-effectiveness in homes [15].

Wang *et al.* [16] applied machine learning methods on the information gathered in the recent released ASHRAE global thermal solace dataset in anticipating warm solace on private structures. They made use of support vector machine and logistic regression to anticipate warm admissibility and warm inclination with warm sensation and thermal comfort. The forecast accuracy is 87% for warm worthiness and 64% for warm inclination [16].

III. METHODOLOGY

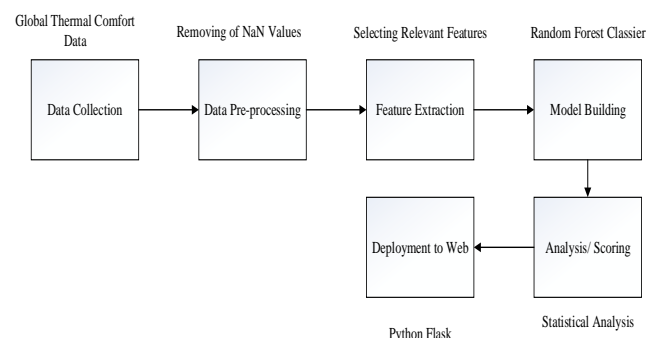


Fig 1. Architecture of the Proposed System.

The architecture of the proposed system shows the processes involved in building a smart system for predicting thermal comfort in residential buildings. More detailed description of the processes can be discussed below.

A. Data Collection

The data collected in this research is a thermal global comfort data. This data comprises of different expressions and data of human expressions about their external environment. The dataset comprises of 70 columns ranging from air temperature, relative humidity, velocity, building types, seasons, countries, cities, to thermal comfort.

B. Data-Preprocessing

In order to get a better training performance of our proposed model, the global thermal data need to pass through the stage of pre-processing. The collected data has some rows that has missing data, so we removed the missing values, in order to get a balanced data. Secondly, we converted some rows with string values (Values that are written in characters) to be zeros (0s) and ones (1s).

C. Feature-Extraction

Out of the 70 columns present in the global thermal comfort dataset, we will be selecting twelve (12) relevant features for the training of our trained model. Therefore, the processing of selecting the ten columns or features out of the 70 columns is called feature extraction. So, by features extraction, we will be creating a refined and processed dataset from the original global thermal comfort data.

D. Model Building

The model will be trained using Random Forest Classifier. The model will be trained by passing 70% of the thermal global data to the random forest classifier and 30% of the dataset will be used for testing. In order to get a better training accuracy, we will be changing the number of estimators until will finding a better training result.

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2 \dots \text{Mean Squared Error}$$

where

N represents the number of points in the data,
f_i is the value returned by the model and
y_i is the original value for data point i.

$$Gini = 1 - \sum_{i=1}^C (p_i)^2 \dots (2) \text{ Classification Problem}$$

where p_i represents the relative frequency of the classes thermal comfort data and c represents the numbers of classes

E. Analysis/ Scoring

In order to get a clearer picture of the thermal comfort various individuals in their residential areas and offices, we will be analyzing the warm comfort data by carrying out some statistical analysis like plotting of graphs, histograms, classification report and confusion metrics.

F. Deploying to Web

The trained thermal comfort model will be saved into file and will be deployed to web for conducting real time analysis

and classification of thermal comfort on residential buildings. We will be using python flask in designing a user-friendly interface where users can input various data like air temperature, relative humidity, velocities, etc.

Algorithm of the proposed system

```

Step 1: Load thermal data
Step 2: Pre-processing the dataset (similar scale and range)
Step 3: x_process = Pre-processing(X)
Step 4: y_process = Pre-processing(y)
Step 5: //Split the data into training (70%) and test (30%) set
Step6: X_train, X_test, y_train, y_test = train_test_split
Step7: (x_process, y_process, test size = 0.30)
Step8: //Import Random Forest Classifier
Step9: Rfc = RandomForestClassifier()
Step10: Rfc.fit (X_train, y_train)
Step11: predicT = Random.predict(X_test)
Step12: //Evaluate model accuracy
Step13: Evaluate (Random)
Step14: Return Random

