Enhancing Thermal Comfort in Buildings with Machine Learning-Based Overheating Prediction

Mithun Vijayakumar Rajalekshmi
Intelligent Technologies Research Group
University of East London
London, United Kingdom
u2428040@uel.ac.uk

Arman Hashemi
Dept. Of Architecture and Visual Arts
University of East London
London, United Kingdom
a.hashemi@uel.ac.uk

Mhd Saeed Sharif
Intelligent Technologies Research Group
University of East London
London, United Kingdom
s.sharif@uel.ac.uk

***Abstract*—The goal of this project is to improve the application of machine learning techniques in the summertime prediction of thermal comfort in residential structures (for both present and future weather situations). Using DesignBuilder's integrated simulation engine and simulated data, the research creates strong prediction models with Random Forest and XGBoost algorithms. Essential factors like building orientation, window-to-floor ratios, U-values, and operating temperatures were examined using exploratory data analysis, feature engineering, and thorough data preparation. Mean Absolute Error (MAE) and R-squared values were applied for the accurate and effective validation of the models. The findings demonstrate significant potential for early-stage decision making on building designs, for reducing risk of overheating and opening the door to more sustainable and comfortable living spaces. Future research endeavors aim to enlarge the dataset, explore different Machine learning modeling techniques, and enhance the models' capability to predict and mitigate overheating in different building kinds and climatic conditions.**

***Index Terms*— Machine Learning, Random Forest, XGBoost, DesignBuilder, EnergyPlus, Building Orientation, Window to floor ratio, Thermal Comfort, U-values, Overheating Prediction**

I. Introduction

The rapid urbanization and climate changes are having significant impact on the modern urban landscape. As cities grow and their climates become more severe, maintaining healthy living and working conditions in buildings within urban settings is becoming increasingly difficult. Conventional cooling methods such as air conditioning and passive cooling techniques often fail to address the complex interrelationship of building materials, architectural design and environmental factors [1 - 3]. In addition to negatively affecting occupant comfort, this reduction dramatically raises energy use and carbon emissions [4].

With the help of data driven algorithms, Machine Learning (ML) provides a revolutionary way to more accurately predict and control building environments [5]. ML models can process large amounts of data to identify the correlation, patterns and make predictions from it, which are very useful for optimizing building operations and thereby ensuring thermal comfort inside the buildings. This research uses machine learning to address the problem of buildings becoming warmer as global temperature rises [6]. The main objective is to develop a predictive model that integrates dynamic simulation standards (CIBSE TM52 and CIBSE TM59) with real-time data analysis [7], [8].

Latest innovations in ML have shown that thermal comfort modeling holds great promise. High accuracy indoor environmental condition predictions have been achieved through the use of ML techniques like Random Forest, XGBoost [5, 6]. Moreover, it appears that integrating adaptive models that respond to real-time data is necessary for dynamically controlling thermal environments [9, 10]. By combining these advanced ML techniques, this research aims to improve the predicted accuracy of thermal comfort model by focusing on important variables including building orientation, material properties and operative temperature.

II. Literature Review

A complicated mental state known as "thermal comfort" is expressed as happiness with the thermal environment and is impacted by a number of variables, such as garment insulation, air speed, humidity, air temperature, radiant temperature and metabolic rate [1]. The impact of thermal comfort on energy usage and occupant well-being in building design has been well studied [2], [3]. Many recommendations and standards, including ASHRAE Standard 55, ISO 7730, and CIBSE TM 52/53, have been created over time to define and quantify thermal comfort. These standards and guidelines include extensive requirements for acceptable thermal settings [1], [2], [7], and [8]. Conventional methods of preserving thermal comfort in buildings frequently depend on HAVC (Heating, Ventilation, and Air Conditioning) systems having setpoints. However, the changing character of thermal comfort which is impacted by occupant behavior as well as environmental factors is creating an increasing threat to these techniques [3]. Recent developments in building technologies and the expansion of data availability have resulted in more advanced techniques for predicting controlling thermal comfort using machine learning [4]. A data-driven method for modelling and predicting thermal comfort is provided by machine learning, allowing for more flexible and customized control over interior conditions. In this domain, several kinds of machine learning techniques have been examined; each has advantages and disadvantages. The use of deep learning models to increase the precision of thermal comfort prediction was shown by

