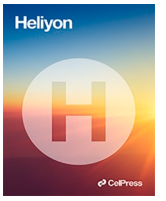
Heliyon 11 (2025) e41572

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Enhanced ensemble learning-based uncertainty and sensitivity analysis of ventilation rate in a novel radiative cooling building

Majid Mohsenpour [a], Mohsen Salimi [b], Atieh Kermani [a], Majid Amidpour [a,*]

[a] Department of Energy System Engineering, Faculty of Mechanical Engineering, K.N. Toosi University of Technology, No. 15, Pardis St., Molasadra Ave., Vanak Sq., Tehran, Iran
[b] Renewable Energy Research Department, Niroo Research Institute (NRI), Tehran, Iran

ARTICLE INFO

Keywords:
Global warming
Natural cooling
Enhanced ensemble learning
Uncertainty analysis
Sensitivity analysis

ABSTRACT

The rising global demand for air conditioning systems, driven by increasing temperatures and urbanization, has led to higher energy consumption and greenhouse gas emissions. HVAC systems, particularly AC, account for nearly half of building energy use, highlighting the need for efficient cooling solutions. Passive cooling, especially radiative cooling, offers potential to reduce cooling loads and improve energy efficiency. However, most studies focus on idealized conditions, neglecting the real-world variability of indoor and outdoor environments. This study proposes a novel machine learning-based ensemble stacking model to predict ventilation rates in passive cooling buildings, addressing the challenges of black-box modeling. The model's performance is improved across key metrics such as R^2, RMSE, and MAE. For the first time, uncertainty and sensitivity analysis is applied to assess the impact of indoor and outdoor conditions on ventilation rates. Sensitivity analysis shows that the reference model's ventilation rate highly depends on inlet air temperature, internal temperatures at 0.1 and 0.2 m, and internal wall heat flux, with optimization of these parameters having a significant impact on building performance. In contrast, the test building relies on fewer parameters, with external temperature, outlet air temperature, and net roof radiation being notable factors; as ambient temperature increases, so does the ventilation rate. The analysis reveals that uncertainties have minimal impact in the reference building, while the test building demonstrates greater sensitivity during warmer months, emphasizing the importance of accounting for seasonal variations. This research underscores the significance of optimizing key features to enhance natural cooling and ventilation, contributing to sustainable climate control solutions and providing an interpretable, robust model for predicting ventilation rates in energy-efficient buildings.

1. introduction

In recent years, energy consumption has surged significantly due to the increasing reliance on air conditioning (AC) systems. This trend is expected to escalate sharply, particularly in middle-income countries, as family incomes rise and global temperatures increase [1]. Given that HVAC systems, including air conditioning, account for approximately half of the total energy consumed in buildings, minimizing their energy consumption is of paramount importance [2,3]. As illustrated in Fig. 1, the global trend in AC usage since 1990, with projections extending to 2050 [4], underscores the urgency of addressing this issue.

* Corresponding author.
E-mail address: amidpour@kntu.ac.ir (M. Amidpour).

https://doi.org/10.1016/j.heliyon.2024.e41572
Received 12 September 2024; Received in revised form 21 December 2024; Accepted 29 December 2024
Available online 31 December 2024

The heightened energy demand associated with AC systems has precipitated energy crises, particularly in developing nations, exacerbating environmental challenges [5]. The heavy reliance on fossil fuel-based energy sources for air conditioning not only leads to increased greenhouse gas emissions [6] but also contributes to air pollution and climate change [7–9]. Therefore, it is critical to advance sustainable and energy-efficient air conditioning technologies to mitigate these adverse effects [10]. Although there has been notable growth in the development of environ mentally friendly cooling technologies, these advancements remain insufficient to meet the growing demand [11].

One of the proper ways to mitigate the adverse effect of active cooling is passive cooling in buildings [12,13]. By doing so, the required cooling load can be reduced, enhancing the efficiency of buildings [14,15]. In recent years, radiative cooling has piqued the interest of scholars, as it can be easily employed in buildings to reduce cooling load demand [16]. Zeyghami et al. [17] Proposed principles of radiative cooling in buildings for both daytime and nighttime cooling. Xu et al. [18] Proposed an emitter composed of six nanospheres embedded in a Polyvinylidene fluoride (PVDF) matrix to enhance radiative cooling. Tang et al. [19] proposed a radiative cooling composite for window glass. This study demonstrated that the proposed material effectively reduced buildings' indoor temperatures. Gong et al. [20] introduced an innovative design for a self-regulating radiative cooler. In this study, paraffin is combined with a radiative cooler to enable dynamic adaptation to temperature changes through liquid-solid transitions. The results indicate that the proposed coolers can lead to energy savings in buildings across various cities and climates. Luo et al. [21] introduced an innovative method for passive radiative cooling during the day, utilizing a coating made of yttrium oxide (Y_2O_3) and polydimethylsiloxane (PDMS). This design effectively enhances solar reflectance and emissivity on the building window, providing high radiative cooling efficiency even under intense solar radiation. Tang et al. [22] Evaluated the effectiveness of Magnesium oxide (MgO) radiative cooling in buildings and conducted a life-cycle assessment of the material. Zhang et al. [23] assessed the effectiveness of super-amphiphobic in passive cooling of buildings. In another study, Feng et al. [24] produced an inorganic coating to cool a building passively. Sun et al. [25] utilized a woody laminate to make a building efficient in terms of energy.

This body of work represents significant advancements in passive cooling technology, effectively reflecting solar radiation and facilitating the transfer of interior heat to the cooler outdoor environment through building openings [26–32]. These innovative materials can substantially increase the rate of heat rejection from building interiors, contributing to the mitigation of urban heat island effects [33–38]. However, while promising, the majority of these studies have primarily focused on evaluating cooling performance of new materials under idealized conditions, often neglecting the substantial variability inherent in real-world building environments where both interior and exterior conditions can fluctuate significantly.

Achieving an optimal cooling rate, which necessitates the efficient removal of heat, presents challenges that are often underestimated [39]. On particularly hot days, these surface materials can indeed lower the temperature differential across external insulation. However, if the temperature gradient reverses, inhibiting outward heat flow becomes problematic due to insulation effects. Moreover, the post-COVID-19 emphasis on healthy ventilation further complicates the rejection of heat from indoor environments. As efforts to cool indoor spaces below ambient temperatures increase, outdoor air exchanges can quickly become the predominant source of heat [40].

Consequently, an alternative approach must be considered. Fortin et al. [40] conducted an experimental study proposing a novel design for a single-occupant building that integrates radiative cooling with buoyancy-driven airflow, targeting energy efficiency in hot and dry climates. The study involved creating a temperature gradient between the interior and exterior, utilizing aluminum sheets coated with white polyvinylidene fluoride (PVDF) mounted on a T-shaped aluminum frame above the roof as radiative cooling surfaces. Additionally, water-filled chambers served as thermal storage to enhance the temperature differential, generating buoyancy forces to drive airflow. This innovative design was benchmarked against a conventional reference box, where results demonstrated that the model maintained an interior temperature 3.9 °C cooler than the external average and 8.9 °C lower than traditional standards. The

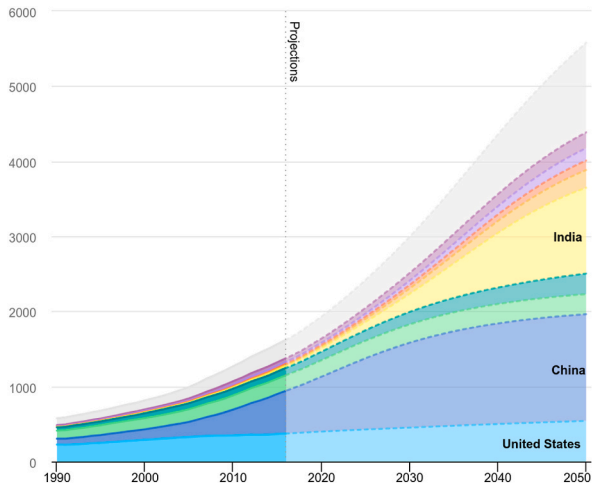


Fig. 1. Global Air Conditioner Stocks, per million units, 1990–2050 [4].

findings highlight that this model can effectively reduce cooling loads across various building types, leveraging the synergy between radiative cooling and buoyancy forces. With its use of simple, inexpensive, accessible, and environmentally friendly materials, this approach offers a viable cooling solution, particularly impactful in enhancing ventilation rates and minimizing energy demands in hot and humid climates.

However, considerable uncertainties exist regarding practical building variables, including indoor and outdoor temperatures, humidity, heat flux, and the thermal mass of walls. These uncertainties play a critical role in the performance of ventilation rates and can profoundly impact indoor air quality [41,42]. Although the results of the Fortin study has shown potential, it has yet to comprehensively assess how these indoor and outdoor conditions influence ventilation rates. Addressing this gap is essential for the effective design and control of passive cooling systems.

To illustrate, Bijarniya et al. [43] evaluated the feasibility of passive cooling across various Indian cities through a comparative analysis of three types of radiative cooler surfaces, yet did not consider indoor factors such as the building's internal temperature. Similarly, Wu et al. [44]examined the cooling effect of radiative sky cooling within indoor environments, while neglecting other critical influences, such as solar irradiance.

A review of existing literature reveals that previous research has not fully addressed the impact of both internal and external conditions on ventilation rates in passive cooling buildings. While studies have made significant strides in predicting ventilation rates through numerical simulations and experimental measurements [45–48], discrepancies often arise between the results of numerical models and experimental data [47]. This inconsistency underscores the need for alternative approaches to improve prediction reliability. In this context, data-driven methods, particularly machine learning models, have shown promise due to their ability to capture complex linear and nonlinear relationships in datasets [49,50]. Despite their potential, the application of single machine learning models—commonly referred to as "black-box" methods—remains widespread, especially for energy-related predictions in buildings [51,52].

One major limitation of single-model approaches is their inherent instability. For instance, algorithms like XGBoost (Xgb) are prone to fluctuations in output when small changes are made to the input data, which limits their practical application in real-time scenarios [53]. To address this, ensemble learning techniques have recently been introduced as a means to combine multiple models and reduce instability, enhancing prediction accuracy for both classification and regression tasks [54]. However, while ensemble methods have been widely used in various domains due to their reliability and accuracy [55–58], their application in estimating ventilation rates for passive cooling buildings has not been thoroughly explored. Moreover, an important gap in the literature lies in determining the optimal combination of base learners for ensemble methods. While previous studies have demonstrated the use of ensemble models, such as in gas turbine efficiency forecasting [56] or performance detection in aviation [55], they do not detail how base learners were selected, which may result in overfitting and increased computational time. Additionally, the use of limited base learners, such as in energy consumption forecasting [59], can limit the generalizability of the model to other applications.

