UDC 004.891.2
IRSTI 28.23.29

https://doi.org/10.55452/1998-6688-2025-22-3-59-74

[1]*Assymkhan N.,
Master's student, ORCID ID: 0009-0002-0398-5645,
*e-mail: anb.asymhan@gmail.com
[1]Momynkul N.,
Master's student, ORCID ID: 0009-0009-0555-3636,
e-mail: n_momynkul@kbtu.kz
[1]Kartbayev A.,
PhD, ORCID ID: 0000-0003-0592-5865,
e-mail: a.kartbayev@kbtu.kz

[1]Kazakh-British Technical University, Almaty, Kazakhstan

OPTIMIZING INDOOR THERMAL COMFORT PREDICTION USING MACHINE LEARNING MODELS

Abstract

Predicting thermal comfort in indoor environments is important for improving residents' well-being, productivity, and energy efficiency. This study explores machine learning approaches, specifically Support Vector Machines (SVM) and Random Forest (RF), to improve thermal comfort prediction. Traditional methods rely on subjective assessments, whereas our approach leverages data-driven models trained on large thermal comfort datasets. The dataset underwent rigorous preprocessing, with 80% used for training and 20% for testing. The integration of the Internet of Things (IoT) further enhances predictive accuracy by enabling adaptive control in smart building systems. A comparative analysis of SVM and RF reveals that while both models effectively capture the complex interactions between environmental parameters and resident comfort, RF demonstrates greater stability and higher accuracy in most scenarios. The paper proposes potential strategies for integrating additional predictive features to further enhance model accuracy, demonstrating the advancement of machine learning in optimizing indoor comfort.

Keywords: heating systems, energy management, thermal comfort, support vector machine, random forest, machine learning

Introduction

Optimizing indoor environments for human habitation is essential for ensuring comfort, health, and productivity. Thermal comfort, a key aspect of indoor environmental quality, is significantly influenced by climate change, which has led to more frequent and severe extreme weather events. The ability to predict and manage thermal conditions effectively has become increasingly important as it directly impacts human well-being. Poor indoor thermal conditions, whether excessively hot or cold, can cause discomfort, fatigue, and health issues, ultimately reducing productivity in workplaces, educational institutions, and homes.

Beyond individual well-being, the economic impact of inadequate thermal comfort is substantial. When indoor conditions are not optimized, occupants rely more on heating, ventilation, and air conditioning (HVAC) systems to maintain comfort, leading to increased energy consumption and higher utility costs. This, in turn, contributes to environmental degradation due to higher carbon emissions. Consequently, there is an urgent need for advanced predictive models capable of accurately forecasting occupants' thermal comfort preferences in diverse environmental conditions and architectural settings. These models must integrate multiple factors, including ambient temperature, humidity, clothing insulation, metabolic rate, and personal preferences, to provide precise thermal comfort assessments.

Developing reliable predictive models requires collaboration across multiple disciplines, including architecture, engineering, psychology, and data science. By combining insights from environmental science, human physiology, and behavioral psychology, researchers can create more effective models. Advances in sensor technology, data analytics, and machine learning have made it possible to generate real-time insights into thermal comfort, allowing building managers and occupants to adjust indoor environments dynamically. This not only improves occupant well-being but also contributes to energy efficiency and sustainability.

To understand the real-world impact of thermal comfort, consider a simple scenario. Suppose a student is studying in a closed room during a hot summer day. The student shuts the door to minimize noise and closes the window to block out the heat. However, this leads to a buildup of carbon dioxide, reducing oxygen levels and increasing the temperature inside the room. As a result, the student experiences discomfort, distraction, and fatigue. The simple act of opening the door to improve ventilation can significantly enhance comfort. This example highlights the importance of managing thermal comfort effectively to maintain productivity and overall well-being.

Predicting thermal comfort involves analyzing multiple factors, including air temperature, humidity, air velocity, and clothing insulation. Traditional methods rely on human comfort models, such as the Predicted Mean Vote (PMV), but these approaches can be subjective and time-consuming. Recent advancements in machine learning have provided new opportunities to develop more accurate predictive models. The goal of this study investigates the use of Support Vector Machines (SVM) and Random Forest (RF) algorithms for thermal comfort prediction, comparing their performance across various conditions.

By following this structured approach, we aim to validate these goals and assess the effectiveness of SVM and RF in predicting thermal comfort. These findings could offer valuable insights for building designers and facility managers to optimize indoor environments. Both SVM and RF are supervised learning algorithms that can be trained on datasets containing thermal comfort parameters alongside human feedback, allowing them to predict thermal comfort levels in new conditions.

The Internet of Things (IoT) is transforming building management systems (BMS), with projections indicating that the number of connected devices will reach 125 billion by 2030. Despite these advancements, current BMS solutions remain limited in flexibility, particularly in feedback control mechanisms. To address these limitations, researchers have explored adaptive control algorithms and modular architecture. One proposed solution, the "Semantically-Enhanced IoT-enabled Intelligent Control System" (SEMIoTICS), leverages redundancy in control system capabilities and dynamically adjust configurations based on quality-of-service criteria [1]. Additionally, Model Predictive Control (MPC) has gained popularity in optimizing HVAC systems for energy efficiency and occupant comfort. However, the computational complexity of nonlinear MPC models has driven researchers to investigate linear controllers using Jacobian linearization. A bilinear model for nonlinear MPC was proposed to minimize energy costs while maintaining comfort, but its high computational demands present challenges for real-time implementation [2].

To further optimize HVAC systems, reinforcement learning (RL)-based approaches have been developed. One study implemented Deep Deterministic Policy Gradients (DDPG) within the Transactive Energy Simulation Platform (TESP) to achieve intelligent and granular HVAC control. The approach optimizes a cost function that balances electricity expenses with occupant dissatisfaction, incorporating a market price prediction model using Artificial Neural Networks (ANN) and a DDPG-based RL control algorithm [3]. Another study presented a simulation model integrating both high- and low-level controllers for a vehicle's air conditioning system, focusing on occupant thermal comfort while ensuring system efficiency. The study also introduced an Eco-Cooling Strategy using MPC to optimize cooling performance while minimizing energy consumption [4].

Fuzzy logic-based models have also been explored for variable-speed air conditioning load control, allowing for energy-efficient HVAC operation while improving thermal comfort. These controllers, implemented in microcontrollers, VLSI chips, and EDA tools, precisely regulate temperature and humidity levels. By integrating fuzzy logic with other control methods, system performance and energy efficiency can be significantly improved [5]. Another study examined classical HVAC control strategies, such as PID controllers, and more advanced approaches like MPC. The research introduced the LAMDA controller, designed to enhance real-time responsiveness and dynamically adjust parameters based on contextual information [6].

Increasing energy consumption in commercial buildings, particularly HVAC systems, has driven research into energy optimization. Although HVAC technologies have improved Demand Response (DR) programs, challenges remain in implementing model predictive control strategies. Machine learning techniques such as Reinforcement Learning and Supervised Learning have been investigated for their potential in improving HVAC efficiency [7]. Research on DR in Building Energy Management (BEM) has primarily focused on optimizing HVAC operations through various methods, including dynamic demand response controllers, mixed-integer nonlinear optimization models, and occupancy-based controllers. Additional strategies, such as event-based control, mutual information frameworks, and MPC, have also been explored [8]. A recent study proposed a three-layered model for optimizing energy consumption in smart homes, incorporating data collection, prediction, and optimization stages. The model uses an Alpha Beta filter for noise reduction, DELM for dynamic parameter prediction, and fuzzy controllers for refined control decisions, ultimately improving both energy efficiency and occupant comfort [9].