Zhang et al. [4]. The necessity for machine learning models that can take into consideration unique thermal preferences and spatial variability was highlighted by Lala et al. [5]. Non-intrusive techniques were investigated by Ghahramani et al. [6] for excellent prediction accuracy in personal comfort. To maxmise thermal comfort and indoor air quality, Ma et al. talked about combining machine learning techniques with empirical and deterministic models [11]. The integration of adaptive models that respond to real-time data is seen as crucial for managing thermal environments dynamically. Jiang et al. developed a personalized HVAC control system using machine learning [9]. Their system learns from occupant feedback and environmental data to adjust HVAC settings dynamically, ensuring optimal thermal comfort while minimizing energy consumption. Similarly, Peng et al. applied machine learning to HVAC system control, demonstrating significant energy savings and improved occupant comfort [10]. Research by Guenter et al. as well as Huchuk et al. emphasise the necessity of occupancy prediction and intelligent building management for energy savings [12], [13]. Huchuk et al. focused on predictive control strategies that anticipate changes in occupancy and environmental conditions, allowing HVAC systems to adjust proactively [12]. To optimize building system functioning Guenter et al. [13] utilizes occupancy patterns and environmental data and able to achieve significant energy savings. The imbalance between HVAC system energy consumption and occupant happiness, especially in over-cooled conditions is another major concern. Chaudhuri et al. [14] developed a model for predicting thermal comfort which are based on gender-specific for solving the problem. Also, their research shows that different population groups have different requirements for thermal comfort, which should be consider while developing and using HVAC systems [15]. Studies by GAO et al. [16], [17] made contributions to the field by implementing cutting-edge machine learning methods like deep deterministic policy gradients and transfer learning and improving the ability of the models to predict thermal comfort in various building and climatic conditions. Zhao et al. checked how AI-based predictive control can help out in smart buildings [18]. Their research showed us that even when environmental conditions changes, machine learning can effectively optimise HVAC system efficiency and keep thermal comfort. Song et al. [19] examined the impact of psychological features on the thermal comfort who shown the importance for models that consider both the psychological and physical aspects of comfort. By using artificial intelligence (AI) to evaluate thermal comfort in learning environments, López-Pérez et al. showed how machine learning can enhance both comfort and learning results [20]. The integration of physiological and psychological aspects in HVAC control was studied by Turhan et al. [21] who highlight the significance of a comprehensive strategy for thermal comfort management. To sum up the combination of machine learning and building environmental data presents an ideal way for efficient and sustainable management of indoor atmosphere [17], [18]. This study highlights how important it is to create models that can be interpreted, adjusted to different building contexts and predicted [21], [22]. While earlier research has shown that machine learning models can be used to predict thermal comfort, many studies focused on just specific building types or circumstances. Studies such as those conducted by Zhang et al. (2019) [4] examined the use of deep learning to increase the prediction accuracy of thermal comfort prediction, but they did not investigate the prediction of overheating risks in the context of future weather patterns. Similarly, Ghahramani et al. (2020) [6] examined non-intrusive methods for predicting personal comfort but did not examine the influence of window-to-floor ratios and building orientation. This study builds upon such previous research by offering multiple machine learning models to predict overheating risks specifically in residential building during supper conditions. In addition, the incorporation of future climatic scenarios into this research enhances its value for building managers and urban planners by offering a useful tool (GUI) for real-time projections.

III. Methodology

The steps to create and check machine learning model that predict thermal comfort in residential buildings are described in the methodology section. The four main steps included in this part are Data collection, data preprocessing, model building and training, and. A flow chart summarises these steps are shown in Fig.1.