A recent study by Wang et al. [60] addressed this gap by employing an exhaustive search method to evaluate all possible combinations of six base models for building energy prediction. This approach identified an optimal ensemble configuration that reduced the Mean Absolute Percentage Error (MAPE) by 4.9%–6.6 % compared to the best single model while improving computational efficiency by eliminating redundant base learners. However, despite its thoroughness, Wang et al.'s method only used a limited set of six base models, leaving potential for further improvement with a more diverse selection of learners and by adopting advanced optimization techniques.

The present study fills this research gap by extending the principles of exhaustive search with a more extensive ensemble framework and a Genetic Algorithm (GA) optimization approach. Specifically, we incorporate a broader array of ten base models—K-Nearest Neighbors (K-NN), Multilayer Perceptron (MLP), Random Forest (RF), Support Vector Regression (SVR), Linear Regression (LR), AdaBoost, Lasso, XGBoost, Decision Tree (DT), and LightGBM—allowing for greater flexibility and more tailored model configurations. GA is employed to systematically optimize the combination and selection of base learners, enhancing model robustness. By doing so, we bridge the research gap related to the optimal selection and configuration of ensemble models, providing a scalable, high-performing ensemble solution adaptable to diverse applications beyond the limited scope of previous studies.

This expanded framework not only addresses the limitations in base model selection highlighted by Wang et al. [60] but also advances ensemble modeling by enhancing predictive accuracy, establishing a robust foundation for ventilation rate prediction and other related fields.

Another critical challenge in using machine learning models is the lack of interpretability, which can obscure the understanding of underlying patterns in the data. To overcome this, the Shapley Additive Explanations (SHAP) method, widely used for providing interpretability in complex systems, is integrated into this study [61–63]. However, the application of SHAP to assess the influence of indoor and outdoor conditions on ventilation rates remains unexplored. This research fills this gap by presenting the first comprehensive model for predicting ventilation rates in passive cooling buildings while identifying the key parameters influencing these rates.

This study's primary contributions include:

1. Proposing a novel ensemble learning framework for the sensitivity and uncertainty analysis of indoor and outdoor conditions affecting ventilation rates in passive cooling systems.
2. Utilizing Genetic Algorithms to optimize base learner selection and combination in the stacking ensemble model, enhancing both accuracy and robustness.
3. Identifying critical factors that influence ventilation rates, which can inform the design of more effective passive cooling solutions and open new avenues for future research.

To validate the proposed model, we utilize four datasets extracted from a study by Fortin et al. [41], focusing on two buildings during November and September. These datasets represent different ventilation rate patterns, ensuring the model's robustness across varying conditions. Recursive feature elimination is used to prioritize the most influential variables. The findings of this research are organized into five sections: methodology, results of the enhanced ensemble model incorporating uncertainty and sensitivity analysis, conclusions, and research limitations, challenges and future prospects.

2. Methodology

Topanga Canyon, known as hot and dry area, is in the Santa Monica Mountains in Los Angeles County, California. The data of the research were collected from the recent work of Fortin et al. [40]. The extracted data are divided into two categories: reference (as the

Table 1
Dataset description.

Abbreviation	Name	Unit	Explanation
VR	Ventilation Rate	m/s	The recorded ventilation rate in both buildings (Reference and Test), which is the target of modeling in this study. All below parameters are used as inputs in the machine learning model.
T	Air Temperature	°C	The ambient air temperature around the building, which influences heat exchange with the structure.
WS	Wind Speed	m/s	The speed at which wind is moving across the building's exterior, affecting ventilation and cooling through convection.
NLR	Net Longwave Radiation	W/m²	Net thermal radiation emitted from the building surface, driven by temperature differences between the building and the atmosphere.
EST	Effective Sky Temperature	°C	Apparent temperature of the sky as perceived by the building, impacting the building's radiative heat exchange.
ARH	Ambient Relative Humidity	%	The relative moisture content of the air surrounding the building, influencing condensation and heat transfer.
WD	Wind Direction	–	Directional source of wind affecting the building, determining which sides are exposed to airflow.
AI	Atmospheric Irradiance	W/m²	Intensity of solar radiation impacting the building's exterior, contributing to heat gain.
BP	Barometric Pressure	mbar	Atmospheric pressure around the structure, influencing weather patterns and the rate of heat transfer.
DP	Dew Point Temperature	°C	The temperature where air reaches full saturation and condensation begins, relevant for moisture control on building surfaces.
WG	Wind Gust	m/s	Measures peak wind speeds occurring in short bursts, which can affect the building's stability and cooling patterns.
SI	Solar Irradiance	W/m²	The intensity of direct solar energy on building surfaces, contributing to internal heat gain.
HFTSR	Heat Flux Top Side Radiator	W/m²	Rate of heat transfer measured on the radiator's top surface, showing heat gain or loss to the surrounding air.
HFISR	Heat Flux Interior Side Radiator	W/m²	Heat transfer occurring on the radiator's indoor surface, indicating the amount of heat released into the room.
HFIWR	Heat Flux Interior Wall Reference/Test	W/m²	Heat flux across interior walls in reference or test setups, showing heat retention or dissipation inside the building.
STTMR/T	Surface Temp. Thermal Mass Reference/Test	°C	Temperature at the surface of the thermal mass in each setup, showing heat storage capability.
ATCG	Air Temp. at Top Convection Guard	°C	Temperature at the convection guard's top edge, which helps control convective heat transfer rates in both building.
FVIR	Flow Velocity at Inlet Reference/Test	m/s	Velocity of air entering the building in either setup, affecting ventilation and cooling efficiency.
HFTMR/T	Heat Flux Thermal Mass Reference/Test	W/m²	Heat flow through the thermal mass in reference and test buildings, indicating comparative heat retention properties.
STTSR	Surface Temp. Top Side Radiator	°C	Temperature on the top side of the radiator's surface, indicating heat exchange between the radiator and the ambient air.
STIWR/T	Surface Temp. Interior Wall Reference/Test	°C	Temperature of interior wall surfaces in reference and test buildings, showing differences in heat absorption and retention.
IT0.1 R/T	Interior Temp. at 0.1m Reference/Test	°C	Measures indoor temperature at a height of 0.1 m, showing thermal conditions near the floor level in each setup.
FVOR/T	Flow Velocity at Outlet Reference/Test	m/s	Airflow velocity exiting the building, which helps assess exhaust rates and ventilation effectiveness in each setup.
STTSR/T	Surface Temp. Thermal Mass Reference/Test	°C	Surface temperature of the thermal mass, reflecting the efficiency of heat retention across the different setups.
STISR	Surface Temp. Interior Side Radiator	°C	Temperature on the indoor side of the radiator, showing how much heat is transferred into the room.
HSR/T	Heat Source Reference/Test	W/m²	Heat load applied within each building setup.
IT0.2 R/T	Interior Temp. at 0.2m Reference/Test	°C	Indoor temperature at a height of 0.2 m, indicating thermal distribution across vertical levels in each building.
ATIR&AT0R/T	Air Temp at Inlet and Outlet Reference/Test	°C	Temperature of air entering and exiting each building, showing the overall heating or cooling effect along the ventilation path.

baseline) and test (as the proposed model). Each dataset was recorded daily in September and November at 15-min intervals. These data include both internal and external conditions. Table 1 provides a description of the recorded variables in the datasets used for this study. In this revision, we have specified the removal of FVOR and FVIR to prevent any potential correlation effects that could compromise model accuracy. Additionally, certain time-dependent indoor and outdoor variables have been excluded, as these could introduce correlations that might affect the model's effectiveness. The primary goal of this research is to present an enhanced ensemble model that combines stacking and a genetic algorithm, forming the basis for sensitivity and uncertainty analysis. This model assesses the influence of various features on ventilation rates across multiple datasets. Fig. 2 outlines the research methodology.

To achieve this, the data are processed, and essential features are selected using the recursive feature elimination method. The dataset is divided into 80 % for training and 20 % for testing, ensuring that the test set remains unseen throughout the training process. From the 80 % training set, 20 % is used as a validation set, following the concept of k-fold cross-validation to enhance model robustness and avoid overfitting.

Several base learners, including KNN, RF, LR, SVR, DT, MLP, AdaBoost, GBM, LightGBM, and Lasso, are employed to construct the stacked ensemble model. Details of the hyperparameters for the base and meta-learners are shown in Tables A4 and A5 in Appendix A. The GA optimizes the selection and combination of these base learners to minimize the RMAE, thereby improving the accuracy of predictions. Finally, LR serves as the meta-learner to consolidate predictions from the base learners, creating a robust model for sensitivity and uncertainty analysis.

Given the uncertainty in the internal and external variables of the building, the impact of uncertain parameters on the distribution of the output, which is the instantaneous ventilation rate, will be examined. Finally, SHAP sensitivity analysis is performed on the learned model to determine the impact of each feature on the ventilation rate. This study is valuable as it offers engineers deep insights into innovative building designs and aids in their development by identifying key parameters that affect ventilation rates.

2.1. Ventilation rate description

The datasets from different months reveal notable variations in reference and test data values, with the majority of the data centered around zero, as shown in Fig. 3. In September, 673 reference data entries were utilized, presenting an average ventilation rate of 6.87e-5 m/s and a maximum rate of 0.00015 m/s. Correspondingly, 673 test data points, with an average ventilation rate of 2.43e-5 m/s and a peak rate of 0.00013 m/s, were employed for model evaluation. For November, a total of 1345 reference data entries were used, showing an average ventilation rate of 7.81e-5 m/s and a maximum rate of 0.00016 m/s, while the 1345 test data points exhibited an average ventilation rate 5.08e-5 m/s with a maximum rate of 0.00017 m/s.