In the field of thermal comfort assessment, a novel model has been introduced that excludes demographic factors such as gender and age. Instead, it considers six key thermal variables: air temperature, mean radiant temperature, relative humidity, air speed, clothing insulation, and metabolic rate. This model was developed using supervised machine learning and applied in a commercial building setting [10]. Another study conducted in Bilbao, Spain, utilized the KUBIK energy efficiency research facility to analyze human thermal perception in response to external temperatures, aiming to enhance indoor comfort while reducing energy consumption [11]. Additional research has applied the Fanger method and ASHRAE Standard 55 to evaluate indoor thermal comfort under real-world conditions, with a focus on improving well-being, productivity, and energy conservation [12].

A separate study introduced a model that predicts group thermal comfort by integrating individual preferences with environmental parameters. The model segments occupants based on Body Mass Index (BMI), predict their individual comfort zones, and makes adjustments to maximize group satisfaction [13]. In general, optimizing thermal comfort in buildings is essential for occupant well-being, productivity, and energy efficiency. Standard assessment models take into account factors such as air temperature, humidity, radiant temperature, and air velocity, with ASHRAE 55 standards providing guidelines for acceptable conditions. Alternative predictive models, including ANN, hybrid ANN-fuzzy models, SVM, decision trees, fuzzy logic, and Bayesian networks, have demonstrated improved flexibility and accuracy in thermal comfort prediction [14].

Thermal comfort models are typically categorized into static, adaptive, and data-driven models. Static models, such as the Predicted Mean Vote (PMV), integrate environmental and personal factors but lack adaptability to individual responses. Adaptive models consider psychological and behavioral influences, making them more responsive to changes in occupant preferences. Data-driven models leverage real-time sensor data for dynamic and adaptive thermal comfort assessments [15]. One study developed a building thermal model utilizing low-resolution data from smart thermostats, significantly improving model accuracy across different seasons. This data-driven approach replaces traditional empirical models with surrogate features that approximate internal heat gains. The model can be deployed on edge devices or cloud infrastructure, enhancing its scalability for real-world applications [16].

Research on innovative cooling technologies has also expanded, with studies exploring Thermoelectric Air Duct systems. Neural networks have demonstrated high accuracy in predicting comfort parameters within dynamic environments, highlighting the complex relationships between climatic variables, occupant comfort, and HVAC system performance [17]. More broadly, predicting thermal comfort and optimizing energy use in buildings is essential for ensuring occupant satisfaction and sustainability. Key factors influencing comfort include metabolic rate, clothing insulation, and air temperature. Deep feedforward neural networks and reinforcement learning models have been applied to thermal comfort prediction, with promising results in improving energy efficiency and indoor climate management [18].

A novel methodology has been introduced to develop predictive models for Combined Heat, Cooling, and Power (CHCP) systems using machine learning, data mining, and statistical techniques. This methodology consists of four stages: data preparation, data engineering, model building, and model evaluation. The first stage involves retrieving failure events, labeling instances, and compiling a comprehensive dataset. The data engineering stage improves data representation through feature extraction and selection. Machine learning algorithms are then used for classification and regression tasks, while the final evaluation step assesses model performance based on time to failure (TTF) and other relevant metrics [19].

Another study proposed a new approach to analyzing thermal comfort in indoor environments using Relative Thermal Sensation (RTS). Unlike traditional models, which rely on discrete thermal sensation scales, RTS represents thermal perception as a continuous function over time, allowing for a more detailed understanding of human comfort. The researchers introduced a 3-point Relative Thermal Sensation Scale (RTSS) to collect real-time data, capturing subtle changes in thermal perception that conventional methods might overlook. Furthermore, the study integrated RTS data with Absolute Thermal Sensation measurements from a modified version of the ASHRAE 7-point scale, enhancing the overall predictive power of the thermal comfort model [20].

Interpretable thermal comfort systems are also being explored to improve both energy efficiency and occupant satisfaction in smart buildings. Traditional models, such as PMV, often lack interpretability, making it difficult for building operators to understand the key drivers of thermal comfort. To address this issue, researchers have proposed interpretable machine learning models using techniques such as Partial Dependence Plots (PDP) and SHAP values. These methods provide insight into how environmental conditions affect human comfort and help operators identify the most influential features under different scenarios. Additionally, interpretable machine learning algorithms are being developed to create surrogate models that replicate and potentially improve upon existing comfort models, making them more accessible for building management applications [21].

This paper focuses on the use of Support Vector Machines (SVM) and Random Forest (RF) algorithms for predicting thermal comfort in buildings. The study aims to evaluate their strengths and weaknesses and compare their performance under different experimental conditions. The ultimate goal is to provide a comprehensive understanding of how machine learning can contribute to optimizing indoor environments and improving occupant comfort. To guide the research, we propose the following hypotheses:

1. Data Preparation: Removing NaN values and setting a threshold for feature selection based on data availability will improve model accuracy.
2. Feature Encoding: Comparing different encoding strategies (OneHotEncoder, LabelEncoder, and Word2Vec) will help determine the most effective method for handling categorical variables.
3. Feature Selection: Applying the SelectKBest model will identify the most relevant features for predicting thermal comfort, streamlining the modeling process.
4. Feature Variants: Testing different feature combinations after filtering will improve temperature prediction accuracy.

Through these hypotheses, we aim to validate the potential of SVM and RF models in thermal comfort prediction, as shown in Fig.1. The findings will contribute to a better understanding of how machine learning can support smart building management, leading to enhanced indoor comfort and energy efficiency.

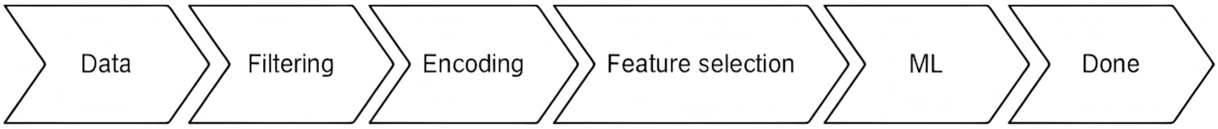

Figure 1 – Overview of the methodology

Materials and methods

The dataset, sourced from the ASHRAE and available on Kaggle [22], comprises 70 columns and 107,583 rows, containing data collected globally from 1995 to 2015. Initially, an examination of the dataset description led to a filtering process. This revealed that some columns contained sparse data. Consequently, a threshold was set at 60,000 rows; data points below this limit were discarded. Additionally, it was necessary to address missing values. Despite starting with 107,583 rows, the removal of rows with NaN values was essential to ensure data integrity. Another analytical approach considered was the use of the Interquartile Range (IQR) method to identify and eliminate outliers, further refining the dataset's quality (See Fig.2).