```

IV. RESULT AND DISCUSSION

This system uses a global warm comfort data which was downloaded from kaggle.com. The dataset comprises 70 columns ranging from seasons to thermal comfort column. The dataset was pre-processed by removing duplicate values and missing values. To check for missing data, we used data.isnull().sum() functions to check for the missing values in the dataset. We removed these missing values by using data.dropna() function. After these processes, we selected twelve (12) columns out of the 70 columns by means of feature_extraction. Therefore, creating a new dataset with ten columns. These columns are seasons, koopen climate classification, building type, cooling strategy_building level, clo, met, air temperature, relative humidity, air velocity, outdoor monthly air temperature, thermal preference, and thermal comfort. Fig. 2 shows a picture of the newly extracted columns, which we used as our new dataset. The dataset was divided and allocated into two variables called x and y. The x variable contains columns ranging from season to outdoor monthly air temperature, while the y variable contains just the thermal comfort column. The thermal comfort columns were used as our target or label class. The variables being x and y were further split into a training and a testing data. The training set is made up 70% of the thermal dataset, and the testing set made up 30% of the testing data. The training data was passed to our Random Forest Classifier algorithm which will imported from sklearn.ensemble. We trained our thermal comfort model using n_estimators=100. By n_estimators, we mean the number of nodes present in the Random Forest Classifier. We set the number of nodes in the Random Forest Classifier to be 100. We trained our model by passing the training data. We got a training accuracy of about 99.99%. This can be glimpse in the classification report in Fig. 11. Analysis of the thermal dataset can be seen from Fig. 3 to Fig. 10. We then saved our trained model and deploy to web for real time thermal comfort prediction. This can be seen in Fig. 12.

	Season	Koppen climate classification	Building type	Cooling strategy_building level	Clo	Met	Air temperature (C)	Relative humidity (%)	Air velocity (m/s)	Outdoor monthly air temperature (C)	Thermal preference	Thermal comfort
74	Summer	Cfa	Office	Air Conditioned	0.57	1.0	24.3	36.8	0.27	32.8	cooler	4.5
75	Summer	Cfa	Office	Air Conditioned	0.57	1.1	25.7	33.1	0.09	32.8	cooler	5.2
76	Summer	Cfa	Office	Air Conditioned	0.57	1.1	24.6	34.9	0.06	32.8	cooler	2.5
77	Summer	Cfa	Office	Air Conditioned	0.57	1.0	26.4	31.7	0.13	32.8	cooler	2.7
78	Summer	Cfa	Office	Air Conditioned	0.57	1.0	25.0	33.3	0.07	32.8	cooler	2.2
...
81746	Winter	Cfa	Senior center	Mixed Mode	0.94	1.0	22.4	52.5	0.06	13.7	no change	5.0
81747	Winter	Cfa	Senior center	Mixed Mode	0.66	1.0	22.1	54.1	0.08	13.7	no change	5.0
81748	Winter	Cfa	Senior center	Mixed Mode	0.69	1.0	22.0	54.0	0.08	13.7	no change	5.0
81749	Winter	Cfa	Senior center	Mixed Mode	0.82	1.0	20.7	58.0	0.11	13.7	no change	5.0
81750	Winter	Cfa	Senior center	Mixed Mode	0.86	1.2	20.8	57.4	0.09	13.7	no change	5.0

10794 rows x 12 columns

Fig. 2. New thermal dataset.

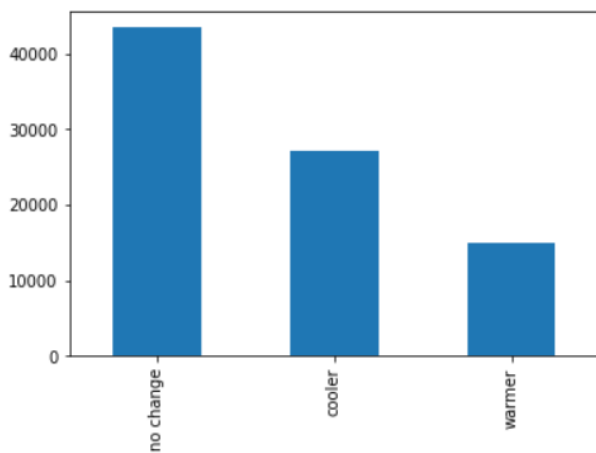


Fig. 3. Thermal preference.

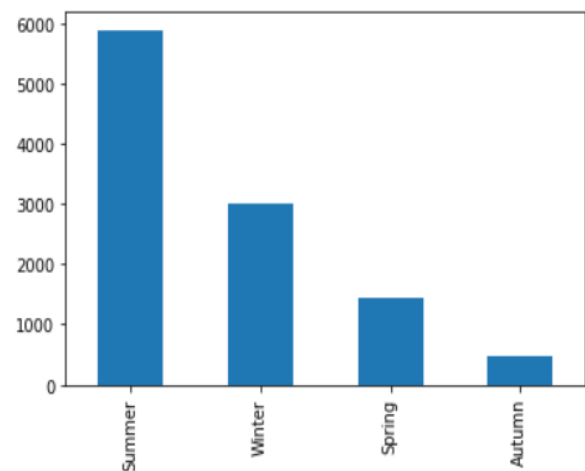


Fig. 5. Countplot of seasons in the thermal comfort data.