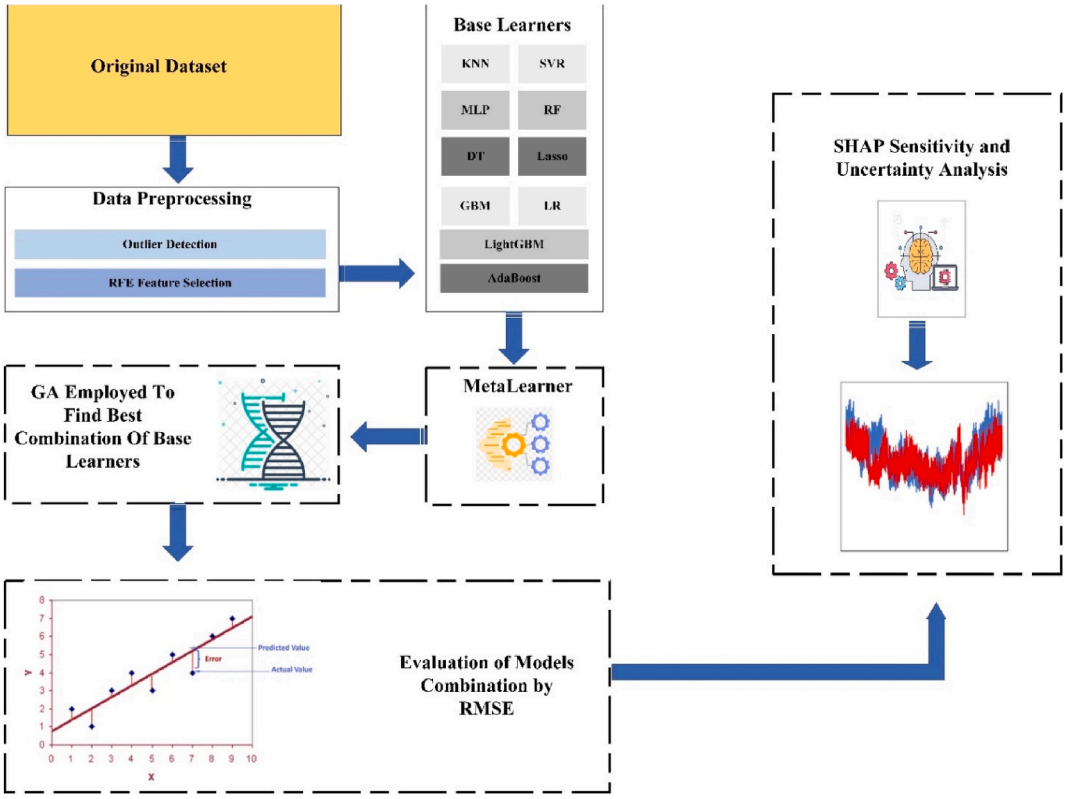


Fig. 2. Research workflow

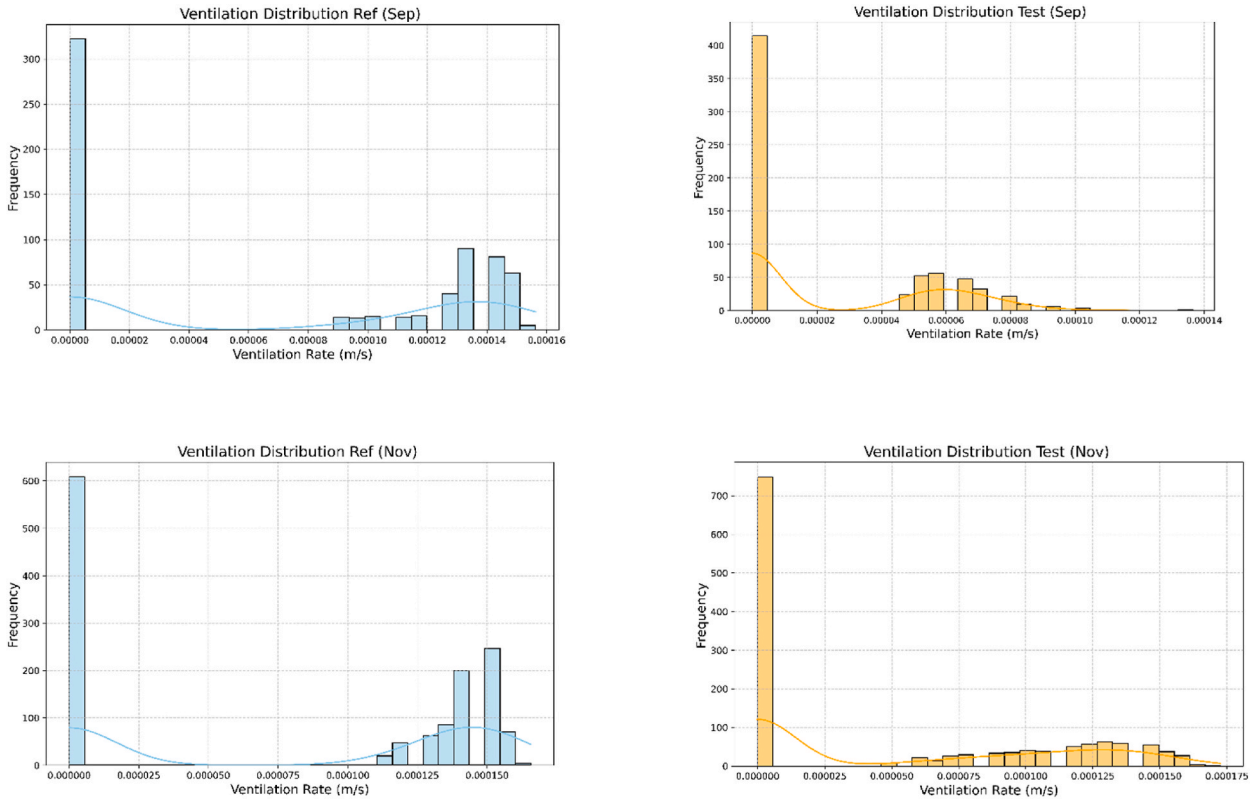


Fig. 3. Ventilation rate distribution analysis in reference and test buildings in September and November

It is also observed that where the data is non-zero, there is an imbalance in distribution, indicating that the dataset suffers from significant imbalances. This imbalance presents challenges in model training, as it intensifies the difficulty of accurately capturing the variations within the data.

2.2. *Recursive elimination feature selection*

Given the number of features in this study, ranging from 15 to 30, we were compelled to employ appropriate variable selection. To achieve this, we implemented the Recursive Feature Elimination (RFE) method to obtain a smaller subset of the dataset. This approach allows the model to identify complex patterns in the data with greater accuracy and speed [64,65]. Firstly, it trains a machine learning model on all available features. Then, these features are ranked based on their importance scores [66]. Afterward, the least important features are removed from the datasets. Subsequently, the model is retrained using the reduced set of features [67]. Finally, Steps 1 and 2 are iteratively executed until the desired number of features is obtained [68].

2.3. *Stacked ensemble model*

The stacked ensemble method is a model integration strategy that combines statistical theoretical knowledge with machine learning. This method includes several base models and a meta-model that fully leverages the advantages of each algorithm to achieve better predictive performance [60,69]. To ensure model diversity, we utilized a variety of base models, including K-NN, RF, DT, MLP, Lasso, GBM, LightGBM, SVR, LR, AdaBoost. Subsequently, a meta-model, specifically LR, was employed to integrate these base models. Initially, the stacking method trains single models using k-fold cross-validation on the training set at the base level. As shown in Fig. 4, for validation, the training data are divided into five equal parts, and in each computational run, one of the five parts is used for model validation. By repeating this process up to five times, the likelihood of encountering overfitting is minimized. Then, the outputs of the base models, as shown in Fig. 4, are used as new input data for the meta-model [70]. The new data are used to train the meta-model. When constructing a stacked ensemble model, it is crucial to consider both the accuracy and diversity of the base models. Stacked learning employs robust base models to ensure high accuracy. By utilizing various algorithms, the base models achieve diversity, which allows for a thorough examination of the internal relationship between indoor and outdoor conditions and the ventilation rate, thereby enhancing predictive performance. The meta-model is usually a well-stable model, combining different base models. The specific steps of the stacked ensemble algorithm are as follows:

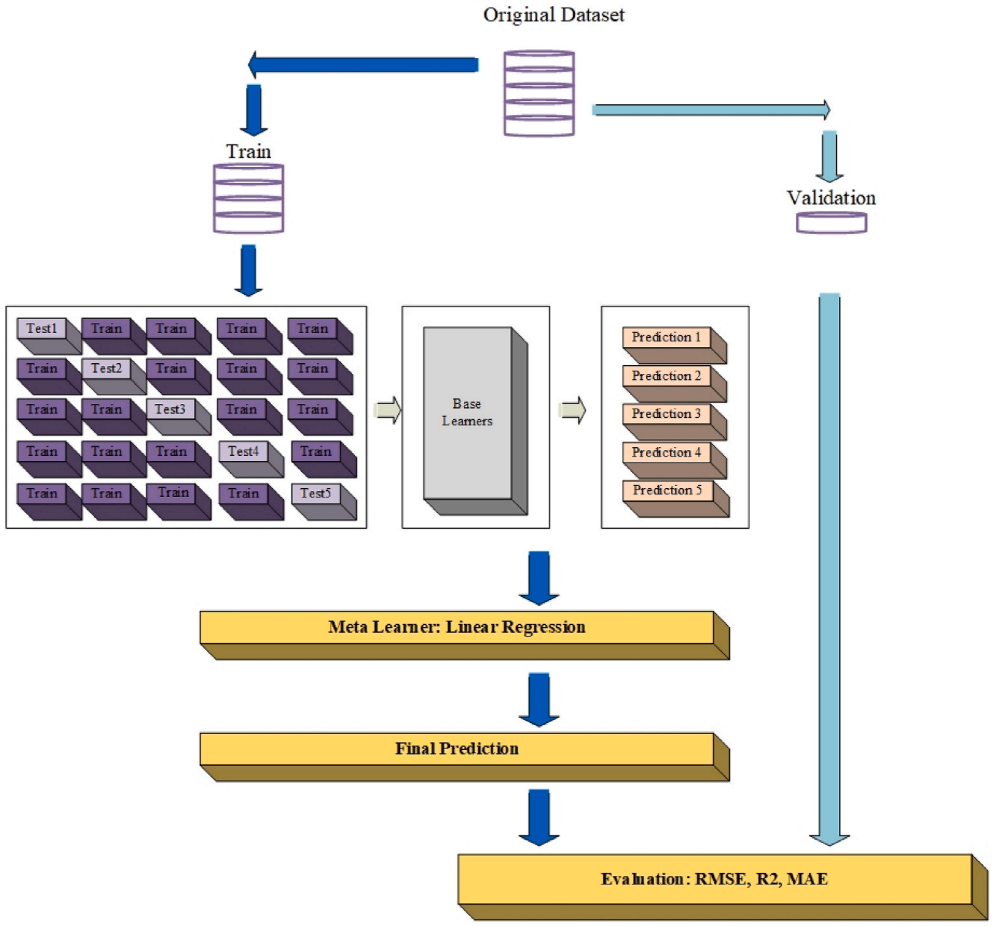


Fig. 4. Workflow of the stacked ensemble model

1. Sample data ($X_{n\times s}$), where 'n' is the sample number and 's' is the features number, and building ventilation rate is ($Y_{n\times 1}$), the target of the study. The datasets are divided into training and test sets [71,72].
2. For each base model, k-fold cross-validation is performed to avoid overfitting. The training set is randomly divided into five equal parts. Four out of the five parts are used to train the base models [73]. Fig. 3 illustrates the operation of the stacked ensemble model. Then, by predicting the base models, the outputs of the models are placed side by side in parallel to create new data for modeling. Meanwhile, the data are fed to the meta-model to make the final prediction, considering the most accurate initial predictions. In the final step, the test data set aside at the beginning is used for validation.