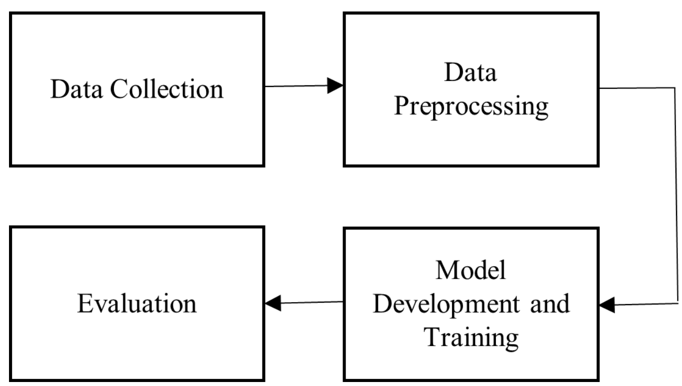

Fig. 1: Steps of Methodology

A. Data Collection

We used DesignBuilder software along with EnergyPlus simulation engine to create thermal comfort data for the current year and future (2050) weather. This setup helps generate detailed environmental data for a residential flat that has materials and design features from mid-century. The simulations are mainly focused on summer conditions. They kept track of hourly operating temperatures and the outside temperature from May all the way through September. To make sure the data really showed typical summer weather patterns, we used CIBSE Design Summer Year (DSY) meteorological files.

a) DesignBuilder Software: DesignBuilder gives you really useful information about how a building performs in the environment. It looks at things like carbon emissions, energy use and thermal comfort. How does it do all this? Well, it uses EnergyPlus, which is a strong simulation tool. This tool helps with detailed thermal modelling. It even works with time steps that are shorter than hour. This program follows the TM59 guidelines too. These guidelines are a special standard created by CIBSE to help predict if a residential building might get overheat.

b) Simulation Setup and Evaluation: For this study, we checked the risk of overheating using the CIBSE TM59 rules during summer conditions. It's all about making sure building stay comfy and cool. The CIBSE DSY weather file was utilised for simulations from May to September. The TM59 methodology requires that results for living spaces should not exceed 3% of occupied hours with temperatures above comfort thresholds. For bedrooms, the operative temperature should not rise above 26°C for more than 1% of the hours from 10 p.m. to 7 a.m. to ensure comfort during sleeping hours. Internal gains profiles, including occupancy, equipment, heat gain, and heating, followed TM59 standards, with natural ventilation applied when room temperatures exceeded 22°C during occupied hours.

c) Description of the Case Study and Building Simulation Modelling: The case study focused on a top-floor flat in a seven-story building block in the London Borough of Newham, southeast England. Constructed between 1950 and 1966, the building includes 100 flats with one- and two-bedroom units. Typical 1960s construction materials' specifications were used for the simulation model. The U-values were set to 2.3 W/m²K for the roof, 1.2 W/m²K for internal floors, 1.5 W/m²K for external walls, and 2.8 W/m²K for glazing. The flat has two exposed external surfaces, with the living room facing south and a shaded terrace, and the kitchen facing north with a shaded open corridor. The bedroom has south and east exposures with a large unshaded window area on the south side.

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

Fig. 2: (a,b) a picture of the case study house. (c) the 3D model of the case study house. (d) flat floor plan from Designbuilder.

The primary variables collected included:

1) *Temporal Features*: Hour of the day, day of the week, and month.
2) *Building Physical Features*: U-values for the roof, walls, floor, and windows; window-to-floor ratio; building orientation.
3) *Environmental Interaction Features*: Window openable percentage, external temperature.

These features were chosen based on their relevance to thermal comfort and their availability from the simulation tools. In this study, external temperature serves as the prediction variable, while the remaining variables are considered input features.

B. Data Preprocessing

To prepare the gathered data for model development, data preprocessing is essential. The steps included in the preprocessing are follows:

1) *Data Cleaning*: Data cleaning is really about getting rid of any wrong or messy information like numbers that just don't fit or places where info is missing. It helps make sure that the data is correct and reliable which is super important.
2) *Feature Engineering*: In this step we will create new features to make the models prediction capability even better. For example, adding new features like the percentage of the windows can be opened etc.
3) *Normalization*: In this step we scale the features to a standard range which will ensure that all features equally contribute to the models learning process.

C. Model Development and Training

In this study, we used four machine learning models for predicting thermal comfort which are Random Forest Regressor, XGBoost Regressor, Linear Regression, and Decision Tree Regressor. The main reason for choosing these models are due to their capacity to manage complex dataset with different variety of features and their dependability. During the training time Random Forest Regressor will create several decision trees and take the average forecast of each tree.. Since it has the capability to handle high-dimensional data, it is very resistant to overfitting and provide insights into the relevance of the features. Also the complexity of the model is less when compared to the other black-box models which will help us to understand how the features are learned for the accurate prediction.