This shows a countplot of the number of persons that choose a particular weather over the other.

<AxesSubplot:xlabel='Thermal preference,Sex'>

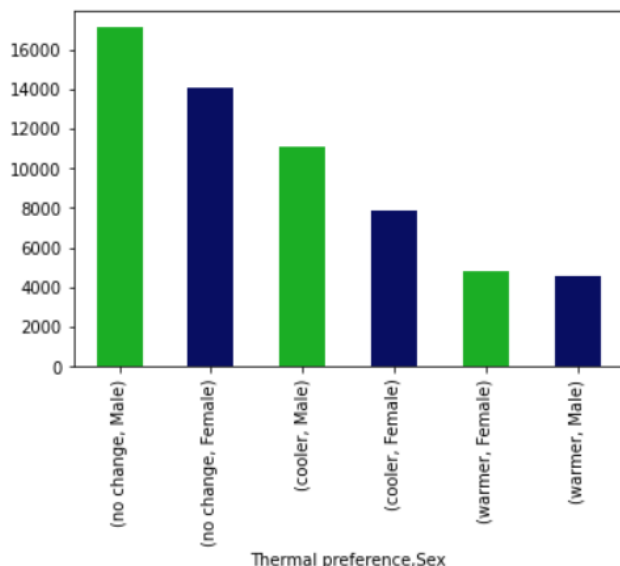


Fig. 4. Comparison of sex and thermal preference.

This shows the preferred choices made by both male and female on the particular weather they prefer.

<AxesSubplot:>

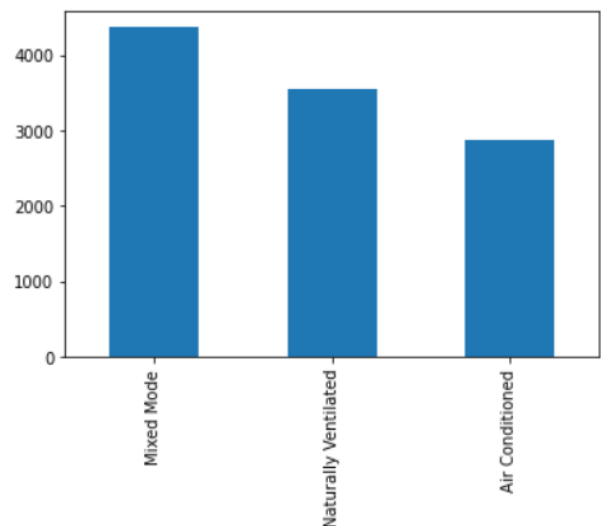


Fig. 6. Countplot on cooling strategy.

This shows the number of persons and their preferred cooling strategy.

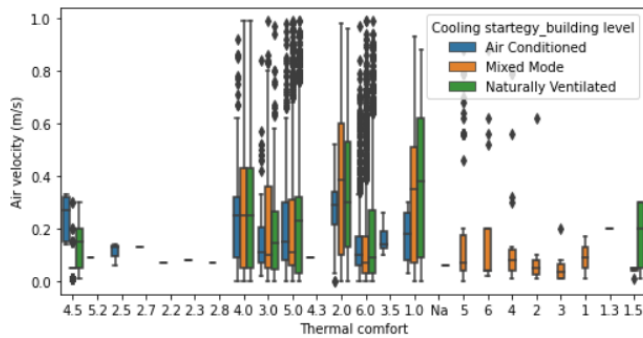


Fig. 7. Velocity vs thermal Comfort.

This shows that the lower the velocity, the better the thermal comfort.

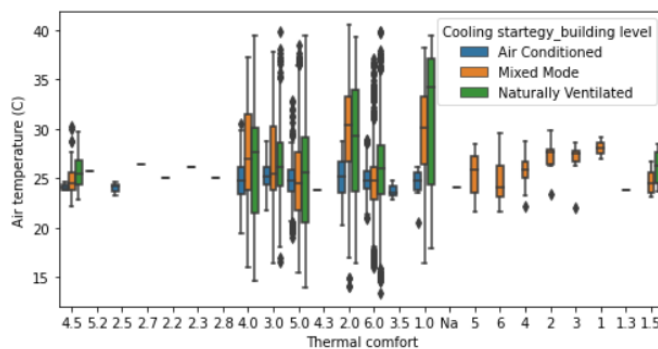


Fig. 8. Temperature vs thermal Comfort.