2.4. Improved stacked ensemble model

The ensemble learning method includes two primary steps: making base models and combinations them [74]. The selection of base models should have appropriate accuracy and diversity, while meta-learning models are usually simple and stable algorithms. Therefore, in the following study, linear regression has been used for meta-learning. However, the selection of base and meta-learning models directly affects the prediction results in the ensemble learning method [75,76]. Therefore, choosing the appropriate settings for model combination is very important.

In this study, a genetic algorithm was employed to determine the optimal combination of models within the ensemble learning framework. The process by which the genetic algorithm selects the appropriate model combination is illustrated in Fig. 5 [77,78]. To identify the optimal combination of initial models, the RMSE is defined as the objective function of the GA. This approach enhances the accuracy of the proposed model.

2.5. Evaluation parameters

The performance and accuracy of machine learning models in each case study will be evaluated using three criteria: Equation (1), the coefficient of determination (R^2); Equation (2), the mean absolute error (MAE); and Equation (3), the root mean square error

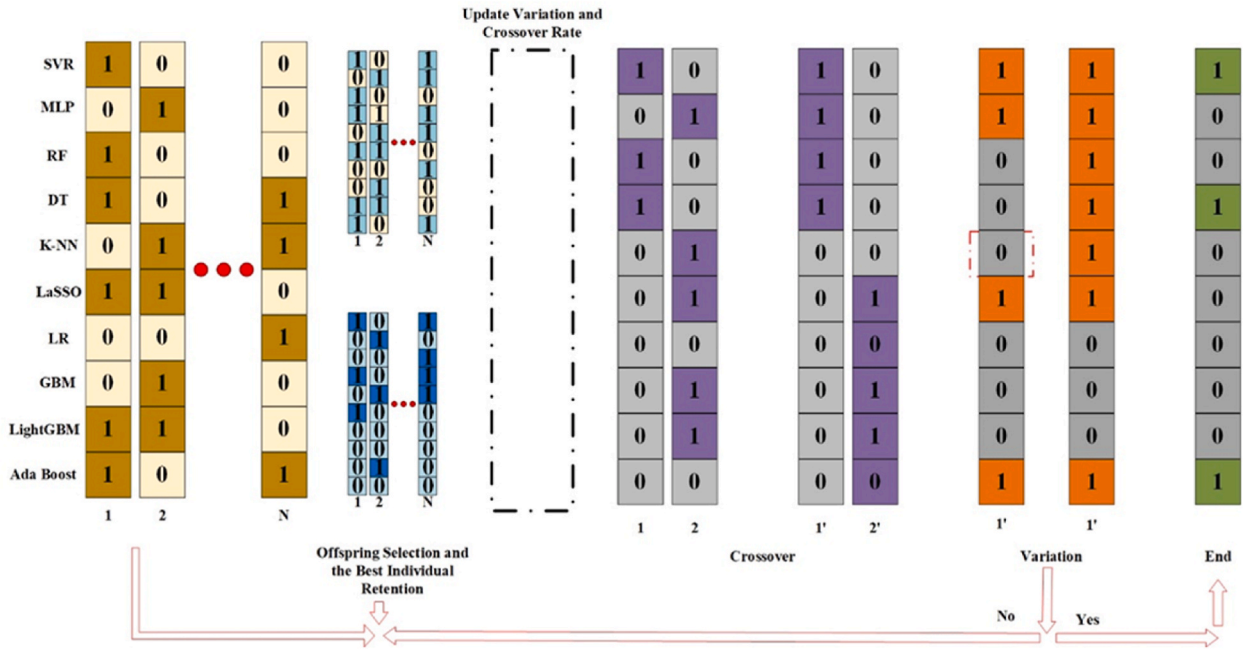


Fig. 5. Workflow of genetic algorithm to find the optimum combination of base learners

(RMSE) [79].

$$R^2 = 1 - \frac{\sum_{j=1}^{N} \left(Y_{j,a} - Y_{j,P}\right)^2}{\sum_{j=1}^{N} \left(Y_{j,a} - \overline{Y_a}\right)^2} \tag{1}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left|Y_{j,a} - Y_{j,p_i}\right| \times 100\% \tag{2}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(Y_{j,a} - Y_{j,p}\right)^2} \tag{3}$$

2.6. Uncertainty analysis

In this section, 1000 random samples were generated using the Latin Hypercube Sampling (LHS) method [80,81]. The data distribution for key parameters was then analyzed, accounting for recording uncertainties. External variables, such as air temperature, radiation level, and relative humidity, were incorporated. Initially, the input variables were classified into two categories: indoor and outdoor uncertain variables, to assess the influence of both internal and external factors on the distribution of ventilation rate data. The enhanced model subsequently predicts the ventilation rate, and these predictions are ranked in ascending order, focusing on the 95 % confidence interval. The results will reveal variations in the ventilation rate distribution and pinpoint the uncertainties that most significantly impact model output.

2.7. Ensemble-based sensitivity analysis

The SHAP method is based on game theory. This method can explain how the output of any machine learning model is interpreted by linearly adding a value to all input variables [82,83]. Therefore, this method is used for sensitivity analysis. The SHAP method is described as follows: According to Equation (4), A_0 can be considered a constant that plays a role in the equation. Then, another variable (A_i) is considered to be multiplied by each recorded data point to examine the change in the prediction made by the learned model. In this method, when the absolute value of the model's output is larger, it indicates a greater impact on the output [84]. During this calculation, each input is analyzed individually to study its impact precisely.

$$Y(x') = A0 + \sum_{i=1}^{N} Ai \times Xi = Y(X) \quad (4)$$

3. Result and discussion

3.1. RFE

RFE method is a widely used technique in machine learning for selecting essential features from a dataset. It works by initially training a model on the entire dataset and then iteratively removing less important features. This process helps in minimizing the input data while retaining the most significant features for the model's performance. The effectiveness of this method in reducing the dataset's complexity is demonstrated in Table 2, which highlights the selected features after applying RFE. The RFE method indicates that the number of features in both the reference and test data has been reduced by up to 40 %. Furthermore, analysis of the retained features in the reference data suggests that the ventilation rate in the reference building is more strongly influenced by internal factors. Specifically, 60 % of the retained features are related to internal conditions, including internal heat flux, thermal storage heat flux, internal temperature at a distance of one-tenth meter, internal temperature at a distance of two-tenths meter, incoming air temperature, and outgoing air temperature. In contrast, the retained features in the test data for November show an equal dependence on both internal and external factors.

Considering the increase in recorded features in the data provided by Fortin et al. [40], it can be observed that November exhibits similar conditions to September in terms of the importance of external condition features. As shown in Table 3, ambient temperature, atmospheric radiation, and effective temperature have a greater impact during this month. However, the extent of their influence will be further explored in the sensitivity analysis section. Table 2 demonstrates that the number of features for modeling in both datasets has been reduced to 10, suggesting that accurate and reliable modeling can be achieved with only around 30 % of the original data. Descriptions of the remaining features used as inputs for the ML models are provided in Table 3."

3.2. Standard stacked ensemble model

In this study, we aimed to utilize a stacked learning model to accurately and precisely model the ventilation rate in buildings based on internal and external parameters. To incorporate diverse models within the ensemble learning framework and capture complex patterns within the provided data, we employed ten classical models as base models. These base models include K-NN, RF, SVR, MLP, DT, Lasso, GBM, LR, LightGBM, and AdaBoost. The primary task of these models is to provide initial predictions for modeling. Subsequently, a linear regression model was used for the final prediction in the meta-model.

The results in Table 4 indicate that the standard stacked model significantly improved the performance of modeling in September for both the reference and test models. Table 4 shows that the R^2 value in the reference dataset reached 0.96 for the ensemble model. In contrast, the best classical model in the reference dataset reported an R^2 value of 0.955 for the LightGBM model. Although the improvement in the ensemble model is only 0.5 %, this improvement is significant due to the nature of the ventilation rate, which ranges between 0 and 0.0002. Based on other evaluation parameters, such as MAE and RMSE, the ensemble model has some weaknesses, as its error rate is higher than that of the gradient boosting model. The gradient boosting model has the lowest error in MAE and RMSE. According to Table 4, the MAE is 4.22e-06, and the RMSE is 1.15e-05, while for the ensemble model, the values are 5.2e-06 and 1.18e-05, respectively. However, based on the table, it can be seen that the MLP, SVR, and Lasso models showed the weakest performance.

Furthermore, in the test dataset (September), the result are slightly different. The result in Table 4 shows that the ensemble model performs better than other base algorithms in terms of R^2 and RMSE. The calculated results for these two parameters are 0.85 and 1.29e-05, respectively. These results indicate the presence of error even in the best-presented model. It means that complicating the modeling process is ineffective, and the amount of recorded data should be rose. It is noteworthy that the GBM model performed better in the MAE evaluation. Consistent with the reference results, the MLP, SVR, and Lasso models demonstrate notable deficiencies in achieving accurate modeling.