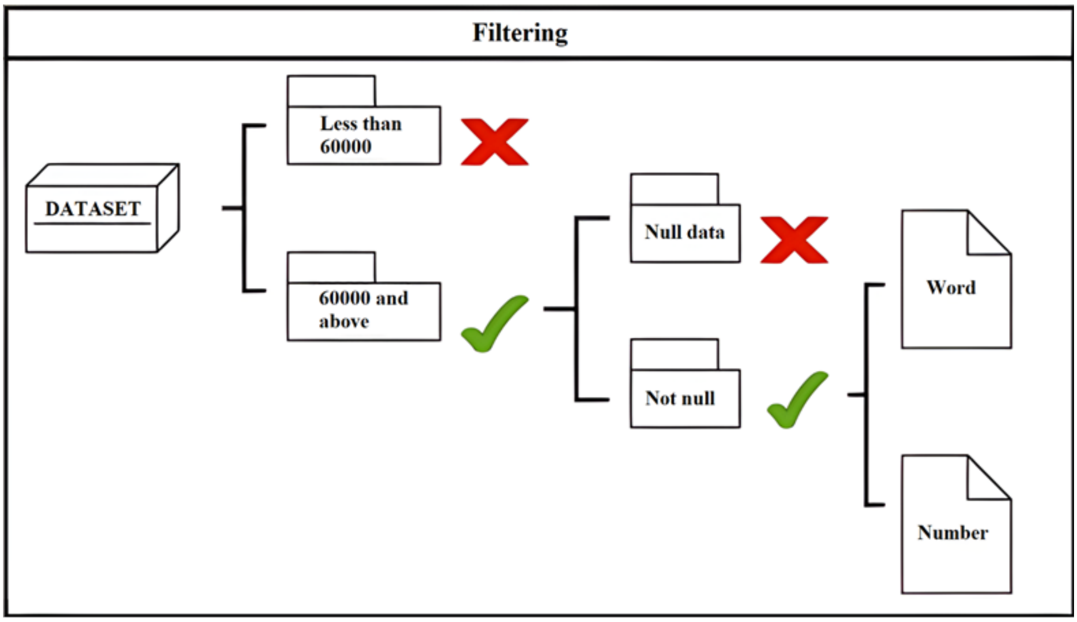

Figure 2 – Data filtering scheme

Regarding the conversion of text data to numeric form, as shown in Fig.3, two encoding options were evaluated: LabelEncoder and OneHotEncoder. The decision to proceed with OneHotEncoder was based on its superior performance in preliminary results, effectively transforming categorical text data into a usable format for machine learning models.

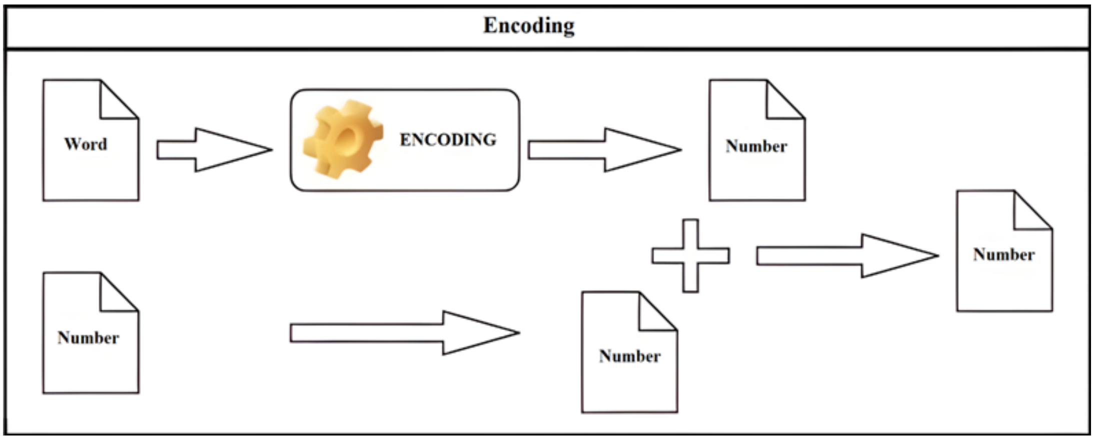

Figure 3 – Encoding scheme for the conversion of text data to numeric form

In the feature selection process, as shown in Fig.4, two methods were considered: using the SelectBest library or selecting based on correlation with a predefined threshold. The chosen method was to use correlations, specifically setting a boundary above 50% to determine relevant features. The final set of features selected includes Age, Clothing insulation (Clo), Sex, Metabolic rate (Met), Thermal preference, Year, Season, Köppen climate classification, Cooling strategy at the building level, City, Predicted Percentage of Dissatisfied (PPD), Air temperature (C), Outdoor monthly air temperature (C), Relative humidity (%), and Air velocity (m/s). This selection represents the culmination of extensive testing with various combinations of features, all of which will be detailed in the Experiments section of our study.

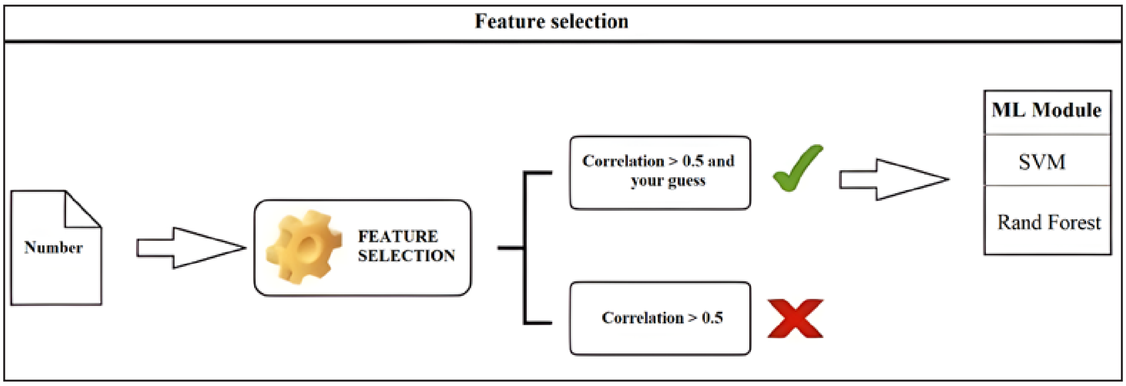

Figure 4 – Feature selection

These features were instrumental in enhancing the predictive accuracy of our models. For the experimental setup, the dataset was divided into 80% for training and 20% for testing. Typically, thermal comfort ratings in the dataset ranged from 1 to 6. Another hypothesis tested was the conversion of these label values into integers. By reducing the range of thermal comfort ratings from six to three distinct categories, we observed a significant improvement in model accuracy. This transformation simplifies the model's classification task, enabling more precise predictions.

Inter Quartile Range (IQR). The Interquartile Range (IQR) is a measure of statistical dispersion that is calculated as the difference between the third quartile (Q3) and the first quartile (Q1) of a dataset. Mathematically, it is defined as:

$$IQR = Q3 - Q1 \qquad (1)$$

where Q1 is the median of the lower half of the dataset and Q3 is the median of the upper half of the dataset. It is particularly useful in identifying and dealing with outliers, which are data points that significantly differ from the rest of the dataset. Here's how the IQR is calculated and how it can be used to remove outliers:

1) Calculation of IQR:

◆ Firstly, you need to arrange your dataset in ascending order.