This shows that, the lower the temperature, the higher the thermal comfort.

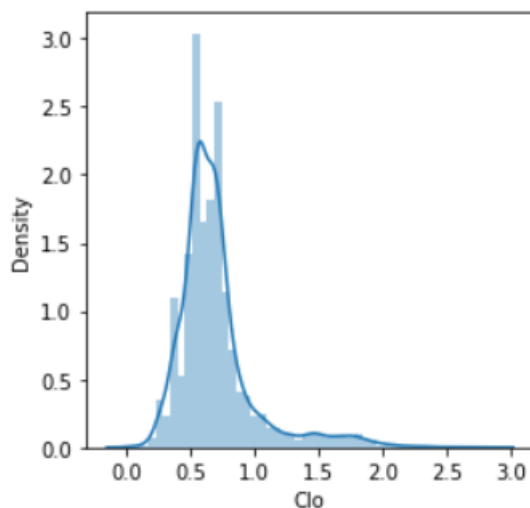


Fig. 9. Clothing Level vs Density.

	precision	recall	f1-score	support
1	1.00	1.00	1.00	113
2	1.00	0.99	1.00	328
3	1.00	1.00	1.00	547
4	1.00	1.00	1.00	1420
5	1.00	1.00	1.00	2483
6	1.00	1.00	1.00	2365
accuracy			1.00	7256
macro avg	1.00	1.00	1.00	7256
weighted avg	1.00	1.00	1.00	7256

Fig. 10. Classification Report.

This shows the classification report of the trained model with about 100% of training accuracy.

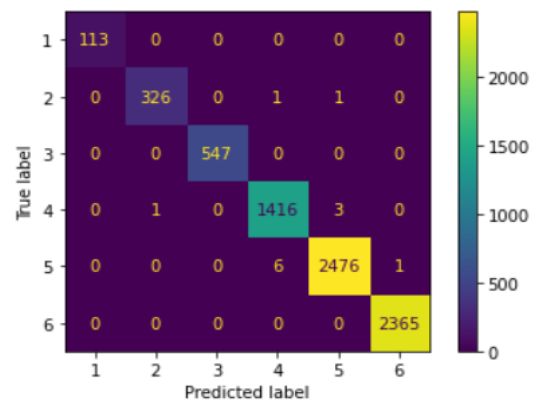


Fig. 11. Confusion Matrix Report.

This shows the True label vs the Predicted Label.

Thermal Data

Clo

0.94

Metabolism

1.0

Air Temperature

22.1

Relative Humidity

54.0

Air Velocity

0.08

Outdoor Monthly Temperature>

13.7

Season

2

Kopean Climate Classification

Building Type

1

Cooling Strategy

2

PREDICT COMFORT

Result

Very Comfortable

Fig. 12. Predicted thermal comfort data.

V. CONCLUSION AND FUTURE WORK

This paper presents a smart system for thermal comfort prediction on residential buildings using data driven model with Random Forest Classifier. The system starts by acquiring a global thermal comfort data, pre-processed the acquired data, by removing missing values and duplicated values, and also reduced the numbers of features in the dataset by selecting just twelve columns out of 70 columns in total. This process is called feature extraction. After the pre-processing and feature extraction, the dataset was split into a training and testing set. The training set was 70% while the testing set was 30% of the original dataset. The training data was used in training our thermal comfort model with Random

Forest Classifier. After training, Random Forest Classifier had an accuracy of 99.99% which is about 100% approximately. We then save our model and deployed to web through python flask, so that users can use it in predicting real time thermal comfort in their various residential buildings. This work and further be improved by using Raspberry pi in generating thermal comfort data and using Deep Feed Forward Neural Network Algorithm for model training. This paper can further be improved by creating an android application for thermal comfort analysis and prediction.