To assess the performance of the proposed model, it was tested on two additional datasets in November. The results, presented in Table 5, demonstrate that the ensemble model can accurately predict the ventilation rate with an R^2 accuracy of 0.963 in the reference dataset. Notably, the MAE and RMSE are calculated as 1.1e-5 and 5.9e-6, respectively. In contrast, for the GBM, these values are 3.93e-6 and 8.78e-6. This highlights the ensemble model's limitations in reducing these two parameters, a problem also observed in the previous two datasets. Furthermore, the MLP, SVR, and Lasso models exhibited the weakest performance, underscoring their lack of capability and efficiency in this domain.

The result in Table 5 shows that the GBM and decision tree-based models, like RF, have suitable capabilities for modeling. However,

Table 2
Number of initial and remained features.

Name	Initial (Sep)	Initial (Nov)	Selected (Sep)	Selected (Nov)	Reduction (Sep)	Reduction (Nov)
Reference	16	28	10	10	37.5 %	65 %
Test	15	26	10	10	0.33 %	62 %

Table 3
The selected feature description.

Name	Selected (Sep)	Selected (Nov)
Reference	Ambient Temperature	Ambient Temperature
	Wind Speed	Atmospheric Radiation
	Solar Radiation	Effective Sky Temperature
	Barometric Pressure	Thermal Storage Heat Flux
	Internal Heat Flux	Internal Wall Heat Flux
	Thermal Storage Heat Flux	Air Temperature at Bottom of Roof Cavity
	Internal Temperature at 0.1m	Internal Temperature at 0.1m
	Internal Temperature at 0.2m	Internal Temperature at 0.2m
	Inlet Air Temperature	Inlet Air Temperature
	Outlet Air Temperature	Outlet Air Temperature
Test	Ambient Temperature	Ambient Temperature
	Dew Point Temperature	Wind Speed
	Wind Speed	Gust Speed
	Wind Direction	Net Radiation
	Solar Radiation	Effective Sky Temperature
	Internal Heat Flux	Atmospheric Radiation
	External Heat Flux	Thermal Storage Surface Temperature
	Thermal Storage Heat Flux	Internal Wall Temperature
	Outlet Air Temperature	Internal Temperature at 0.2m
		Outlet Air Temperature

Table 4
Comparison of ML models and the standard stacked model (Sep).

Algorithm	Reference			Test		
	R^2	RMSE (m/s)	MAE	R^2	RMSE (m/s)	MAE
K-NN	0.91	1.17E-05	4.37E-06	0.83	1.36E-05	7.15E-06
RF	0.89	1.24E-05	4.45E-06	0.84	1.34E-05	6.99E-06
SVR	0.041	1.24E-05	6.62E-05	0.012	5.20E-05	0.043
MLP	0.165	0.077	0.118	0.005	0.057	0.0256
DT	0.87	1.41E-05	5.07E-06	0.66	1.97E-05	9.82E-06
Lasso	0.06	6.72E-05	6.60E-05	0.39	3.46E-05	3.33E-05
GMB	0.955	**1.15E-05**	**4.22E-06**	0.81	3.46E-05	**3.19E-06**
LR	0.931	1.38E-05	8.15E-06	0.78	1.59E-05	8.06E-06
LightGbm	0.956	1.16E-05	5.54E-06	0.81	1.46E-05	8.13E-06
AdaBoost	0.92	1.46E-05	6.68E-06	0.73	1.76E-05	1.30E-05
StackedEN	**0.96**	1.18E-05	5.20E-06	**0.85**	**1.29E-05**	7.14E-06

Table 5
Comparison of ML models and the standard stacked model (Nov).

Algorithm	Reference			Test		
	R^2	RMSE (m/s)	MAE	R^2	RMSE (m/s)	MAE
K-NN	0.92	1.38E-05	5.54E-06	0.95	1.30E-05	6.95E-06
RF	0.94	9.29E-06	4.06E-06	0.97	9.75E-06	4.80E-06
SVR	0.31	7.11E-05	6.96E-05	0.3	7.01E-05	6.39E-05
MLP	0.11	4.55E-02	3.20E-02	0.11	1.15E-01	8.63E-02
DT	0.925	1.33E-05	4.60E-06	0.95	1.25E-05	5.85E-06
Lasso	0.04	7.10E-05	7.01E-05	0.007	6.14E-05	5.81E-05
GMB	0.9578	**8.78E-06**	**3.93E-06**	0.959	1.20E-05	5.52E-06
LR	0.95	1.57E-05	9.75E-06	0.956	1.05E-05	5.52E-06
LightGbm	0.94	1.12E-05	5.42E-06	0.96	1.01E-05	5.52E-06
AdaBoost	0.9448	1.43E-05	7.87E-06	0.49	1.35E-05	1.14E-05
StackedEN	**0.963**	1.10E-05	5.90E-06	**0.973**	**9.63E-06**	**4.74E-06**

in modeling the test dataset, the ensemble model's results are more robust. The results in Table 5 show that the ensemble model excels in all three evaluation parameters compared to other models. The results presented in Table 5 for R^2, MAE, and RMSE are 0.973, 4.74e-6, and 9.63e-6, respectively. Even in the GBM and LightGBM models, lower values were estimated. These results indicate that the proposed modeling approach can provide acceptable results.

It is noteworthy that, as in the previous three datasets, the MLP, SVR, and Lasso models cannot model the ventilation rate precisely. This indicates that due to the limited data in modeling, such models cannot predict the ventilation rate accurately. Therefore, these models are not suitable for this domain. However, this does not preclude their inclusion in the base model, as even weaker models in an

ensemble can significantly contribute by capturing certain data patterns. To enhance the ensemble model, the next step involves selecting the correct and accurate combination of base models using GA.

3.3. GA-based stacked ensemble model

Based on the results presented in the previous section, it can be observed that the ensemble model suffers from performance improvement issues in key parameters, including MAE and RMSE. This can be attributed to errors in the base models. Therefore, the importance of optimization algorithms in this context becomes evident. To fulfill the research objectives and enhance the model, a GA was utilized to identify the optimal base models. The GA not only improves the evaluation parameters but also reduces the functional complexity of the ensemble model. A single-objective optimization approach was used to implement the GA. Thus, the RMSE parameter was considered the objective function of the optimization algorithm to provide the minimum RMSE by finding the best number of base learners.

Table 6 shows the variables of the genetic algorithm for implementing the optimization.

After employing the GA to the reference and test datasets, the correct combination of base models was revealed. The proposed method includes models such as K-NN, DT, MLP, GBM, and LightGbm. Despite the weak performance of the MLP model in all datasets, it significantly aids in modeling the ventilation rate in buildings. However, the test data results differ significantly, primarily due to the absence of the Random Forest (RF) model. This discrepancy highlights the inherent complexity of the test dataset, as established in the previous section. Consequently, despite the application of the GA, 9 out of the initial 10 models are retained for modeling.

The results presented in Table 7 indicate that selecting the optimal combination of base models significantly enhances the performance of the ensemble model. As shown in Table 7, the improved model surpasses the standard ensemble and GBM models, which were previously identified as the best models, across all three evaluation parameters. Specifically, the improved model increases the R^2 parameter by over one percent, achieving a value of 0.9717 in the reference dataset. Furthermore, there is a notable improvement in the RMSE and MAE parameters. The improved model not only demonstrates growth but also outperforms the previously best models, namely the ensemble and GBM models. In Table 7, the MAE and RMSE values are 4.13e-6 and 1.13e-5, respectively. Fig. 6A illustrates the high accuracy of the proposed model.

The proposed method also shows good performance in the test data. However, some errors in the test data can be attributed to the inherent imbalance in the dataset, mentioned in the ventilation rate description section. Despite this issue, the improved model substantially enhances the evaluation parameter results, demonstrating its robustness in handling the complexities of the test data. According to Table 7, the R^2 parameter has increased by 4 percent, reaching a value of 0.89, which demonstrates the accuracy and performance of the model. Additionally, there has been a significant improvement in the RMSE and MAE, with values of 1.12e-5 and 2.67e-6, respectively. It is noteworthy that the proposed method, due to its substantial improvement, can be applied to any dataset. Fig. 6B illustrates the predictions of the improved model compared to the actual data in the reference dataset for September.

For validation of the proposed method, the reference and test data were also examined in November to highlight the importance and robustness of the method in modeling. By employing the GA to select the most suitable base algorithms, the number of models in the reference data was reduced to 5. For modeling in November, the presence of K-NN, RF, LR, SVR, and DT models in the initial predictions provided acceptable results. It is noteworthy that the selection of the SVR model by the GA drew our attention, despite its considerable error. This indicates that a model should not be dismissed solely based on weaknesses in evaluation parameters, as it may still capture complex data behaviors. Consequently, the significance of the proposed method in this research is evident.

The results presented in Table 8 demonstrate that the improved ensemble model, which retains 5 out of the initial 10 base models, delivers precise and acceptable outcomes compared to the ensemble and the GBM model. In the reference dataset, the improved model achieved an excellent R^2 value of 0.983, representing nearly a two percent improvement. Additionally, the model attained very suitable values in the MAE and RMSE metrics, with error rates of 3.97e-6 and 9.11e-6, respectively. These are the lowest errors compared to other models. Fig. 7A illustrates the accuracy of the proposed model in correct detection.

The improvement in the test dataset by the optimized model was quite evident, retaining 8 out of the initial 10 base models. The R^2 value reached 0.988, RMSE was 8.67e-6, and MAE was 3.79e-6. The improvement in the first metric was more than one percent, and the error rates in the other two metrics showed a significant improvement. Indeed, the proposed model results illustrated that by increasing the amount of data, it can achieve outstanding results across all evaluation parameters. Fig. 7B illustrates the accuracy of the proposed model in correctly predicting the actual data.

Table 6
Defined genetic algorithm variables.

Variables	Data
Population Size	50
Number of Generations	10
Selection Strategy	Tournament Selection
Crossover Rate	0.5
Crossover Strategy	Two-Point Crossover
Mutation Rate	0.2
Mutation Strategy	Bit Flip (0.05)

Table 7
Comparison between enhanced ensemble model and the best models (September).