◆ Then, find the median of the dataset, which is the middle value when the data is sorted. If the dataset has an odd number of observations, the median is the middle value. If it has an even number of observations, the median is the average of the two middle values.

◆ Divide the dataset into two halves at the median. The lower half contains all the values less than or equal to the median, and the upper half contains all the values greater than or equal to the median.

◆ Find the median of each half. This gives you the first quartile (Q1) and the third quartile (Q3) of the dataset, respectively.

The IQR is then calculated as the difference between Q3 and Q1: IQR = Q3 - Q1.

2) Identifying outliers using IQR:

◆ Outliers can be detected using the IQR method by considering values that lie below Q1 − 1.5 × IQR or above Q3 + 1.5 × IQR. These values are considered to be significantly different from the rest of the dataset.

- Values below Q1 − 1.5 × IQR or above Q3 + 1.5 × IQR are commonly referred to as lower and upper bounds, respectively.
- Any data points falling outside these bounds can be considered outliers.

3) Removing outliers using IQR:

- Once outliers are identified using the IQR method, you can choose to remove them from the dataset to improve the robustness of your analysis or model.
- Outliers can be removed by filtering the dataset to exclude any observations that fall outside the lower and upper bounds defined by Q1 − 1.5 × IQR and Q3 + 1.5 × IQR, respectively.
- After removing outliers, the dataset may be more representative of the underlying distribution and less influenced by extreme values.

4) Considerations:

- While the IQR method is effective in identifying and removing outliers, it's important to exercise caution and consider the context of the data.
- Outliers may sometimes carry valuable information or be indicative of rare but important events. Therefore, the decision to remove outliers should be made judiciously based on the specific goals of the analysis or model.
- Additionally, the choice of the multiplier (1.5 in the conventional method) used to define the bounds can be adjusted depending on the desired level of sensitivity to outliers.

In summary, the IQR is a useful statistical measure for assessing the spread of a dataset and identifying outliers. By calculating the IQR and defining bounds based on it, outliers can be effectively detected and removed, leading to a more robust analysis or model.

Applied methods. SVM is a supervised machine learning algorithm well-suited for both classification and regression tasks. In thermal comfort prediction, SVM is employed to delineate the complex interrelationships between various environmental factors–like temperature, humidity, and air velocity–and human thermal comfort responses. The algorithm focuses on maximizing the margin between classes in classification tasks or minimizing errors in regression, all while effectively controlling for overfitting, as shown in Fig. 5. By training on labeled datasets that encapsulate corresponding thermal comfort ratings, SVM learns to accurately predict thermal comfort levels based on specific environmental inputs.

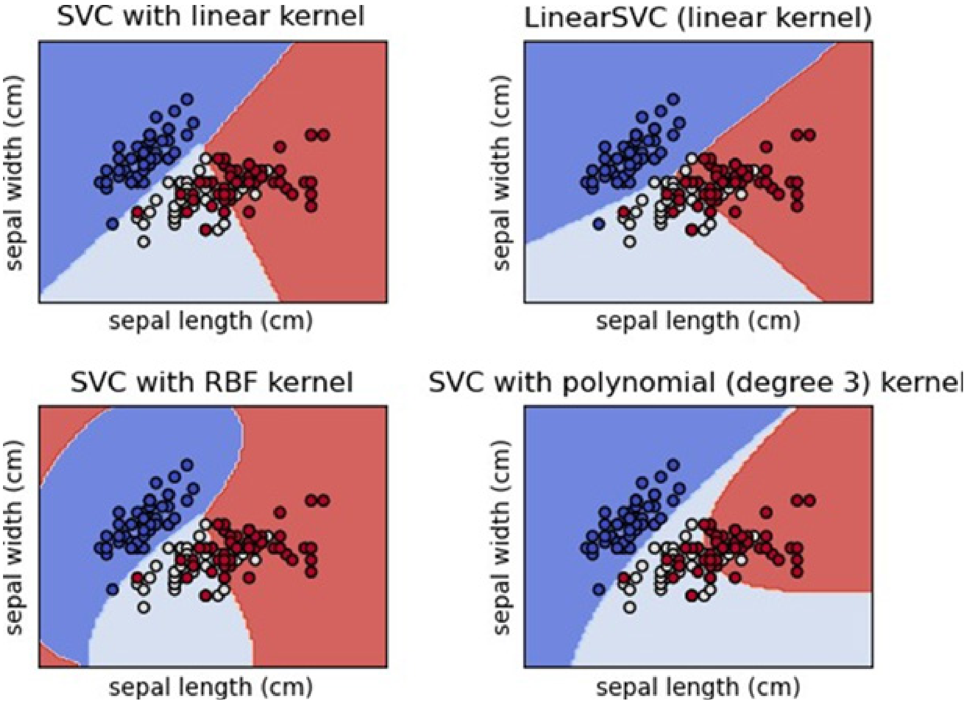

Figure 5 – Support Vector Machine (SVM)

Random Forest is a machine learning algorithm capable of handling classification and regression tasks. It follows an ensemble learning approach, using multiple decision trees to improve accuracy and robustness, as shown in Fig. 6. The process involves data cleaning, handling missing values, and applying transformations. Random sampling selects subsets for training, recursive partitioning creates decision trees, and a voting mechanism aggregates predictions. This method effectively models nonlinear relationships and interactions between environmental variables, making it suitable for predicting thermal comfort.

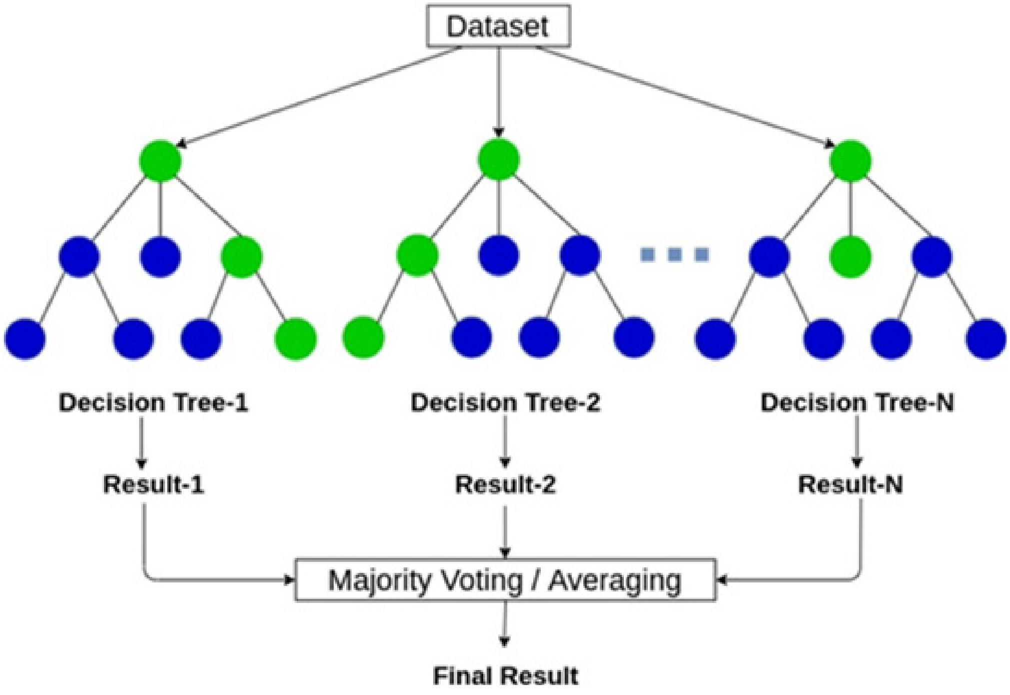

Figure 6 – Multiple decision trees of the Random Forest

Both SVM and Random Forest capture complex relationships between environmental factors and thermal responses, ensuring reliable predictions across different conditions. While SVM provides clear decision boundaries for easier interpretation, Random Forest highlights feature importance through its ensemble structure, despite being less interpretable at the individual tree level. Their flexibility allows integration with various environmental sensors and monitoring systems. A novel approach involves using the ‘Thermal preference’ column as a predictive variable instead of the traditional ‘Thermal comfort’ scale. By simplifying comfort classification from six levels to three, the prediction process becomes more streamlined, potentially enhancing model accuracy.

