Architecture and Planning Journal (APJ)

Volume 28 Issue 3 *ASCAAD 2022 - Architecture in the Age of the Metaverse – Opportunities and Potentials*
ISSN: 2789-8547

Article 8

March 2023

A HYBRID APPROACH BASED ON BUILDING PHYSICS AND MACHINE LEARNING FOR THERMAL COMFORT PREDICTION IN SMART BUILDINGS

MONA M. SALEM
Faculty of Architecture - Design & Built Environment, Beirut Arab University, Lebanon, m.abdelkadersalem@bau.edu.lb

RAMY M. MOUSSA
Faculty of Computer Science and Artificial Intelligence, Pharos University in Alexandria, Alexandria, Egypt, ramy.moussa@pua.edu.eg

A HYBRID APPROACH BASED ON BUILDING PHYSICS AND MACHINE LEARNING FOR THERMAL COMFORT PREDICTION IN SMART BUILDINGS

Abstract

One of the most important challenges facing the world is the application of modern technology in order to create smart buildings that achieve sustainable development goals (SDGs). Thermal comfort and reduction of energy consumption in buildings are considered important factors which, in turn, are reflected in creating a healthy environment and improving human productivity. Internet of Things (IoT) provides an ideal solution for collecting real-time data on the factors affecting indoor thermal comfort and energy consumption. However, comfort level is subjective and depends on many factors, which may not be learned by conventional models, an integrated model depending on thermal comfort factors is needed. In this work, a hybrid physics-based model incorporated with machine learning techniques is used for the prediction of thermal comfort inside buildings. XGBoost (eXtreme Gradient Boost) algorithm method was used due to its abilities to handle complex problems. A calculated dataset was extracted from the physics-based model gathered with the environmental variables data such as humidity, moisture, temperature, and air velocity collected from IoT devices. The results show an improvement in the prediction of the thermal comfort approach as compared with the conventional models. The XGBoost algorithm can exhibit an effective solution for eliminating deficiencies of traditional models and can be used when designing smart buildings, simulating, and evaluating the designed buildings, controlling energy consumption, and achieving thermal comfort.

Keywords

Thermal comfort, machine learning XGBoost, Smart buildings, building physics, IoT.

A HYBRID APPROACH BASED ON BUILDING PHYSICS AND MACHINE LEARNING FOR THERMAL COMFORT PREDICTION IN SMART BUILDINGS

MONA M. SALEM[1] AND RAMY M. MOUSSA[2]

[1] Assistant Professor, Faculty of Architecture - Design & Built Environment, Beirut Arab University, Lebanon
m.abdelkadersalem@bau.edu.lb
[2] Assistant Professor, Faculty of Computer Science and Artificial Intelligence, Pharos University in Alexandria, Egypt
ramy.moussa@pua.edu.eg

ABSTRACT

One of the most important challenges facing the world is the application of modern technology to create smart buildings that achieve sustainable development goals (SDGs). Thermal comfort and reduction of energy consumption in buildings are considered important factors which, in turn, are reflected in creating a healthy environment and improving human productivity. Internet of Things (IoT) provides an ideal solution for collecting real-time data on the factors affecting indoor thermal comfort and energy consumption. However, comfort level is subjective and depends on many factors, which may not be learned by conventional models, an integrated model depending on thermal comfort factors is needed. In this work, a hybrid physics-based model incorporated with machine learning techniques is used for the prediction of thermal comfort inside buildings. XGBoost (eXtreme Gradient Boost) algorithm method was used due to its abilities to handle complex problems. A calculated dataset was extracted from the physics-based model gathered with the environmental variables data such as humidity, moisture, temperature, and air velocity collected from IoT devices. The results show an improvement in the prediction of the thermal comfort approach as compared with the conventional models. The XGBoost algorithm can exhibit an effective solution for eliminating deficiencies of traditional models and can be used when designing smart buildings, simulating, and evaluating the designed buildings, controlling energy consumption, and achieving thermal comfort.

Keywords: Thermal comfort, machine learning XGBoost, Smart buildings, building physics, IoT.