Algorithm	Reference			Test		
	R^2	RMSE (m/s)	MAE	R^2	RMSE (m/s)	MAE
GBM	0.955	1.15E-05	4.22E-06	0.81	3.46E-05	3.19E-06
Stacked EN	0.96	1.18E-05	5.20E-06	0.85	1.29E-05	7.14E-06
GA-Stacked	0.9717	1.13E-05	4.13E-06	0.89	1.12E-05	2.67E-06

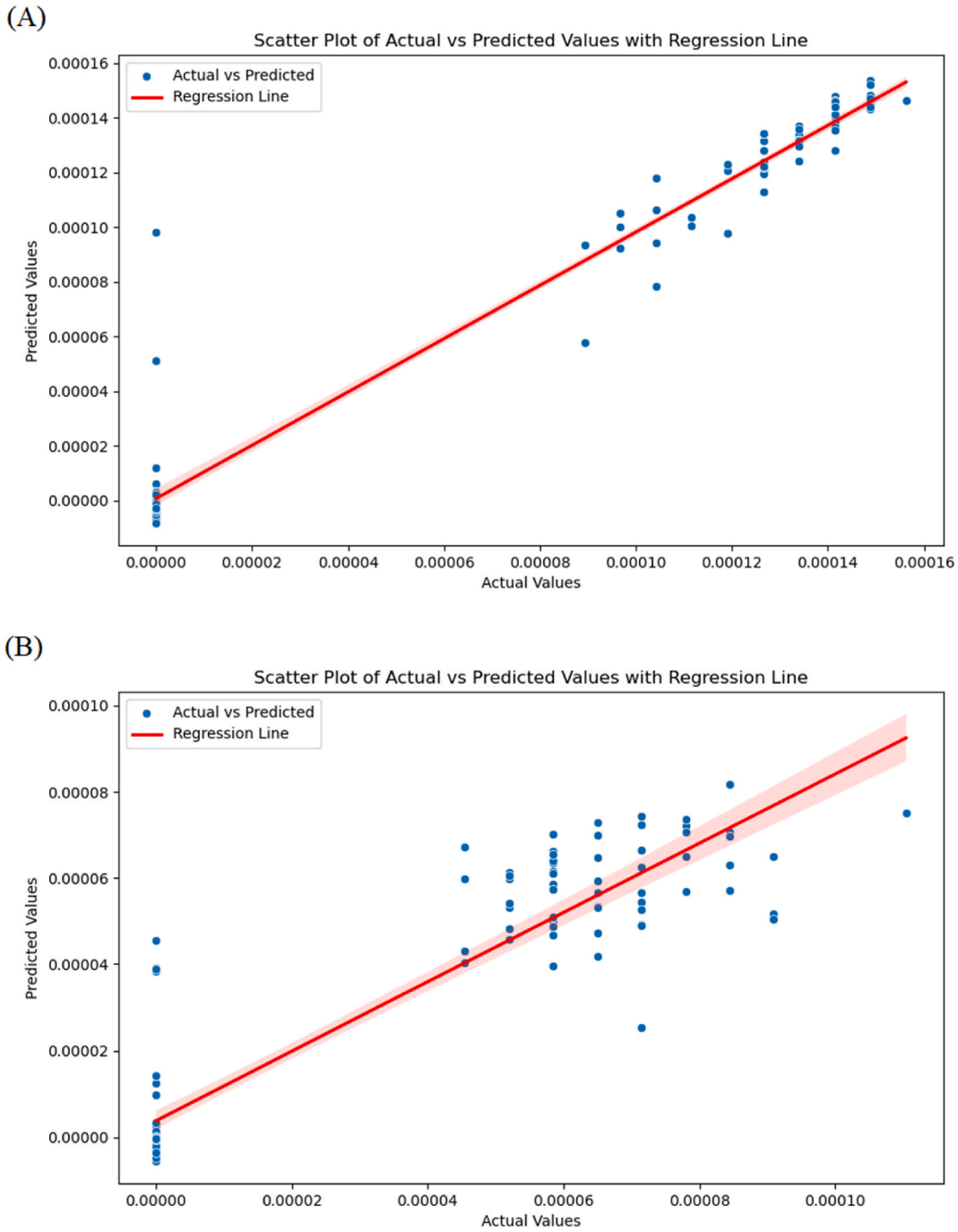


Fig. 6. Actual vs. Predicted Ventilation Rate (m/s) for November: Reference (A), Test (B).

Table 8
Comparison between enhanced ensemble model and the best models (November).

Algorithm	Reference			Test		
	R^2	RMSE (m/s)	MAE	R^2	RMSE (m/s)	MAE
GBM	0.955	1.15E-05	4.22E-06	0.957	8.78E-06	3.93E-06
Stacked EN	0.963	1.35E-05	5.90E-06	0.973	9.91E-06	4.74E-06
GA-Stacked EN	0.983	9.11E-06	3.97E-06	0.988	8.67E-06	3.79E-06

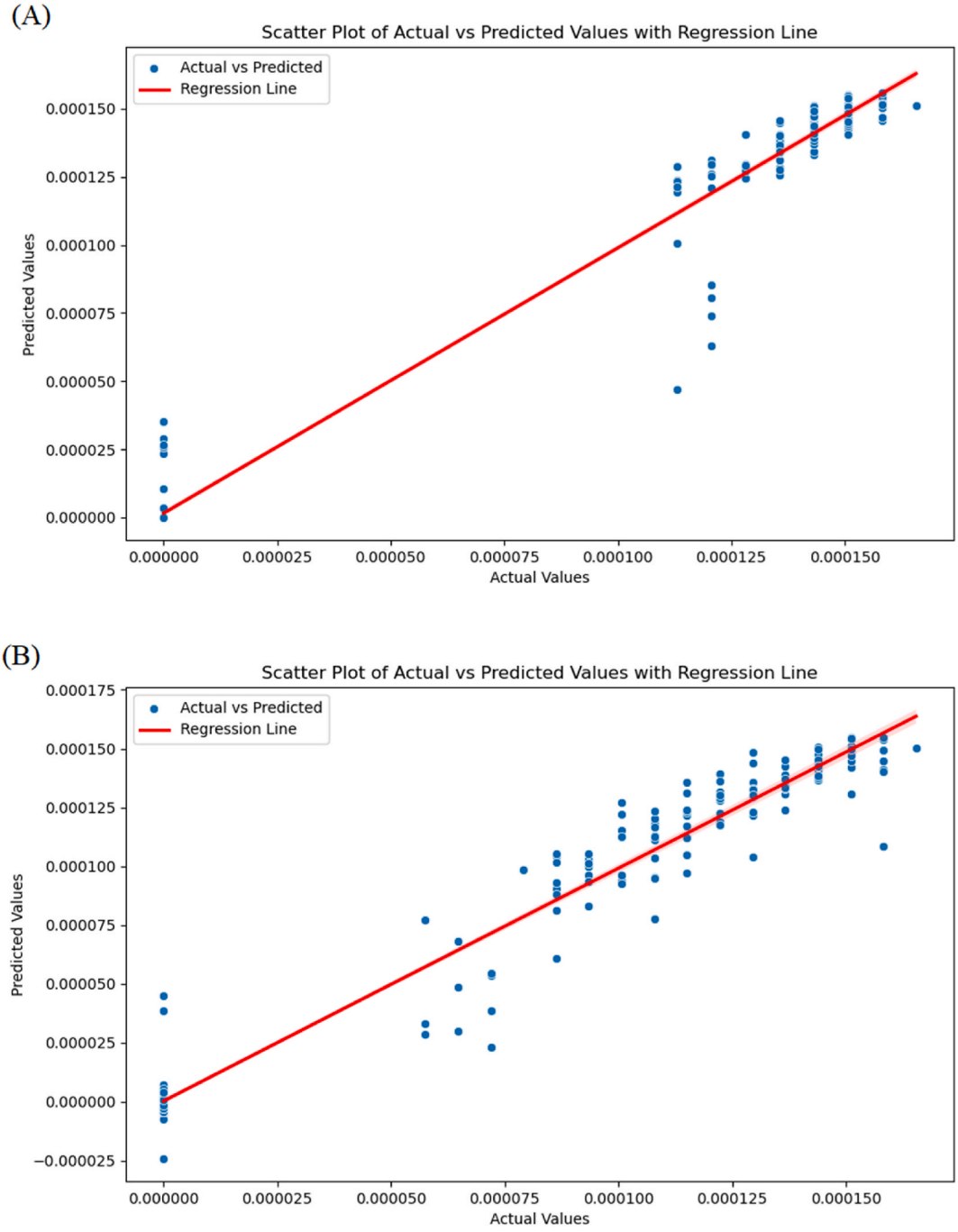


Fig. 7. Actual vs. Predicted Ventilation Rate (m/s) for November: Reference (A), Test (B).

3.4. Uncertainty analysis

Uncertainty analysis is a crucial tool in evaluating and managing the risks and uncertainties inherent in models and predictions. This analysis helps us assess the accuracy and reliability of model results, thereby preventing incorrect decision-making. Typically, statistical and probabilistic methods are employed in uncertainty analysis to predict various possible outcomes and examine the impact of various factors on the final results.

In this research, following the development of an improved model, uncertainty analysis is conducted in the model's input parameters. Considering that climatic data and data recording devices are inherently subject to uncertainties, the application of this method is deemed essential and significant.

Table 9 presents the distribution, mean, minimum, and maximum values of each feature in the reference dataset for the month of September. Features of the other datasets, used for uncertainty analysis, are documented in Appendix A. To facilitate a more precise and appropriate analysis, the parameters with uncertainties were categorized into indoor and outdoor variables. This categorization was employed to examine the impact of input parameter uncertainties on the standard deviation of the ventilation rate.

The results in Table 10 indicate that the standard deviation of the ventilation rate in September in the reference dataset is 6.68e-05. After generating 1000 datasets with identical distributions in the input parameters, the results extracted from the improved model show that the uncertainties in internal and external parameters generally have a minimal impact on the ventilation rate. The results suggest that despite the uncertainties in internal and external conditions, the data distribution will ultimately be affected by up to 2 %.

In contrast, the test dataset reveals a distinct impact of parameters with uncertainties. As shown in Table 10, uncertainties in internal variables, such as the heat flux from the top of the radiator, internal heat flux of the radiator, internal heat flux, and the outlet temperature from the building, significantly affect the ventilation rate. The results indicate that, despite these uncertainties, it is possible to reduce the standard deviation of the ventilation rate by up to 11 %. This suggests that accurate recording and precise control of these parameters can substantially influence the building's ventilation rate.