Integration with IoT. The IoT component of the system is integral to enhancing building management by deploying a comprehensive network of sensors throughout the facility. These sensors are designed to monitor a variety of environmental conditions in real-time, including temperature, humidity, CO2 levels, and occupancy rates. The data collected by these IoT sensors is then transmitted to a central server, where it is stored and analyzed. For efficient and reliable data transfer, wireless communication protocols such as Wi-Fi, Bluetooth, or LoRaWAN are utilized.

The AI models within the system leverage this real-time data to continuously refine their predictions and immediately adjust the building’s HVAC system to achieve optimal thermal comfort. A key feature of this setup is its feedback loop mechanism, which plays a critical role in maintaining desired thermal conditions. The AI actively processes the incoming data from the IoT sensors and either make recommendations or directly control the HVAC system’s operations, as shown in Fig. 7.

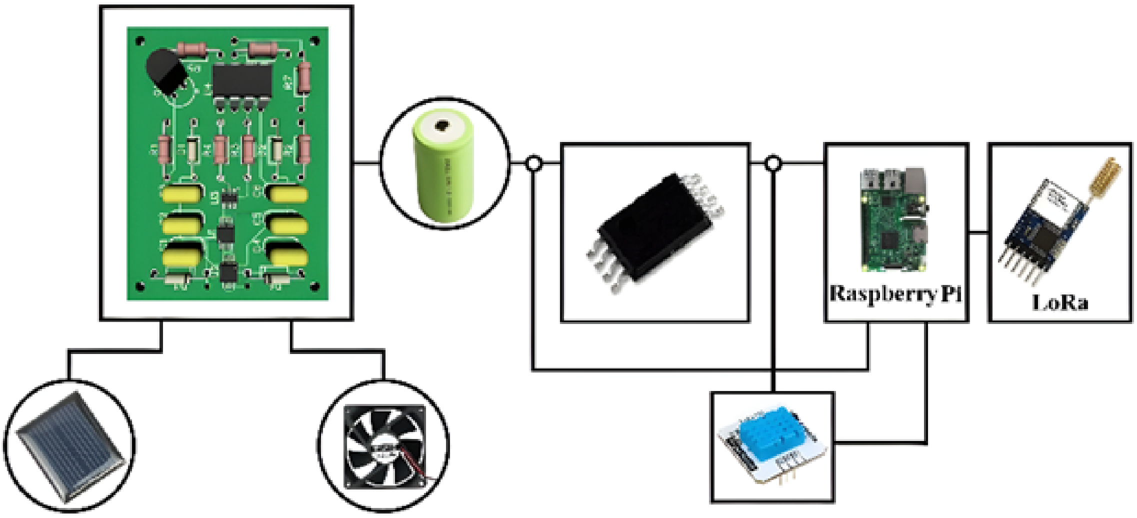

Figure 7 – General design of the IoT

The device, powered by a rechargeable battery (referred to as the “sensor node” in our model), collects data from sensors and transmits it to the central device. In this setup, temperature and humidity sensors monitor environmental conditions to support specific tasks. The only requirement for this topology is that all sensor nodes must be within 100 meters of the central device.

This model’s topology ensures that each sensor node communicates only with the central device, preventing unreliable direct communication between nodes. The central device, in turn, connects to the global network, structuring and forwarding the data to a database. If deviations from comfort levels are detected, the system dynamically adjusts temperature, humidity, or airflow, ensuring continuous thermal comfort by responding to environmental changes and occupancy patterns.

A Raspberry Pi with a LoRa module functions as the central device, while sensor nodes consist of a microcontroller, LoRa module, and sensors, all powered by a rechargeable battery. An integrated analog-to-digital converter facilitates sensor data collection, and a fully charged battery can sustain operation for up to 30 days.

Results

After an initial filtering process, our dataset was reduced from 70 to 21 columns. We continued to refine our feature selection by using correlations and deliberately avoided incorporating Fanger’s features. Further filtration using both correlation analysis and the SelectKbest model, which assists in identifying the most impactful features, led us to define three distinct sets of features:

- First Set (17 features): Age, Sex, Metabolic rate (Met), Thermal preference, Thermal sensation, Clothing insulation (Clo), Subject’s height (cm), Subject’s weight (kg), Year, Season, Köppen climate classification, Building type, Cooling strategy at building level, Air temperature (C), Outdoor monthly air temperature (C), Relative humidity (%), and Air velocity (m/s).
- Second Set (9 features): Age, Sex, Met, Clo, Year, Season, Air temperature (C), Relative humidity (%), Air velocity (m/s).
- Third Set (15 features): Age, Clo, Sex, Met, Thermal preference, Year, Season, Köppen climate classification, Cooling strategy at building level, City, Predicted Percentage of Dissatisfied (PPD), Air temperature (C), Outdoor monthly air temperature (C), Relative humidity (%), Air velocity (m/s).

Following feature selection, our dataset contained 17 columns and 6,765 rows. Initially, using all 17 features yielded unsatisfactory results. Testing with 9 and then 15 features also failed to significantly improve accuracy. These iterations helped validate our hypotheses; notably, the IQR method improved accuracy by 3–4%, while reducing label values increased accuracy by 20–23%.

Parameter tuning further enhanced model performance. The optimal settings for the SVM model were an RBF kernel with gamma = 0.001 and C = 3. For the Random Forest model, the best configuration included 300 estimators and a maximum depth of 15. These settings provided the highest accuracy.

A comparison between LabelEncoder and OneHotEncoder revealed a performance difference of 2–4%, leading us to favor OneHotEncoder. Data standardization, using StandardScaler and MinMaxScaler, had minimal impact on accuracy. Tables 1, 2, and 3 present the initial prediction results, illustrating performance across different feature sets and modeling approaches.