ملخص

من أهم التحديات التي تواجه العالم تطبيق التكنولوجيا الحديثة من أجل إنشاء مبانٍ ذكية تحقق أهداف التنمية المستدامة. وتعتبر الراحة الحرارية وتقليل استهلاك الطاقة في المباني من العوامل المهمة التي تنعكس بدورها على خلق بيئة صحية وعلى تحسين إنتاجية الإنسان. يوفر إنترنت الأشياء (IoT) حلاً مثاليًا لجمع البيانات بطريقة متزامنة حول العوامل التي تؤثر على الراحة الحرارية الداخلية واستهلاك الطاقة. ولكن مستويات الراحة لا تقاس بشكل موضوعى وتعتمد على العديد من العوامل، والتي قد لا تتعلمها بالضرورة النماذج الحوسبية التقليدية. ولذا تظهر هنا الحاجة إلى نموذج متكامل يعتمد على عوامل الراحة الحرارية. تم فى هذا البحث استخدام نموذج هجين قائم على الفيزياء مدمج مع تقنيات التعلم الآلي للتنبؤ بالراحة الحرارية داخل المباني. تم استخدام طريقة خوارزمية XGBoost (eXtreme Gradient Boost) نظرًا لقدرتها على التعامل مع المشكلات المعقدة. تم استخراج مجموعة بيانات محسوبة من النموذج القائم على الفيزياء التي تم جمعها مع بيانات المتغيرات البيئية مثل الرطوبة ودرجة الحرارة وسرعة الهواء والتي تم جمعها عن طريق أجهزة إنترنت الأشياء. تظهر النتائج تحسنا في التنبؤ بنهج الراحة الحرارية مقارنة بالنماذج التقليدية. يمكن أن تعرض خوارزمية XGBoost حلاً فعالاً لإزالة أوجه القصور في النماذج التقليدية ويمكن استخدامها عند تصميم المباني الذكية ومحاكاة وتقييم المباني المصممة والتحكم في استهلاك الطاقة وتحقيق الراحة الحرارية.

الكلمات المفتاحية: الراحة الحرارية، التعلم الآلى XGBoost، المبانى الذكية، فيزياء المبانى، انترنت الأشياء.

1. INTRODUCTION

One of the challenges facing creating smart buildings with respect to climate change that prevails in the world is achieving thermal comfort inside the building in addition to conserving energy during use (Calama-González, León-Rodríguez and Suárez, 2022). Smart educational buildings are considered one of the most construction systems that need thermal comfort during the presence of workers and students, which helps to study and focus better (Balbis-Morejon and Noya-Sambrano, 2020).

These smart buildings need advanced systems to monitor the change in the physical properties that affect the thermal comfort inside the buildings. Physical properties such as temperature, humidity, air flow velocity, and carbon dioxide ratio are among the most important factors that affect thermal comfort and energy conservation in smart buildings (Majewski et al., 2020).

The use of IoT devices which is already supported by sensors to measure changes in these factors inside buildings has become a necessity, as it can collect a lot of information throughout the day and during periods of temperature and energy usage peak (Zang, Xing and Tan, 2019). Managing this huge amount of data is one of the most important requirements to reach the optimal use of energy and achieve thermal comfort within smart buildings.

Physics-based modeling algorithms require significant computing resources and are not suitable for fast predictions. Machine learning is considered as one of the best ways to manage complex processes and link different physical variables to help make a decision to achieve indoor thermal comfort (Qavidel Fard, Zomorodian and Korsavi, 2022).

Different Techniques such as Support Vector Machine (SVM), Artificial Neural Network (ANN), and random forests (RF), decision trees (DT), genetic and Logistic Regression method (LoR), Gradient boosting (GB) and eXtreme Gradient boosting (XGBoost) are used due to their abilities to handle complex problems.

Machine learning models could generate linear equations, and with the reduced number of attributes, the predictive ability will only decline. At the same time, XGBoost, method could potentially be improved with parameter tuning. And so, this study focuses on design an optimised thermal comfort model in educational buildings based on IoT monitoring sensors and hybrid parameters of the physics based and XGBoost machine learning model. The physics-based part can make estimate thermal comfort based on physics equations that builds forecasting to map the parameters that affect the thermal comfort in buildings. XGBoost Algorithm was ran using Mathematica wolfram 13.1 software which can make a correlation between theses parameters and provide a time conserved solution during these complex processes.