Similarly, XGBoost Regressor is popularfor its effectiveness and super performance, particularly when dealing with complex and large-scale data. The main technique involved in this model is Gradient boosting in which the model is a sequential model, it will learn by fixing the errors from previous ones and goes. Also XGBoost can be used for wide range of machine learning applications due to its optimisation capability and scalability. Another main feature of XGBoost model is it has the ability to manage missing data and it can avoid overfitting by using regularisation techniques. As a baseline model comparison, Linear Regression model was selected due to its simplicity and interpretability. Linear Regression shows better performance while used with simpler dataset because the model assumes a linear connection between the features. At last, Decision Tree Regressor was used, because of its capability to handle continuous and categorical features. More like to Random Forest, Decision Tree also have the flow-chart like structure to make decisions and which enables a clear understanding of how the model will predict. Due to the efficiency and the ability to perform well in different domains from healthcare analytics to financial, these models are kept with high respect by the machine learning community. Apart from the precise prediction, a clear understanding of the data patterns is also expected, so that it can contribute for a finer decision-making processes.

To identify the best parameters for each model, grid search with cross-validation was used for hyperparameter tuning. min_samples_leaf, max_depth, min_estimators were the main parameters used for tuning Random Forest model, and min_samples_split. Learning_rate, max_depth, n_estimators, and subsample were the main tuning parameters for the XGBoost model. min samples split and max depth were used for tunning Decision Tree Regressor. Decision Tree Regressor was tuned using min_samples_split and max_depth. For the training process, the dataset was divided into training (80%) and testing (20%) sets. The performance of the model is validated using a 10-fold cross-validation strategy by minimising mean squared error(MSE). Finally, the trained model is tested on the testing set.

D. Evaluation

Accuracy and performance of the models were evaluated using a variety of metrics. Mean Absolute Error metrics was used to measure the average magnitude of the prediction mistakes, which helps in analyzing the accuracy of the forecasts. The metrics Mean Squared Error(MSE) and Root Mean Squared Error(RMSE) were chosen to offer a thorough evaluation of the predicted accuracy and resilience of the model. In MSE by squaring the errors we get a metric that penalize greater errors more than smaller ones, The square root of MSE, known as Root Mean Squared Error(RMSE), shows the error using the same units as the original data.. The percentage of the dependent variables variance that can be predicted from the independent variables is represented by R-squared (R^2) [4]. In particular, the models were compared to determine their relative performance in terms of predictability (Decision Tree Regressor), simplicity (Linear Regression) and predictive strength (Random Forest and XGBoost). Highest model performance is achieved when the model has low MAE, MSE, and RMSE values and high R-squared values. Cross-validation was performed using various random split of the dataset. This method is used to enhance the models robustness. As a results, it provides more accurate prediction and also help to reduce the overfitting.

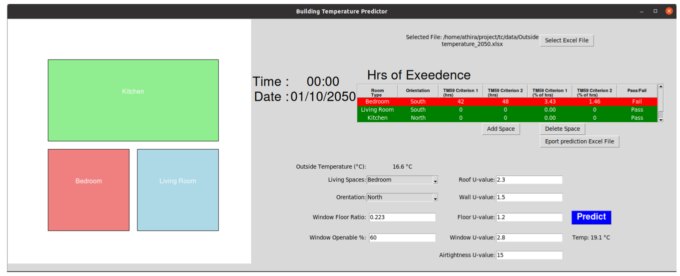

Fig. 3: User Interface for Real-Time Testing and Validation of Predictive Models.

E. UI Development and Real-Time Testing

In this research, we have developed a simple Graphical User Interface (GUI) to ensure the developed models prediction have theoretical soundness ad practical applicability. With the use of GUI we can test and validate all the four models in real time, which will improves user engagement and easy to use for building panners and the experts. The home screen screenshot of the GUI is shown in Fig. 3.