Additionally, as indicated by the results in Table 10, the impact of uncertainties in environmental parameters is estimated to be 2.91e-05. Consequently, the presence of uncertainty can reduce the distribution in the ventilation rate data by less than 7 %. This suggests that the innovative building is more sensitive to uncertainties in the model's input parameters.

The results presented in Table 11 indicate that the ventilation rate in the reference building for November is predominantly influenced by internal variables. These variables include the heat flux of the thermal storage, the heat flux of the internal wall, the air temperature at the bottom of the convective flow guard, the indoor temperature at distances of 0.1 m and 0.2 m, as well as the air temperatures entering and exiting the building. These findings corroborate the results of feature selection conducted using the RFE.

The results indicate that, despite the uncertainties in environmental variables, it is possible to reduce the distribution of the ventilation rate by approximately 1 %. In contrast, uncertainties in internal parameters can reduce this distribution by up to 2.5 %. This behavior was similarly observed in the test dataset. The findings suggest that uncertainties in internal parameters can lead to a reduction in the distribution of the ventilation rate by up to 2 % throughout the day. According to Table 11, with these reductions, the standard deviation for internal and environmental variables is estimated to be 5.84e-05 and 5.90e-05, respectively.

3.5. Sensitivity analysis

Sensitivity analysis using SHAP is an advanced method for evaluating the impact of input variables on the output of machine learning models. This method, grounded in game theory and the concept of SHAP values, enables the precise calculation of each input variable's contribution to the final result.

The results depicted in Fig. 8A indicate that the inlet air temperature of the reference building, as well as the internal temperatures at 0.2 m and 0.1 m, have the most significant impact on ventilation rates. The findings suggest that reducing the inlet air temperature of the reference building in September can increase the ventilation rate. Conversely, increasing the inlet air temperature decreases the ventilation rate.

Furthermore, the internal temperatures at 0.1 and 0.2 m significantly affect the ventilation rate. Notably, there is an inverse relationship between these two parameters: increasing the temperature at 0.2 m increases the ventilation rate, while increasing the temperature at 0.1 m decreases it. This suggests the presence of natural convection flow, which can enhance the ventilation rate,

Table 9
Statical description of inputs.

Category	Parameter	Unit	Distribution	Minimum	Maximum	Mean	Standard Deviation
Outdoor Variables	Ambient Temperature	(C°)	Normal	7.28	38.93	21.91	9.33
	Wind Speed	(m/s)	Normal	0.1	1.42	0.44	0.32
	Solar Radiation	(W/m^2)	Normal	1.91	191.99	54.71	70.57
	Barometric Pressure	(mbar)	Normal	981.76	988.99	985.51	1.67
Indoor Variables	Internal Heat Flux	(W/m^2)	Normal	−29.73	28.31	−9.25	15.49
	Thermal Storage Heat Flux	(W/m^2)	Normal	−33.72	15.57	−3.49	12.37
	Indoor Temperature at 0.1m Distance	(C°)	Normal	20.52	32.23	25.84	2.97
	Indoor Temperature at 0.2m Distance	(C°)	Normal	21.17	32.98	26.48	2.99
	Air Temperature Entering the Building	(C°)	Normal	9.2	40.5	23.98	9.15
	Air Temperature Exiting the Building	(C°)	Normal	20.08	43.84	30.07	5.96

Table 10
Uncertainty analysis of indoor and outdoor variables (September).

Data	Parameter	Actual Std	New Std	Impact on Standard Deviation (%)
Reference	Outdoor Variables	6.68 e−05	6.55 e−05	−2.002
	Indoor Variables	6.68 e−05	6.60 e−05	−1.28
Test	Outdoor Variables	3.12 e−05	2.91 e−05	−6.85
	Indoor Variables	3.12 e−05	2.78 e−05	−11.11

Table 11
Uncertainty analysis of indoor and outdoor variables (November).

Data	Parameter	Actual Std	New Std	Impact on Standard Deviation (%)
Reference	Outdoor Variables	7.17e-05	7.11 e−05	−0.75
	Indoor Variables	7.17e-05	6.99 e−05	−2.44
Test	Outdoor Variables	5.96 e−05	5.90 e−05	−1.05
	Indoor Variables	5.96 e−05	5.84 e−05	−1.97

particularly since the ventilation system is installed in the roof of the building. The results for the outlet air temperature in the reference building corroborate this finding. Additionally, solar radiation positively impacts the ventilation rate. As shown in Fig. 8A, controlling solar radiation can enhance ventilation and natural cooling.

In contrast, the test building in September is increasingly influenced by ambient temperature. Due to the building's efficiency, the ventilation rate rises with the increase in ambient temperature. Solar radiation and gust speed also positively impact the building's ventilation, indicating its suitability for hot and dry environments, as it can facilitate ventilation through pressure differences. The second most influential parameter on the ventilation rate in the test building is the outlet air temperature. The results in Fig. 8B show that a decrease in temperature can potentially increase ventilation. It is worth noting that other input parameters, except for the thermal storage heat flux, have minimal impact on the ventilation rate. Given the importance of heat flux, there is potential to expand this part of the research, considering that the thermal storage in the provided data is assumed to be water.

Analysis of the SHAP method results for the November datasets indicates that the reference building is influenced by external temperature during this month. An increase in external and inlet temperatures leads to a decrease in the ventilation rate. However, it is important to note that the ventilation rate is dependent on several other parameters, as illustrated in Fig. 9A. The internal temperatures at 0.1 m and 0.2 m exhibit an inverse relationship, corroborating the findings from the previous section. Additionally, the heat flux in the internal walls and thermal storage significantly impacts the ventilation rate. Specifically, increasing the heat flux in these two parameters can enhance ventilation. Furthermore, as shown in Fig. 9A, other parameters such as effective sky temperature and outlet air temperature have minimal impact on the ventilation rate.

According to Fig. 9B, the test building, unlike the reference building, relies on fewer parameters. The results indicate that an increase in temperature leads to an improvement in the building's ventilation. Comparing the reference and test buildings, creating a temperature difference is an effective method for enhancing ventilation. Therefore, modeling internal temperature in future research would be beneficial. The second most important parameter in the test data is the net radiation emitted on the roof of the building. An increase in this parameter enhances the ventilation rate. In other words, greater radiation reflection on the roof positively impacts the ventilation rate. The results presented in Fig. 9B regarding environmental radiation confirm that a decrease in environmental radiation can improve ventilation, although this improvement is slight compared to the other two parameters.

4. Conclusion

Considering the global climate changes, the demand for air conditioning systems has significantly increased, while the greenhouse gas emissions associated with these systems, including refrigerants, exacerbate global warming. Therefore, employing natural cooling in buildings plays a crucial role in mitigating the problems caused by air conditioning systems. In this research, using the published data [40], the building was modeled by employing the enhanced machine learning model in order to identify the most relevant feature. To reduce the issues of black-box modeling, uncertainty and sensitivity analysis based on machine learning was developed.

The results of this research can be summarized in three points:

1. The improved ensemble stacking model significantly improved the results of the machine learning model in all three parameters, consisting of R^2, RMSE, and MAE. By using the proposed model, the power and accuracy of the method in all data sets have been demonstrated. This means that the proposed method can be applied to all black-box problems.
2. Given the uncertainty in the internal and external environmental conditions in the recorded data sets, for the first time, uncertainty analysis based on the improved model was employed. The results show that the uncertainty in internal and external parameters has a minimal impact on the ventilation rate in the reference model. On the other hand, in September, the test building shows substantial impact of uncertainty on the distribution of ventilation rate data. Therefore, the sensitivity of the test building in September to uncertainty is higher than that of the reference building and even the test building in November.