2. METHOD

2.1. Overall Structure of the Model

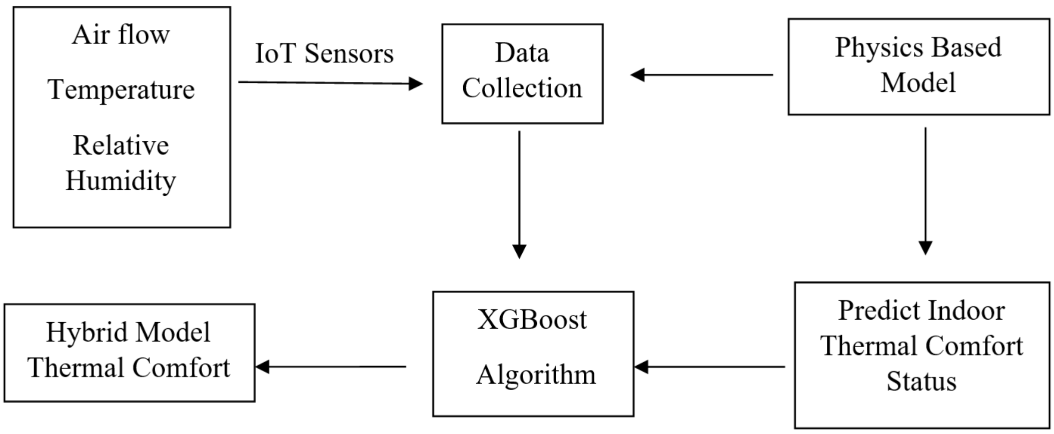

Fig.1: The hybrid physics-based/ML model.

DOI: 10.54729/2789-8547.1203

The hybrid model consists of the physics based and ML part. By integrating these two parts, the developed model can take both advantages of the physics-based method and ML method.

2.1.1 Data set

The studied data set was taken from the indoor monitoring system containing IoT sensors. The system offers an inexpensive way of monitoring the indoor environment with wifi sensors able to make real time monitoring of temperature, humidity, CO2, air flow velocity, room occupancy and light levels. The data set collected every minute during the educational day in 8 hours for three selected variables. The monitoring system was installed in the walls of the classroom. The duplicates were first isolated from the dataset. The procedure of duplicates removal was performed with the use of Python 3.10.5.

2.1.2 IoT System

The 9-wall mounted IoT devices are used to provide a real time information about different parameters that affect indoor thermal comfort in educational building Fig 2. The Data process are gathered through three layers as follow:

1. Perception layer: In this layer the IoT sensors can record the information about indoor humidity, temperature, and air velocity.
2. Network layer: The gateway devices are responsible for data routing from the previous layer.
3. Cloud layer: In this layer the data was collected, and the analysis would take place. The aggregated data that stored in this layer is a real time monitoring data. The visualization and statistics would help in taking the best decision through the statistics provided from this layer.

Fig.2: Educational classroom that used in this study

2.1.3 Extreme gradient boosting (XGBoost)

XGBoost is a type of machine learning algorithms that used as a supervised learning technique due to its efficiency and faster learning process through an ensemble algorithm based on gradient boosted trees (Chen and Guestrin, 2016).

The loss function related to XGBoost provides an additional regularization term that contributes to smoothing the final learning weights and reducing the probability of overfitting. It also considers gradients up to the second order to optimize the loss function. Additionally, to avoid overfitting problems, XGBoost also handles row and column sampling (Wang et al., 2021). The following paragraphs explain how this algorithm works.

XGBoost integrates predictions of "weak" classifiers (tree model) to achieve a "strong" classifier (tree model) via a serial training process. It can avoid over-fitting by adding a regularization term. Parallel and distributed computing makes the learning process faster to give a quicker modeling process. Figure 3 shows a schematic diagram of the computational process of XGBoost and y_i appeared in the process is calculated by Equation 1.

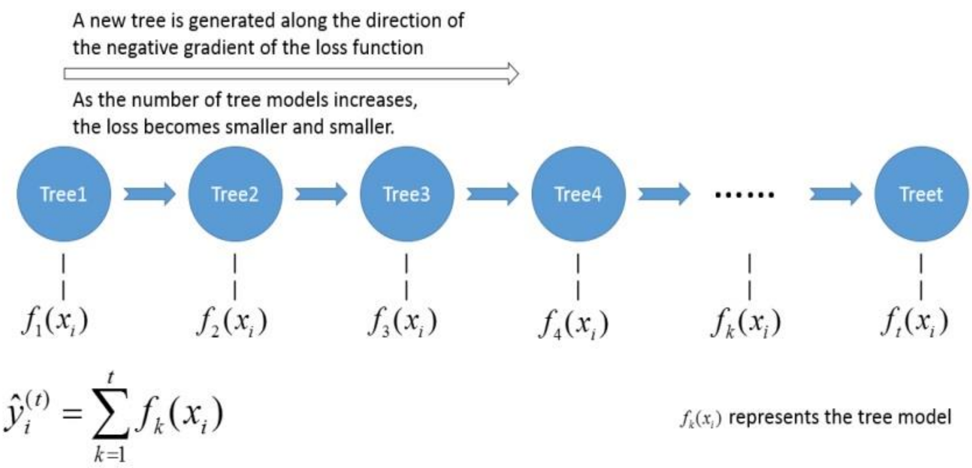

Fig.3: A schematic diagram of XGBoost computational Process

$$\hat{y}_i^{(t)} = \sum_{i=1}^{t} \boldsymbol{f_k(x_i)} = \hat{y}_i^{(t-1)} + \boldsymbol{f_t(x_i)} \tag{1}$$