The GUI offers several key features:

- *Data Integration*: Users can input the external temperature data through the GUI as Excel files, which helps utilize both the historical and real time weather data for predictive modelling.
- *Building Parameters*: The GUI accepts building parameters input manually from users, such as proportion of openable windows, window-to-floor ratios, and U-values for various building components. Because of its adaptability, users can model diverse situations and comprehend how various setups impact interior thermal comfort.
- *Dynamic Living Space addition*: This attribute allows users to allocate new living spaces dynamically with certain parameters. It clears the way to inspect thermal comfort across a range of architectural design and orientations, and also helps to maximise both occupant comfort and energy efficiency.
- *Real-Time Validation and Testing*: GUI allows users to rapidly obtain forecasts of the operating temperatures for various living areas using the input parameters. Due to this real-time testing and validation of the prediction models can be achieved immediately which is essential for determining how environmental and architectural factors affect interior temperatures
- *Measure of Overheating*: The amount of time each living area spends in overheated conditions is measured using an integrated overheating assessment tool. It determines the proportion of time that each space surpasses these requirements and outputs a Pass/Fail signal depending on predetermined cutoff points, like the CIBSE TM59 overheating requirements. A Fail status is displayed by the GUI if the living space overheats beyond the permitted limits.

IV. Results and discussion

The results pf the developed machine learning models for predicting thermal comfort are shown in this part, along with a detailed analysis of the results. The main aim is to assess the performance of the models such as Random Forest Regressor, XGBoost Regressor, Linear Regression and Decision Tree Regressor using the metrics-Mean Absolute Error(MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-Squared (R2)-described in the approach.

Through the GUI the predictive analytics are converted in to actionable insights which helps in improving the accessibility and utility of the machine learning models.. The thermal comfort and energy efficiency in residential buildings is control in a proactive manner by filling the gap between the results of cutting-edge research and practical operational requirements. Also this work make sure that the models created for this research is workable and immediately applicable to the real-world building . Building managers and designers may make well-informed decisions faster using the GUI, which optimises building designs to improve thermal comfort while meeting energy efficient requirements.

A. Model Performance

The performance of all the four models is evaluated using the test dataset.. The assessment metrics for the Random Forest Regressor, XGBoost Regressor, Linear Regression and decision Tree Regressor models are summarised in Table 1. The table shows the values of MAE,MSE, and RMSE for the Random Forest Regressor are lower implying that it has a better predictive capacity and can predict temperatures more accurately, Thus the performance of the Random Forest Regressor is much better than that of the XGBoost Regressor, Linear Regression, and Decision Tree Regressor. Furthermore higher R2 values indicate percentage of variation in the data is greater in Random Forest Regressor compared to XGBoost Regressor.

B. Analysis of Model Performance

a) Mean Absolute Error (MAE): The Random Forest Regressors MAE value (0.12) is significantly low when compared to the XGBoost Regressor (0.35), Linear Regression (0.85), and Decision Tree Regressor (0.14), indicating that the Random Forest Regressor produces predictions with less average error.

b) Mean Squared Error (MSE) and Root Mean Squared Error (RMSE): Compared to XGBoost Regressor, which has MSE and RMSE values of 0.24 and 0.49 respectively, the Random Forest Regressors MSE values are lower at 0.05 and 0.21 respectively. With a MSE of 1.30 and RMSE of 1.14 the Linear Regression model shows the highest error rates, making it less suitable for this application. With a MSE of 0.08 and RMSE of 0.29 the Decision Tree Regressor shows a respectable performance but does not surpass the Random Forest Regressor.

TABLE I: PERFORMANCE METRICS

Model	MAE	MSE	RMSE	R^2
Random Forest Regressor	0.12	0.05	0.21	0.99
XGBoost Regressor	0.35	0.24	0.49	0.92
Linear Regression	0.85	1.30	1.14	0.58
Decision Tree Regressor	0.14	0.08	0.29	0.97

c) R-Squared (R^2): Random Forest Regressor With an R^2 of 0.99 captures 99% of the variation in the dependent variable. In contrast, the XGBoost Regressor captures 92% of the variance with an R^2 value of 0.92, while the Linear Regression only captures for 58% ($R^2 = 0.58$), showing a poor fit. With an R^2 value of 0.97 the Decision Tree Regressor performs well, but still behind the Random Forest regressor. This highlights the Random Forest Regressor's better ability to understand the relationships in the data.