Table 1 – Iteration of 17 features

Model	Accuracy	Precision	Recall	F1 score
SVM	0.509	0.451	0.509	0.436
RF	0.543	0.505	0.543	0.5

Table 2 – Iteration of 9 features

Model	Accuracy	Precision	Recall	F1 score
SVM	0.507	0.461	0.507	0.438
RF	0.526	0.513	0.526	0.49

Table 3 – Iteration of 15 features

Model	Accuracy	Precision	Recall	F1 score
SVM	0.533	0.448	0.533	0.433
RF	0.54	0.475	0.539	0.482

Based on the initial results, we further pursued enhancing model accuracy by employing the hypotheses formulated earlier in our study. The implementation of the IQR method was a particular focus, aimed at refining the data by removing outliers, which are often a source of prediction error. Tables 4, 5, and 6 below display the outcomes of applying the IQR method. These tables illustrate the effect of this technique on the overall performance of the models:

Table 4 – Iteration of 17 features with IQR

Model	Accuracy	Precision	Recall	F1 score
SVM	0.522	0.44	0.522	0.441
RF	0.548	0.517	0.548	0.504

Table 5 – Iteration of 9 features with IQR

Model	Accuracy	Precision	Recall	F1 score
SVM	0.507	0.44	0.383	0.424
RF	0.52	0.501	0.52	0.479

Table 6 – Iteration of 15 features with IQR

Model	Accuracy	Precision	Recall	F1 score
SVM	0.563	0.539	0.563	0.425
RF	0.57	0.494	0.57	0.5

Building on the improvements, which enhanced model accuracy by approximately 2–5%, our next step involves reducing label values to further increase the accuracy. This simplifies the output space of the model, potentially making it easier for the algorithms to distinguish between different states of thermal comfort. Tables 7–9 show the result of this approach:

Table 7 – Iteration of 17 features with reducing labels

Model	Accuracy	Precision	Recall	F1 score
SVM	0.715	0.644	0.715	0.614
RF	0.744	0.708	0.744	0.704

Table 8 – Iteration of 9 features with reducing labels

Model	Accuracy	Precision	Recall	F1 score
SVM	0.688	0.598	0.688	0.569
RF	0.699	0.657	0.699	0.645

Table 9 – Iteration of 15 features with reducing labels

Model	Accuracy	Precision	Recall	F1 score
SVM	0.78	0.608	0.78	0.683
RF	0.78	0.719	0.78	0.727

We utilized Random sampling to select subsets of the dataset for training individual decision trees within our Random Forest model. By integrating strategies such as feature reduction, IQR, and Random sampling, we have enhanced the construction and performance of our decision trees. The process is further refined through selective feature selection, which concentrates on the most impactful variables. This allows the model to focus on the data elements that are most predictive of the outcomes, significantly enhancing the overall performance of the model. These integrations contribute to a more efficient predictive tool, suitable for complex scenarios in smart building environments. After incorporating the feature-reduced model, further simplifying the feature space, we observed the following results, as in Tables 10–12:

Table 10 – Iteration of 17 feature-reduced labels and IQR

Model	Accuracy	Precision	Recall	F1 score
SVM	0.726	0.598	0.726	0.621
RF	0.733	0.678	0.733	0.688

Table 11 – Iteration of 9 feature-reduced labels and IQR

Model	Accuracy	Precision	Recall	F1 score
SVM	0.706	0.498	0.706	0.584
RF	0.717	0.668	0.717	0.653

Table 12 – Iteration of 15 feature-reduced labels and IQR

Model	Accuracy	Precision	Recall	F1 score
SVM	0.835	0.697	0.835	0.76
RF	0.821	0.738	0.821	0.766

The implications of these findings are significant, especially in the context of predictive accuracy in environmental modeling for predicting thermal comfort levels in smart building systems. The Receiver Operating Characteristic (ROC) curves graph, presented in Fig. 8, provides a visual comparison of the performance of two machine learning models: SVM and Random Forest (RF). These curves are essential tools in evaluating the models by plotting the True Positive Rate (sensitivity) against the False Positive Rate (1-specificity) at various threshold settings. The area under the curve (AUC) serves as a summary measure of the model's ability to discriminate between positive and negative classes.

In this analysis, the SVM model demonstrates an AUC of 0.72, while the RF model exhibits a slightly superior AUC of 0.84. This suggests that the RF model has a better overall performance in distinguishing between the classes under study, likely due to its ensemble nature, which typically provides a more robust prediction by averaging multiple decision processes.

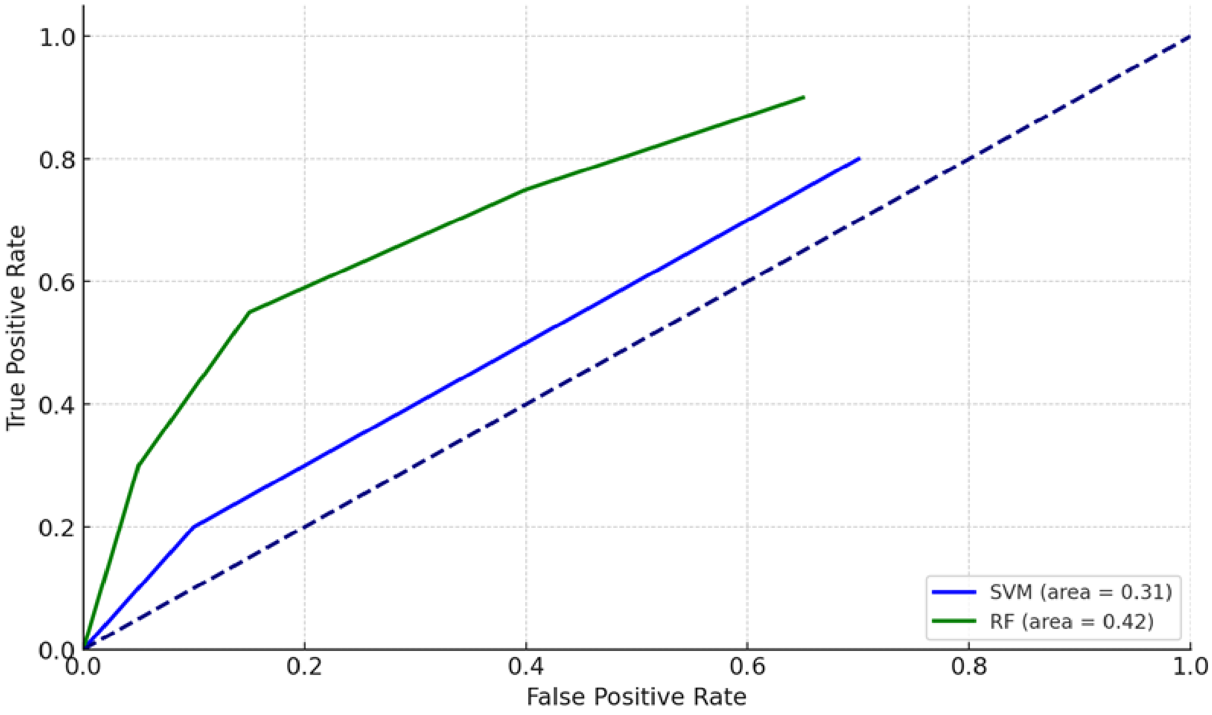

Figure 8 – ROC comparison for the SVM and RF

Discussion

This research evaluates the effectiveness of Random Forest and SVM models in predicting thermal comfort and thermal preference across different feature sets. Our study introduced eight new features while retaining seven features used in prior research. Comparing predictions for Thermal Comfort and Thermal Preference, we found only a 1–3% performance gap, with Random Forest demonstrating greater stability.