where, $\hat{y}_i^{(t)}$ is the final tree model; $\hat{y}_i^{(t-1)}$ is the previously generated tree model; $\boldsymbol{f_t(x_i)}$ is the newly generated tree model, and t is the total number of base tree models. For the XGBoost algorithm, both depth and number of trees are important parameters. The problem of finding the optimal algorithm was changed into finding a new classifier that can reduce the loss function, with the target loss function shown in Equation 2

$$Obj^t = \sum_{i=1}^{t} L(\ y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^{t} \Omega(f_i) \tag{2}$$

where, y_i is the actual value; $\hat{y}_i^{(t)}$ is the predicted value; $L(\ y_i, \hat{y}_i^{(t)})$ is the loss function and $\Omega(f_i)$ is the regularization term.

Substituting Equation 1 into Equation 2 and then following some deduction steps, Equation 3 could be obtained.

$$Obj^t = \sum_{i=1}^{t} L(\ y_i, \hat{y}_i^{\,(t-1)} + \boldsymbol{f_t}(\boldsymbol{x_i})) + \Omega(f_t) + Constant \qquad (3)$$

The final target loss function was then converted into Equation 4, and the model was then trained according to this target loss function.

$$Obj^t = \sum_{i=1}^{t} [g_i\, \boldsymbol{f_t}(\boldsymbol{x_i}) + \frac{\boldsymbol{1}}{\boldsymbol{2}} \boldsymbol{h_i} \boldsymbol{f_t^2}(\boldsymbol{x_i})] \ + \Omega(f_t) \qquad (4)$$

Where $g_i = \partial_{y_i^{(t-1)}}\ l(\ y_i, y_i^{(t-1)})$, $h_i = \partial^2_{\ y_i^{(t-1)}}\ l(\ y_i, y_i^{(t-1)})$ are the first and second order gradient statistics on the loss function.

The regularization term $\Omega(f_t)$ is calculated by Equation 5 to reduce the model's complexity and improve its usability to other datasets.

$$\Omega(f) = \boldsymbol{\gamma}\, \mathbf{T} + \frac{1}{2}\lambda + \|\omega\|^2 \qquad (5)$$

where, T is the number of leaves; ω is the weight of the leaves; λ and γ are coefficients, with default values set as λ=1, γ=0.

The XGBoost algorithm can accept both continues variables and discrete variables as inputs but the output variable has to be discrete, including binary variables. In this study, the XGBoost algorithm was ran in Mathematica wolfram 13.1 software. When using the XGBoost algorithm, Z-statistic is often used for testing the significance of each independent variable, with p-value given at 95% confidence interval. It is calculated and given by the computational package after running the XGBoost algorithm.

2.2. Error Measurements

Normalised Mean Bias Error (NMBE) and the Coefficient of Variation of the Root Mean Square Error (CV(RMSE)) are used as a validation metrics to evaluate the accuracy of the XGBoost model (Eguía-Oller *et al.*, 2021).

$$NMBE \ = \ 100 * \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i^{\,(t)})}{\sum_{i=1}^{N}(y_i)} \qquad (6)$$

$$\mathrm{CV(RMSE)} \ = \ 100 * \frac{SQRT(\sum_{i=1}^{N}(y_i - \hat{y}_i)^2)}{\sum_{i=1}^{N}(y_i)} \qquad (7)$$

3. RESULTS

Figure 4 shows the change in indoor temperature and relative humidity during the period between March 2021 and October 2021 in the educational building. The temperature increases and reach the maximum level at summer.