C. Thermal Comfort Assessment

We compared the AI predictions with the simulated results using PMV index for thermal comfort under various building standards in order to give a deeper understanding of the model's performance.

TABLE II: THERMAL COMFORT IDENTIFIED FROM SIMULATION AND AI PREDICTIONS

Room Type	Base Case		PartL		Passivhaus	
	(Sim)	(AI)	(Sim)	(AI)	(Sim)	(AI)
Bedroom	Fail for 141 h	Fail for 90 h	Fail for 94.5 h	Fail for 43 h	Fail for 90 h	Fail for 43 h
Living Room	Pass	Pass	Pass	Pass	Pass	Pass
Kitchen	Pass	Pass	Pass	Pass	Pass	Pass

- *Living Room and Kitchen*: As indicated by the simulated outcomes and AI predictions the living room and kitchen maintain an appropriate level of thermal comfort in all the conditions. This determines the predictive accuracy and dependability for thermal comforts of AI models. From the findings between the AI predictions and simulation, it is clear that the model captures the critical parameters impacting comfort levels. This helps validate the AI model's effectiveness in accurately predicting thermal comfort in a range of settings and show the model's durability.
- *Bedroom*: 90 hours of discomfort has been predicted by AI, where the simulated results show that for the base case 141 hours of discomfort has been identified. In the case of PartL with cavity insulation 43 hours of discomfort has been predicted by AI and simulated results shows 100 hours of discomfort. Similarly, for the Passivhaus with EWI the AI predicts 43 hours of discomfort and the simulated results shows 84.6 hours of

discomfort. This indicate that the AI model offers a more optimistic assessment of thermal comfort by consistently predicting few hours of discomfort. However, this shows the AI's capacity to more accurately understand complex patterns and adjust quickly to changing circumstances than static models.

D. Practical Implications

The results show that for predicting thermal comfort in residential buildings Random Forest Regressor is the best choice compared to the other 3 models. In addition to its high accuracy, it is likely to be a dependable tool for real time thermal comfort prediction, supporting building managers in their decision-making about occupant comfort enhancement and maintenance in naturally ventilated spaces.

- *Adaptability*: Due to its flexibility to manage a variety of building kinds and climate conditions, it can be used in wide range of applications. Furthermore its adaptability helps to maintain ideal thermal comfort in a variety of situations.
- *Occupant Comfort*: Occupant comfort is very important in residential and commercial building, this is because the increased thermal comfort boosts productivity and occupant happiness. Accurate prediction allows Precise prediction enables immediate adjustment to natural ventilation techniques and which provides constant comfort levels.
- *Energy Efficiency*: Energy saving can be achieved by optimising natural ventilation systems with the use of precise predictions. Resulting more advantages like cheaper operating costs , less dependence on mechanical cooling systems and for environment.

E. Discussion

The ensemble learning approach of the Random Forest Regressor which integrates many decision trees to improve prediction accuracy is responsible for its good performance. Also this approach helps to reduce the overfitting while handling complex data with effectiveness. With its better accuracy in predicting thermal comfort, the Random Forest Regressor model shows the potential for real word use cases in building management and design.

A Comparison between the simulated results and AI predictions offers valuable insights into thermal comfort in residential buildings. The AI model predicts fewer hours of discomfort in the bedroom compared to simulation. This suggests a significant understanding of thermal dynamics due to the models ability to capture difference in ambient factors and behaviour elements that static simulations might miss. This is crucial because thermal comfort has a significant impact on the well-being and quality of sleep of occupants in the bedroom. In contrast , living room and kitchen shows more stable place for keeping thermal comfort due to the consistency between AI and simulated results. This stability is likely the result of well-designed architectural features such as insulation types and ventilation details that the AI model has correctly learned. With an impressive R^2 value of 0.99, the model shows its reliability in predicting thermal comfort across different scenarios and supports preventive measures to enhance occupant comfort.