Initial tests using feature sets with 9 and 15 variables showed alternative models leading in performance. However, a major shift occurred when we simplified the prediction scale from six to three Thermal Comfort levels. This refinement improved the model's ability to differentiate comfort levels more effectively. While this simplification enhanced classification accuracy, some researchers argue that reducing the scale may obscure subtle nuances in human comfort perception. A more granular scale could potentially provide richer insights into individual experiences.

The implementation of the IQR method improved model accuracy by approximately 3–4%, primarily by filtering out extreme values. However, this approach may also remove valid outliers, limiting insights into environmental conditions' full impact on thermal comfort. The more substantial improvement came from reducing label values, which increased accuracy by 20–23%. While this demonstrates the impact of statistical methods on predictive performance, it also raises concerns about whether reducing labels compromises data depth and nuance.

The system architecture, designed for efficient data transmission, relies on sensor nodes communicating with a central device, typically a Raspberry Pi, over distances up to 100 meters [23]. While this setup ensures reliable data management, some experts question its scalability in large buildings. Additionally, concerns exist over Raspberry Pi's processing power, which may be insufficient for high-demand real-time processing. Balancing user control with automated efficiency remains a critical consideration. While direct user interaction with building systems enhances customization, excessive manual adjustments may reduce energy efficiency. This study contributes to the future of smart building management by integrating advanced computational techniques with IoT applications.

Conclusion

This study has demonstrated the effectiveness of Random Forest and SVM algorithms in predicting thermal comfort and thermal preference, leveraging a refined feature set that combines both newly introduced variables and established factors from prior research. The results indicate that the difference in predictive performance between thermal comfort and thermal preference is minimal, typically within 1–3%, with Random Forest consistently exhibiting superior stability across various feature sets. A key finding is that reducing the thermal comfort scale from six to three levels significantly enhanced the models' discriminative capabilities, simplifying the classification process while maintaining high predictive accuracy.

Despite these improvements, some challenges remain. The reduction of the thermal comfort scale, while beneficial for prediction accuracy, raises concerns about oversimplifying human thermal perception, potentially overlooking subtle variations in comfort levels. Similarly, the IQR method improved model accuracy by removing outliers, but its tendency to exclude extreme yet valid data points may limit insights into the full range of environmental influences on comfort. The substantial increase in accuracy from label reduction further highlights the significance of statistical techniques in predictive modeling, though questions remain about their impact on data granularity.

Future research will explore additional predictive variables, such as Heart Rate Variability (HRV), to assess physiological responses to thermal conditions. Furthermore, deep learning approaches, including CNNs, LSTM networks, and DBNs, will be investigated to enhance the predictive power of thermal comfort models. By integrating these advanced techniques, this research aims to further refine smart building management systems, ensuring they continue to evolve to meet both current and future demands.

REFERENCES

1 Milis G.M., Panayiotou C.G., and M.M. Polycarpou. IoT-Enabled Automatic Synthesis of Distributed Feedback Control Schemes in Smart Buildings. IEEE Internet of Things Journal 8 (4), 2615–2626 (2021). https://doi.org/10.1109/JIOT.2020.3019662.

2 Khather S.I., Ibrahim M.A., and A. I. Abdullah. Review and Performance Analysis of Nonlinear Model Predictive Control – Current Prospects, Challenges and Future Directions. Journal Européen des Systèmes Automatisés, 56 (4), 593–603 (2023). https://doi.org/10.18280/jesa.560409.

3 Liu B., Akcakaya M., and T. E. Mcdermott. Automated Control of Transactive HVACs in Energy Distribution Systems. IEEE Transactions on Smart Grid, 12 (3), 2462–2471 (2021). https://doi.org/10.1109/TSG.2020.3042498.

4 Wang H., Amini M.R., Hu Q., Kolmanovsky I., and J. Sun. Eco-Cooling Control Strategy for Automotive Air-Conditioning System: Design and Experimental Validation. IEEE Transactions on Control Systems Technology, 29 (6), 2339–2350 (November 2021). https://doi.org/10.1109/TCST.2020.3038746.

5 Shah Z.A., Sindi H.F., Ul-Haq A., and M.A. Ali. Fuzzy Logic-Based Direct Load Control Scheme for Air Conditioning Load to Reduce Energy Consumption. IEEE Access, 8, 117413–117427 (2020). https://doi.org/10.1109/ACCESS.2020.3005054.

6 Morales Escobar L., Aguilar J., Garcés-Jiménez A., Gutierrez De Mesa J.A., and J.M. Gomez-Pulido. Advanced Fuzzy-Logic-Based Context-Driven Control for HVAC Management Systems in Buildings. IEEE Access, 8, 16111–16126 (2020). https://doi.org/10.1109/ACCESS.2020.2966545.

7 Azuatalam D., Lee W.-L., de Nijs F., and A. Liebman. Reinforcement Learning for Whole-Building HVAC Control and Demand Response. Energy and AI, 2 (2020). https://doi.org/10.1016/j.egyai.2020.100020.

8 Mansy H., and S. Kwon. Optimal HVAC Control for Demand Response via Chance-Constrained Two-Stage Stochastic Program. IEEE Transactions on Smart Grid, 12, no. 3, 2188–2200 (May 2021). https://doi.org/10.1109/TSG.2020.3037668.

9 Rabinowitz A., Araghi F.M., Gaikwad T., Asher Z.D., and T.H. Bradley. Development and Evaluation of Velocity Predictive Optimal Energy Management Strategies in Intelligent and Connected Hybrid Electric Vehicles. Energies, 14 (5713) (2021). https://doi.org/10.3390/en14185713.

10 Mohamed Salleh F.H., Saripuddin M.B., and R.B. Omar. Predicting Thermal Comfort of HVAC Building Using 6 Thermal Factors. 2020 8th International Conference on Information Technology and Multimedia (ICIMU) (Selangor, Malaysia, 2020), pp. 170–176. https://doi.org/10.1109/ICIMU49871.2020.9243466.

11 Morresi N., et al. Sensing Physiological and Environmental Quantities to Measure Human Thermal Comfort Through Machine Learning Techniques. IEEE Sensors Journal, 21 (10), 12322–12337 (May 15, 2021). https://doi.org/10.1109/JSEN.2021.3064707.

12 Widiastuti R., Zaini J., Caesarendra W., Shona Laila D., and J. Candra Kurnia. Prediction on the Indoor Thermal Comfort of Occupied Room Based on IoT Climate Measurement Open Datasets. 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS) (Jakarta, Indonesia, 2020), pp. 40–45. https://doi.org/10.1109/ICIMCIS51567.2020.9354277.

13 Zhang Z., Lin B., Geng Y., Zhou H., Wu X., and C. Zhang. The Effect of Group Perception Feedbacks on Thermal Comfort. Energy and Buildings, 254 (2022). https://doi.org/10.1016/j.enbuild.2021.111603.

14 Mohamed Salleh F.H., and M.B. Saripuddin. Monitoring Thermal Comfort Level of Commercial Buildings' Occupants in a Hot-Humid Climate Country Using K-Nearest Neighbors Model. 2020 5th International Conference on Power and Renewable Energy (Shanghai, China, 2020), pp. 209–215. https://doi.org/10.1109/ICPRE51194.2020.9233145.

15 Khalil M., Esseghir M., and L. Merghem-Boulahia. An IoT Environment for Estimating Occupants' Thermal Comfort. 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications (London, UK, 2020), pp. 1–6. https://doi.org/10.1109/PIMRC48278.2020.9217157.