REFERENCES

- [1] P.H. Shaikh, N.B.M. Nor, P. Nallagownden, I. Elamvazuthi, T. Ibrahim, A review on optimized control systems for building energy and comfort management of smart sustainable buildings, *Renew. Sustain. Energy Rev.* vol. 34, pp. 409–429, 2014. doi:10.1016/j.rser.2014.03.027.
- [2] A.J. Marszal, P. Heiselberg, J.S. Bourrelle, E. Musall, K. Voss, I. Sartori, A. Napolitano, zero energy building – a review of definitions and calculation methodologies, *Energy Build.* vol. 43, pp. 971–979, 2011. doi:10.1016/j.enbuild.2010. 12.022
- [3] D. Mitchell, C. Heaviside, S. Vardoulakis, C. Huntingford, G. Masato, B.P. Guillod, P. Frumhoff, A. Bowery, D. Wallom, M. Allen, Attributing human mortality during extreme heat waves to anthropogenic climate change, *Environ. Res. Lett.*, vol. 11, 074006, 2016. doi:10.1088/1748-9326/11/7/074006.
- [4] Y. Geng, W. Ji, B. Lin, Y. Zhu, The impact of thermal environment on occupant IEQ perception and productivity, *Build. Environ.*, vol. 121, pp. 158–167, 2017. doi:10. 1016/j.buildenv.2017.05.022.
- [5] D. Ormandy, V. Ezratty, Health and thermal comfort: From WHO guidance to housing strategies, *Energy Policy*, vol. 49, pp. 116–121, 2012. doi:10.1016/j.enpol.2011. 09.003.
- [6] Nan G., Wei S., Mohammad S. R., Flora D. S., “Predicting in-class Emotional, Behavioural and Cognitive Engagement in the Wild”. arXiv preprint arXiv:2007.04831 (2020).
- [7] A.K. Melikov, “Advanced air distribution: Improving Health and Comfort While Reducing Energy Use”, *Indoor Air*, vol. 26, pp. 112–124, 2016.
- [8] A. C. Cosma, R. Simha, “Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions”, *Building and Environment*, vol. 148, pp. 372–383, 2019.
- [9] B. Brik, M. Esseghir, L. Merghem-Boulahia and H. Snoussi, “ThermCont: A machine Learning enabled Thermal Comfort Control Tool in a real time,” 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), pp. 294–300, 2019. doi: 10.1109/IWCMC.2019.8766697.
- [10] Q. Chai, H. Wang, Y. Zhai, L. Yang, “Using machine learning algorithms to predict occupants’ thermal comfort in naturally ventilated residential buildings”, *Energy & Buildings*, vol. 217, pp. 2–13, 2020.
- [11] T. Chaudhuri, Y. C. Soh, H. Li and L. Xie, “Machine learning based prediction of thermal comfort in buildings of equatorial Singapore,” 2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC), pp. 72–77, 2017. doi: 10.1109/ICSGSC.2017.8038552.

- [12] D. Li, C.C. Menassa, V., Kamat, “Feasibility of Low-Cost Infrared Thermal Imaging to Assess Occupants’ Thermal Comfort”, *Computing in Civil Engineering: Smart Cities, Sustainability and Resilience*, pp. 58–65, 2019.
- [13] J. Ngarambea, G. Young, M. Santamouris, “The use of artificial intelligence (AI) methods in the prediction of thermal comfort in buildings: energy implications of AI-based thermal comfort controls”, *Energy & Buildings*, vol. 211, pp. 2–15, 2020.
- [14] F. Salamone, L. Belussi, C. Currò, L. Danza, M. Ghellere, G. Guazzi, I. Meroni, I., “Integrated Method for Personal Thermal Comfort Assessment and Optimization through Users’ Feedback, IoT and Machine Learning: A Case Study”, *Sensors*, 18(5), 1602, 2018. doi:10.3390/s18051602.
- [15] R. Lou, K. P. Hallinan, K. Huang, T. Reissman, “Smart Wifi Thermostat-Enabled Thermal Comfort Control in Residences”, *Sustainability* 12(5), pp. 1–15, 2020.
- [16] Z. Wang, J. Wang, Y. He, Y. Liu, B. Lin, T. 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. 2–10, 2020.



Dr. O. E. Taylor obtained his B. Sc, M. Sc and Ph. D degrees all in Computer Science from the Rivers State University of Science and Technology, University of Ibadan and University of Port Harcourt, Nigeria respectively. He is currently Senior Lecturer in the Department of Computer science, Rivers State University, Port Harcourt, Nigeria. He is a chartered member of the Computer Professionals (Registration Council) of Nigeria and Nigeria Computer Society. His research focuses on intelligent systems, smart space, context-aware systems, machines learning algorithms and artificial intelligence. He has over ten years of teaching and research experience.



Mr. P. S. Ezekiel obtained his Bachelor of Science degree from Department of Computer Science, Rivers State University. He has published four research papers in international journals which are available online. His main research work focuses on Machine Learning, Data Science, Deep Learning and Artificial Intelligence.



Mr. V. T Emmah obtained his Bachelor of Science degree and Master of Science degree from Department of Computer Science, Rivers State University and University of Port Harcourt respectively. He is currently pursuing a Ph. D in Computer Science at the University of Port Harcourt. He is a member of the Computer Professional (Registration Council) of Nigeria. His main research work focuses on Machine Learning, Data Science, Deep Learning and Artificial Intelligence.