This project has achieved several important benchmarks, making significant advances in thermal comfort prediction by using sophisticated machine learning models such as Random Forest Regressor. The models high accuracy allows for a real time thermal comfort adjustments, providing a stronger foundation for comfort maintenance than in traditional methods. Artificial cooling can be minimised and proper airflow can be maintained by precise prediction, which in turns reduce the operation costs, energy preservation and benefits the environment. Additionally, occupants comfort can be enhanced and maintained by making proper adjustments by the building managers with the help of precise prediction. This is really important in residential environment, where quality of life is impacted by comfort. The results suggest that Ai can play a crucial role in preserving constant comfort levels and enhancing occupant satisfaction. Our model is useful for wide range of applications and can be applied in different scenarios only because of the model's adaptability.

The dataset used in the study is limited to a single flat in London, which restricts the ability to apply the findings in different kinds of buildings and climate zones, Future research should aim to expand the dataset by adding different building types and different climates etc. Doing so would increase the model's adaptability and effectiveness in a variety of contexts. Furthermore, adding more features like occupant behaviour, building specific features like insulation types, and real-time meteorological data might increase the model's accuracy and wider applicability.

Although the outcomes are promising, there are few areas that could benefit further research and development:

- *Feature Engineering*: Incorporating additional features related to occupant behaviour and external environmental factors could further enhance model accuracy.
- *Real-time Implementation*: Developing a real-time implementation of the model could provide immediate feedback and adjustments to HVAC systems, further optimizing energy use and comfort.

In conclusion, research has shown that the Random Forest Regressor has great potential for predicting residential building thermal comfort. Because of its great precision and reliability it will be a valuable tool for enhancing occupant comfort. Future work will focus on expanding the dataset, adding new features and developing real-time application in order to increase the models accuracy.

V. CONCLUSION

This study focused on developing and evaluation machine learning models such as Random Forest Regressor, XGBoost Regressor, Decision Tree Regressor, and Linear Regression for predicting thermal comfort in the residential buildings. . The Random Forest Regressor shows the best performance With a mean absolute error (MAE) of 0.12, mean square error

(MSE) of 0.005, root mean square error (RMSE) of o.21, and an R-squared value of 0.99. The decision Tree Regressor also performs well similar to Random forest with MAE of 0.14 and R^2 of 0.97 making it a strong alternative to Random Forest. However, the less effective model was Linear Regression.

The high accuracy of the Random Forest Regressor's suggest it has a wide range of practical applications such as adjusting the HVAC settings to improve occupant comfort and save energy. This is really important in terms of rising global temperatures and rising energy costs. Future research should be focused on adding more data, adding more features like, occupant behaviour features, temperature features, building features etc will helps to improve models prediction capability and reliability.. Additionally developing a real time applications and improving the models interpretability will be crucial for its continued success.

The data used in this research is limited to a single residential flat in the London. While the result are promising with in this specific context, the research does not fully address how well these findings can be applied to different building types and climate zones. Future studies should aimed to expand the dataset by adding more features like different building types, different climatic conditions, environmental features, occupant features etc. This extension would ensure that the findings may be implemented globally by enabling more comprehensive model validation over a various range of scenarios. In general, the Random Forest Regressor achieve a lot of potential for enhancing energy efficiency and occupant comfort in residential structures, contributing to the creation of smarter, healthier and more sustainable living environments.

Acknowledgment

This research got support from the medical Research Council (MRC) and UK Research and Innovation (UKRI) under grant number MR/Y503186/1. The authors would like to thank iOpt Limited, Hyde Housing Association, and Newham Council for their support and giving access to the residences involved in the research.

References

[1] ASHRAE Standard 55. (2017). Thermal Environmental Conditions for Human Occupancy.

[2] Fanger, P. O. (1970). Thermal Comfort: Analysis and Applications in Environmental Engineering.

[3] de Dear, R., & Brager, G. S. (1998). Developing an adaptive model of thermal comfort and preference. ASHRAE Transactions, 104(1), 145–167.

[4] Zhang, W., Hu, W., & Wen, Y. (2019). Thermal comfort modeling for smart buildings: A fine-grained deep learning approach. IEEE Internet of Things Journal, 6(2).

[5] Lala, B., Kala, S., Manas, M., Rastogi, A., Dahiya, K., Yamaguchi, H., & Hagishima, A. (2022). Building matters: Spatial variability in machine learning based thermal comfort prediction in winters. Proc. 2022 IEEE Int. Conf. Smart Computing (SMARTCOMP), 342–348. DOI: 10.1109/SMARTCOMP55677.2022.00078.