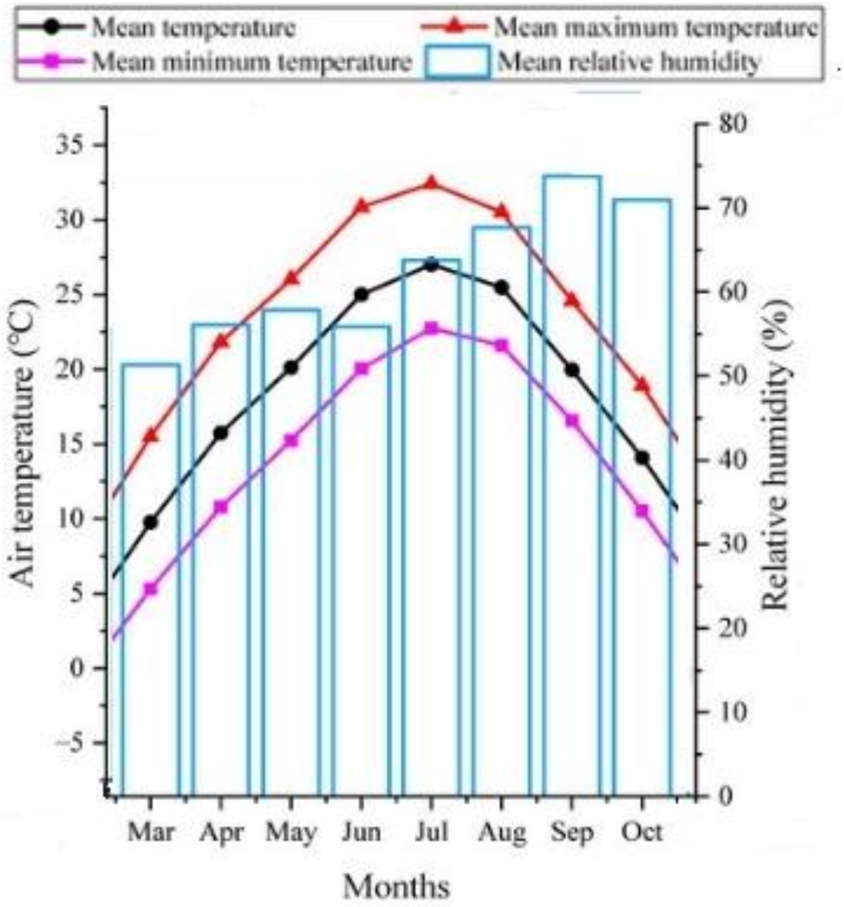

Fig.4: Temperature and relative humidity as measured inside the classroom

The XGBoost model consists of a lot of parameters that must be taken into consideration. These parameters are:

- “n_estimators”: the number of base tree models, the higher the number of iterations, the higher the value;
- “max_depth”: the maximum depth of the base tree model, with a higher value for more complex base tree models;
- ‘gamma”: the minimum loss reduction required to divide further on the leaf nodes of the tree, with higher values for more conservative models;
- “subsample”: the subsample rate of the training instances.

In Table 1 the tuned parameters and evaluation metrics from the XGBoost model are presented after construction.

Table 1: The mean NMBE, CV(RMSE), standard deviations (SDs). and computation time (in seconds), needed to train each of the models

Model	Variables	Regression trees	NMBE [%]	SD	CV(RMSE) [%]	SD	Computational time [sec]	R^2
Physics based	Air Velocity	120	-2.185	7.61	10.43	4.68	2.02	0.93
	Temperature	400	-1.09	2.81	4.57	1.59	5.21	0.91
	Relative humidity	440	0.075	4.52	5.98	2.76	7.31	0.92
Hybrid	Air velocity	120	-2.20	7.33	8.74	4.44	0.91	0.99
	Temperature	400	0.163	3.78	4.47	1.66	0.81	0.99
	Relative humidity	440	0.88	4.37	5.66	3.14	0.92	0.99

According to the results shown in the table and Figure 5, the hybrid model maintains a relative reduction in errors with a slight change in the results between the two models, in a faster processing time for arithmetic operations than the traditional model.