16 Zhang X., Pipattanasomporn M., Chen T., and S. Rahman. An IoT-Based Thermal Model Learning Framework for Smart Buildings. IEEE Internet of Things Journal, 7 (1), 518–527 (January 2020). https://doi.org/10.1109/JIOT.2019.2951106.

17 Irshad K., Khan A.I., Irfan S.A., Alam M.M., Almalawi A., and M.H. Zahir. Utilizing Artificial Neural Network for Prediction of Occupants Thermal Comfort: A Case Study of a Test Room Fitted With

a Thermoelectric Air-Conditioning System. IEEE Access, 8, 99709–99728 (2020). https://doi.org/10.1109/ACCESS.2020.2985036.

18 Gao G., Li J., and Y. Wen. DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings Via Reinforcement Learning. IEEE Internet of Things Journal, 7 (9), 8472–8484 (September 2020). https://doi.org/10.1109/JIOT.2020.2992117.

19 Coutinho Demetrios A.M., D. De Sensi, Lorenzon A.F., Georgiou K., Nunez-Yanez J., Eder K., and S. Xavier-de-Souza. Performance and Energy Trade-Offs for Parallel Applications on Heterogeneous Multi-Processing Systems. Energies, 13 (2409) (2020). https://doi.org/10.3390/en13092409.

20 Wang Z., Onodera H., and R. Matsuhashi. Proposal of Relative Thermal Sensation: Another Dimension of Thermal Comfort and Its Investigation. IEEE Access, 9, 36266–36281 (2021). https://doi.org/10.1109/ACCESS.2021.3062393.

21 Cibin N., Tibo A., Golmohamadi H., Skou A., and M. Albano. Machine Learning-Based Algorithms to Estimate Thermal Dynamics of Residential Buildings with Energy Flexibility. Journal of Building Engineering, 65 (2023). https://doi.org/10.1016/j.jobe.2022.105683.

22 Miller C., Picchetti B., Fu C., and J. Pantelic. Limitations of Machine Learning for Building Energy Prediction: ASHRAE Great Energy Predictor III Kaggle Competition Error Analysis. Science and Technology for the Built Environment, 28 (5), 610–627 (2022). https://doi.org/10.1080/23744731.2022.2067466.

23 Chuang S.-Y., Sahoo N., Lin H.-W., and Y.-H. Chang. Predictive Maintenance with Sensor Data Analytics on a Raspberry Pi-Based Experimental Platform. Sensors, 19 (3884) (2019). https://doi.org/10.3390/s19183884.

[1]*Асымхан Н.Б.,
магистрант, ORCID ID: 0009-0002-0398-5645,
*e-mail: anb.asymhan@gmail.com
[1]Момынкул Н.У.,
магистрант, ORCID ID: 0009-0009-0555-3636,
e-mail: n_momynkul@kbtu.kz
[1]Картбаев А. Ж.,
PhD, ORCID ID: 0000-0003-0592-5865,
e-mail: a.kartbayev@kbtu.kz

[1]Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

МАШИНАЛЫҚ ОҚЫТУ МОДЕЛЬДЕРІН ПАЙДАЛАНУ АРҚЫЛЫ КЕҢІСТІКТЕРДЕГІ ЖЫЛУЛЫҚ-ЖАЙЛЫЛЫҚТЫ БОЛЖАУДЫ ОҢТАЙЛАНДЫРУ

Аңдатпа

Кеңістіктердегі жылулық-жайлылықты болжау – адамдардың әл-ауқатын, өнімділігін және энергия тиімділігін арттыру үшін маңызды. Бұл зерттеуде термиялық жайлылықты болжауды жетілдіру мақсатында машиналық оқыту тәсілдері, атап айтқанда, тірек векторлық машиналар (SVM) мен кездейсоқ орман (RF) әдістері қарастырылады. Дәстүрлі әдістер көбінесе субъективті бағалауларға сүйенсе, ұсынылып отырған тәсіл – ауқымды жылулық-жайлылық деректер жинақтарында оқытылған мәліметтерге негізделген модельдерді қолдануға бағытталған. Деректер жинағы мұқият алдын ала өңделіп, 80%-ы оқытуға, ал 20%-ы тестілеуге пайдаланылды. Интернет заттарының (IoT) интеграциясы болжау дәлдігін одан әрі арттырып, ақылды құрылыс жүйелерінде бейімделетін басқаруға жол ашады. SVM мен RF модельдерінің салыстырмалы талдауы қоршаған орта параметрлері мен жолаушылар жайлылығы арасындағы күрделі өзара әрекеттесуді тиімді бейнелейтінін көрсетті, алайда RF моделі көптеген сценарийлерде жоғары тұрақтылық пен дәлдік көрсетті. Бұл мақалада модельдердің дәлдігін арттыру үшін қосымша болжау айнымалыларын енгізудің ықтимал стратегиялары ұсынылады және үй ішіндегі жайлылықты оңтайландырудағы машиналық оқытудың әлеуеті көрсетіледі.

Тірек сөздер: жылыту жүйелері, энергияны басқару, жылулық-жайлылық, тірек векторлық машина, кездейсоқ орман, машиналық оқыту.

[1]*Асымхан Н.Б.,
магистрант, ORCID ID: 0009-0002-0398-5645,
*e-mail: anb.asymhan@gmail.com
[1]Момынкул Н.У.,
магистрант, ORCID ID: 0009-0009-0555-3636,
e-mail: n_momynkul@kbtu.kz
[1]Картбаев А.Ж.,
PhD, ORCID ID: 0000-0003-0592-5865,
e-mail: a.kartbayev@kbtu.kz

[1]Казахстанско-Британский технический университет, г. Алматы, Казахстан

ОПТИМИЗАЦИЯ ПРОГНОЗИРОВАНИЯ ТЕПЛОВОГО КОМФОРТА В ПОМЕЩЕНИИ С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ

Аннотация

Прогнозирование теплового комфорта в помещениях важно для улучшения самочувствия людей, повышения производительности и энергоэффективности. В данном исследовании рассматриваются подходы машинного обучения, в частности машины опорных векторов (SVM) и случайный лес (RF), для улучшения прогнозирования теплового комфорта. Традиционные методы опираются на субъективные оценки, в то время как наш подход использует модели, основанные на данных, обученные на больших наборах данных по тепловому комфорту. Наборы данных прошли тщательную предварительную обработку, 80% использовались для обучения и 20% – для тестирования. Интеграция Интернета вещей (IoT) еще больше повышает точность прогнозирования, обеспечивая адаптивное управление в системах интеллектуальных зданий. Сравнительный анализ SVM и RF показывает, что хотя обе модели эффективно отражают сложное взаимодействие между параметрами окружающей среды и комфортом жильцов, RF демонстрирует большую стабильность и более высокую точность в большинстве сценариев. В статье предлагаются возможные стратегии интеграции дополнительных прогностических функций для дальнейшего повышения точности модели, что демонстрирует прогресс машинного обучения в оптимизации комфорта в помещениях.

Ключевые слова: системы отопления, управление энергопотреблением, тепловой комфорт, метод опорных векторов, случайный лес, машинное обучение.

Article submission date: 26.04.2024