[6] Ghahramani, A., Galicia, P., Lehrer, D., Varghese, Z., Wang, Z., & Pandit, Y. (2020). Artificial intelligence for efficient thermal comfort systems: Requirements, current applications and future directions. Frontiers in Built Environment, 6:49. DOI: 10.3389/fbuil.2020.00049.

[7] CIBSE. (2013). The Limits of Thermal Comfort: Avoiding Overheating in European Buildings. TM52

[8] CIBSE. (2017). Design Methodology for the Assessment of Overheating Risk in Homes. TM59

[9] Jiang, L., & Yao, R. (2016). Modelling personal thermal sensations using C-support vector classification (C-SVC) algorithm. Building and Environment, 99, 98–106. https://doi.org/10.1016/j.buildenv.2016.01.022.

[10] Peng, Y., Nagy, Z., & Schlüter, A. (2019). Temperature-preference learning with neural networks for occupant-centric building indoor climate controls. Building and Environment, 154, 296–308. https://doi.org/10.1016/j.buildenv.2019.01.036.

[11] Ma, N., Aviv, D., Guo, H., & Braham, W. W. (2021). Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality. Renewable and Sustainable Energy Reviews, 135, 110436. https://doi.org/10.1016/j.rser.2020.110436.

[12] Huchuk, B., Sanner, S., & O'Brien, W. (2019). Predictive control of indoor environments with occupancy forecasting. Building and Environment, 144, 594–604.

[13] Guenter, J., & Sawodny, O. (2019). Feature selection and Gaussian process regression for personalized thermal comfort prediction. Building and Environment, 148, 448–458. https://doi.org/10.1016/j.buildenv.2018.11.019.

[14] Chaudhuri, T., Zhai, D., Soh, Y. C., Li, H., & Xie, L. (2018). Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology. Energy & Buildings, 166, 391–406. https://doi.org/10.1016/j.enbuild.2018.02.035.

[15] Chaudhuri, T., Soh, Y. C., Li, H., & Xie, L. (2019). A feedforward neural network based indoor-climate control framework for thermal comfort and energy saving in buildings. Applied Energy, 248, 44–53. https://doi.org/10.1016/j.apenergy.2019.04.065.

[16] Gao, G. Y., Li, J., & Wen, Y. G. (2019). Energy-efficient thermal comfort control in smart buildings via deep reinforcement learning. arXiv:1901.04693v1 [cs.SY].

[17] Gao, N., Shao, W., Rahaman, M. S., Zhai, J., David, K., & Salim, F. D. (2020). Transfer learning for thermal comfort prediction in multiple cities. In Proc. 6th ACM/IEEE Conf. Internet of Things Design and Implementation (IoTDI '21).

[18] Zhao, X., Li, H., & Yang, J. (2020). AI-based predictive control in smart buildings. Energy and Buildings, 226, 110354.

[19] Song, G., Ai, Z., Zhang, G., Peng, Y., Wang, W., & Yan, Y. (2022). Using machine learning algorithms to multidimensional analysis of subjective thermal comfort in a library. Building and Environment, 212, 108790. https://doi.org/10.1016/j.buildenv.2022.108790.

[20] López-Pérez, L. A., & Flores-Prieto, J. J. (2023). Adaptive thermal comfort approach to save energy in tropical climate educational building by artificial intelligence. Energy, 263, 125706. https://doi.org/10.1016/j.energy.2022.125706.

[21] Turhan, C., Ozbey, M. F., Lotfi, B., & Akkurt, G. G. (2023). Integration of psychological parameters into a thermal sensation prediction model for intelligent control of the HVAC systems. Energy & Buildings, 296, 113404. https://doi.org/10.1016/j.enbuild.2023.113404.

[22] Wang, Z., Liu, J., Zhang, Y., Yuan, H., Zhang, R., & Srinivasan, R. S. (2021). Practical issues in implementing machine-learning models for building energy efficiency: Moving beyond obstacles. Renewable and Sustainable Energy Reviews, 143, 110929. https://doi.org/10.1016/j.rser.2021.110929.