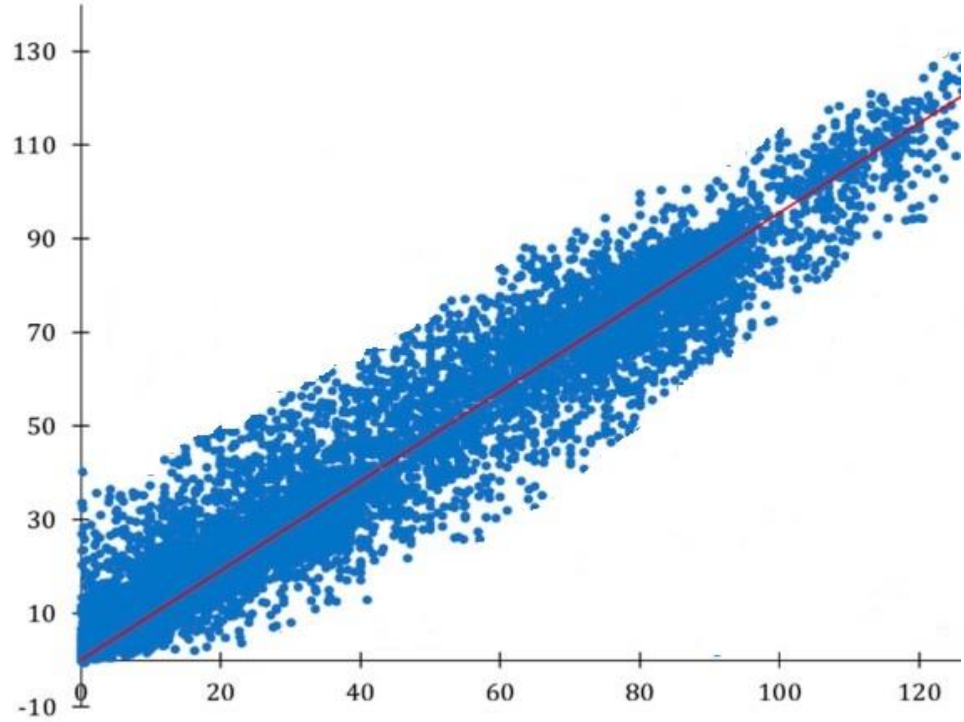

Fig.5: True value (x-axis) vs. predicted value (y-axis) plot for optimized XGBoost

4. CONCLUSIONS

An optimized model consists of two sections, an application of physics-based model for prediction of thermal comfort and a hybrid integrated model for thermal comfort real time calculation using IoT devices within an educational building using XGBoost algortithm can monitor the indoor thermal comfort variables. The results show that the model can accurately predict the indoor environmental parameters. The Error measurements in the three variables was below 5% for temperature, below 6% for relative humidity and below 9% for air flow with very low computational times. This indicate that the hybrid model could be an effective one to help the end user in decision making.

REFERENCES

- BALBIS-MOREJON, M. AND NOYA-SAMBRANO, A. (2020) "Thermal comfort evaluation in an educational building with air conditioning located in the warm tropical climate of Colombia," in *IOP Conference Series: Materials Science and Engineering*. Institute of Physics Publishing. Available at: https://doi.org/10.1088/1757-899X/844/1/012030.
- CALAMA-GONZÁLEZ, C.M., LEÓN-RODRÍGUEZ, Á.L. AND SUÁREZ, R. (2022) "Climate change mitigation: thermal comfort improvement in Mediterranean social dwellings through dynamic test cells modelling," *International Journal of Energy and Environmental Engineering* [Preprint]. Available at: https://doi.org/10.1007/s40095-022-00498-1.
- CHEN, T. AND GUESTRIN, C. (2016) "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, pp. 785–794. Available at: https://doi.org/10.1145/2939672.2939785.
- EGUÍA-OLLER, P. *et al.* (2021) "Empirical validation of a multizone building model coupled with an air flow network under complex realistic situations," *Energy and Buildings*, 249. Available at: https://doi.org/10.1016/j.enbuild.2021.111197.
- MAJEWSKI, G. *et al.* (2020) "Assessment of thermal comfort in the intelligent buildings in view of providing high quality indoor environment," *Energies*, 13(8). Available at: https://doi.org/10.3390/en13081973.
- QAVIDEL FARD, Z., ZOMORODIAN, Z.S. AND KORSAVI, S.S. (2022) "Application of machine learning in thermal comfort studies: A review of methods, performance and challenges," *Energy and Buildings*. Elsevier Ltd. Available at: https://doi.org/10.1016/j.enbuild.2021.111771.
- WANG, YONGJUN *et al.* (2021) "Ticagrelor versus Clopidogrel in CYP2C19 Loss-of-Function Carriers with Stroke or TIA ," *New England Journal of Medicine*, 385(27), pp. 2520–2530. Available at: https://doi.org/10.1056/NEJMOA2111749.
- ZANG, M., XING, Z. AND TAN, Y. (2019) "IoT-based personal thermal comfort control for livable environment," *International Journal of Distributed Sensor Networks*, 15(7). Available at: https://doi.org/10.1177/1550147719865506.