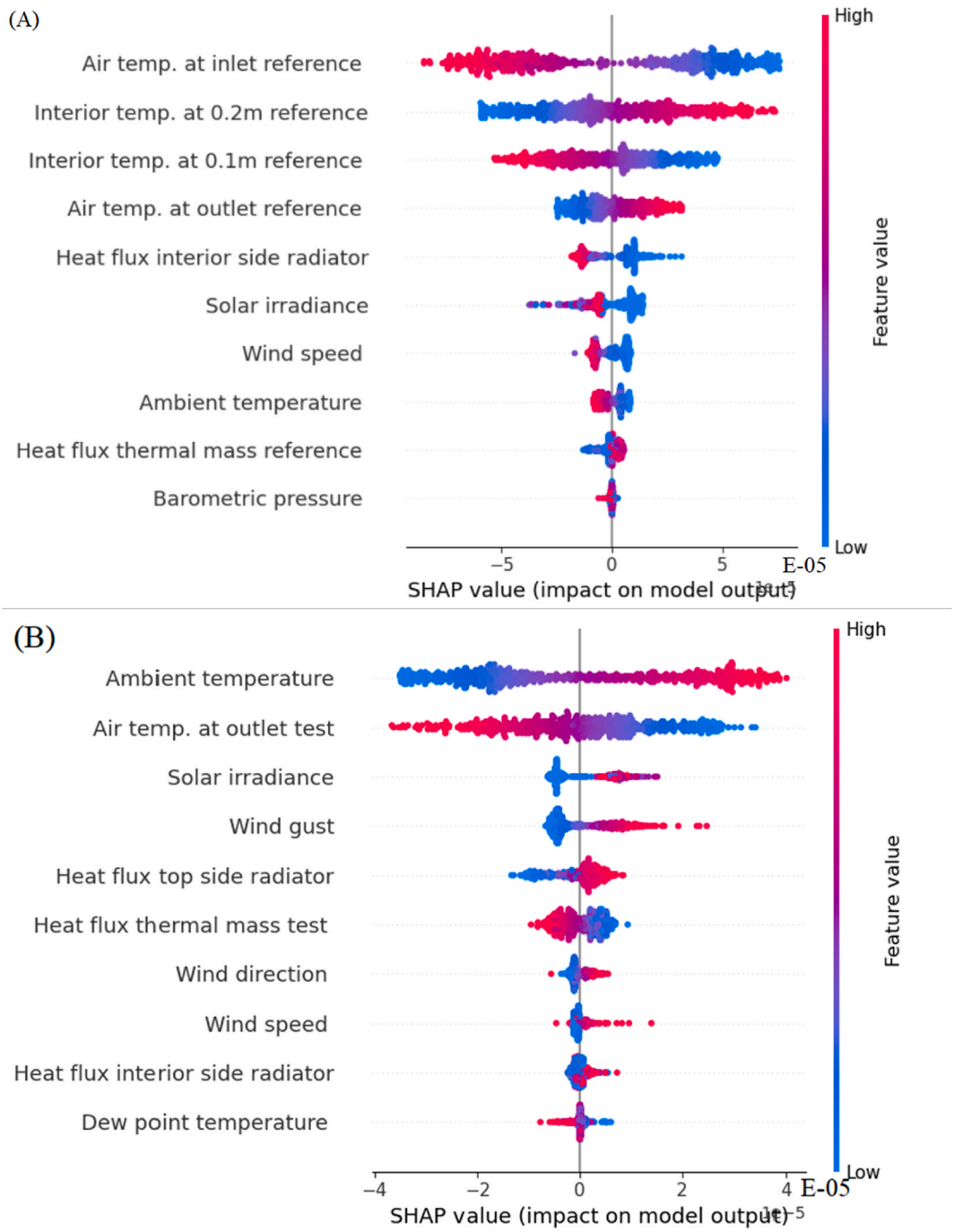
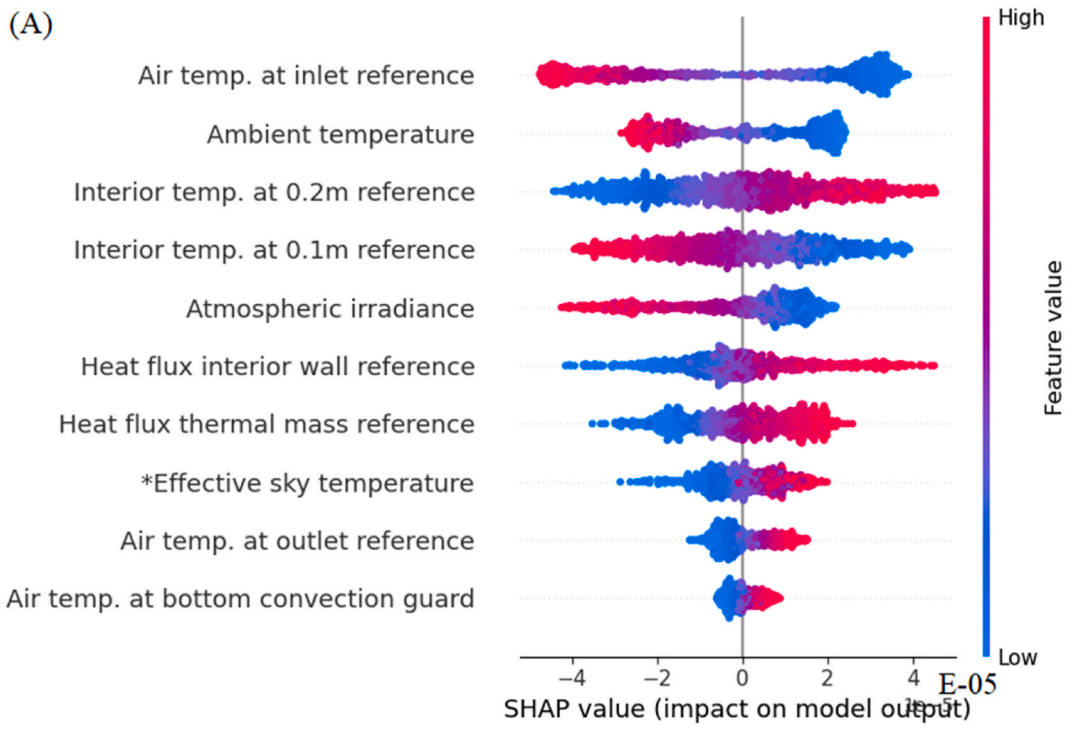


Fig. 8. Machine learning-based SHAP sensitivity analysis (September): Reference (A), test (B)

3. Using sensitivity analysis based on the improved model, it can be concluded that the reference model in both examined months is highly dependent on the inlet air temperature, internal temperatures at 0.1 and 0.2 m, and internal wall heat flux. Optimizing the control of these parameters significantly impacts the development of this building. It is worth noting that, given the special importance of the test building, the significance of input parameters based on the improved model was also examined. The findings indicate that the test building, during the two considered periods, depends on fewer parameters. According to the results in the previous section, the test building is influenced by outdoor conditions, outlet air temperature, and net radiation from the roof. A notable point in these results is that an increase in ambient temperature will enhance the ventilation.



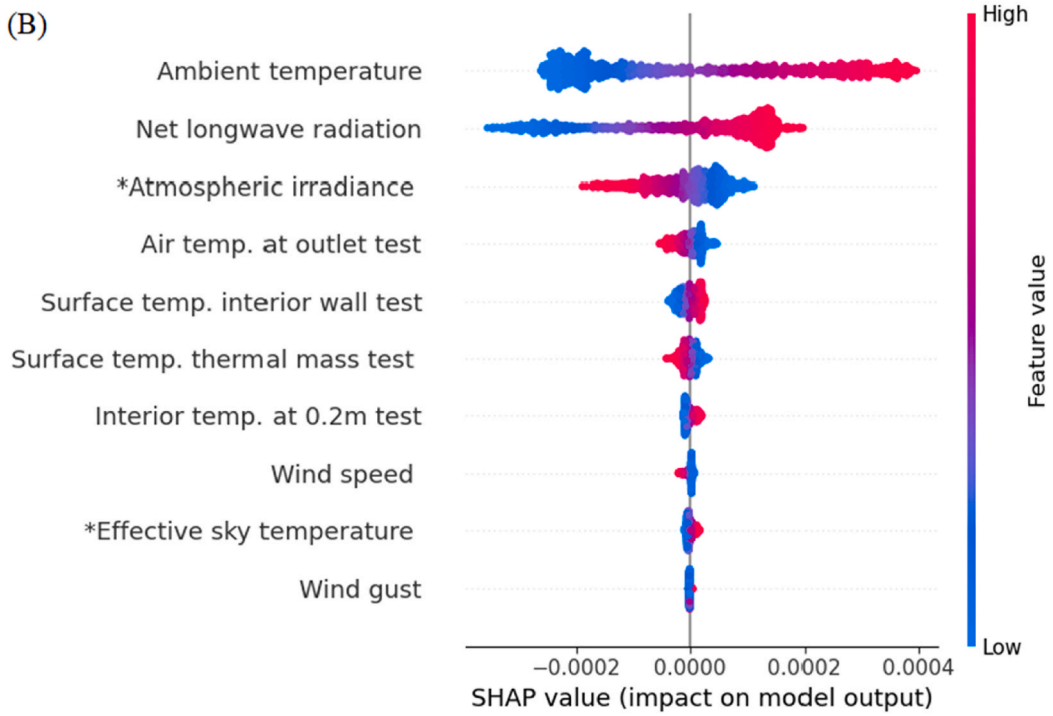


Fig. 9. Machine learning-based SHAP sensitivity analysis (November): Reference (A), test (B)

5. Research limitations, challenges and future prospects

This study has several limitations, notably the imbalance in the datasets used, which include instances with near-zero ventilation rates. This imbalance can lead to significant errors in the models, as the lack of sufficient data for these extreme conditions hinders the model's ability to make accurate predictions. Unfortunately, the necessary data to address this imbalance is not available in this study, posing a challenge to model performance. A potential avenue for future research is to explore techniques for handling data imbalance, such as data augmentation, synthetic data generation, or advanced balancing algorithms, particularly in cases where access to comprehensive datasets is limited.

Additionally, this study does not consider occupant behavior or the impact of harsh weather conditions, both of which can significantly influence ventilation rates and cooling efficiency in passive cooling buildings. These factors, along with varying human interactions with building systems, create a dynamic environment that may affect the model's accuracy and applicability. Furthermore, the study does not assess the feasibility of passive cooling systems across diverse environmental conditions globally, which is crucial for understanding how such systems perform in various climates. Future research could expand on these limitations by incorporating occupant behavior, exploring the effects of extreme weather conditions, and evaluating the feasibility and performance of passive cooling buildings in different geographical regions to provide more generalizable and practical insights.

CRediT authorship contribution statement

Majid Mohsenpour: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mohsen Salimi:** Writing – review & editing, Validation, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Atieh Kermani:** Writing – review & editing, Writing – original draft, Investigation. **Majid Amidpour:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Data availability statement

The data supporting this study is sourced from a published dataset, accessible at https://doi.org/10.5281/zenodo.8164886.

Ethics declaration

Review and/or approval by an ethics committee was not required for this study, as it did not involve any experiments, interactions with humans, or use of animal subjects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Table A1
Statical Description of Reference (November)

Category	Parameter	Unit	Distribution	Minimum	Maximum	Mean	Standard Deviation
Outdoor Variables	Ambient Temperature	(C°)	Normal	−0.96	24.82	9.42	8.32
	Atmospheric Radiation	(W/m^2)	Normal	226.98	295.47	255.27	12.02
	Effective Sky Temperature	(W/m^2)	Normal	−21.63	−4.47	−14.16	3.05
Indoor Variables	Thermal Storage Heat Flux	(W/m^2)	Normal	−6.76	5.39	−0.185	3.46
	Internal Wall Heat Flux	(W/m^2)	Normal	−1.38	1.01	−0.31	0.572
	Air Temperature at Bottom of Convection Shield	(C°)	Normal	−5.61	22.17	4.86	8.34
	Indoor Temperature at 0.1m Distance	(C°)	Normal	11.021	17.78	14.024	1.43
	Indoor Temperature at 0.2m Distance	(C°)	Normal	11.58	18.28	14.59	1.46
	Inlet Air Temperature to Building	(C°)	Normal	0.48	28.69	11.9	9.274
	Outlet Air Temperature from Building	(C°)	Normal	10.62	33.79	18.4	6.34

Table A2
Statical Description of Test (September)

Category	Parameter	Unit	Distribution	Minimum	Maximum	Mean	Standard Deviation
Outdoor Conditions	Ambient Temperature	(C°)	Normal	7.28	38.93	21.91	9.33
	Dew Point Temperature	(C°)	Normal	−3.41	13.99	6.9	3.64
	Wind Speed	(m/s)	Normal	0.1	1.42	0.441	0.326
	Wind Direction	(−)	Normal	0.42	359.29	181.91	87.99
	Instantaneous Wind Speed	(m/s)	Normal	0.15	2.9	0.77	0.58
	Solar Radiation	(W/m^2)	Normal	1.91	191.99	54.71	70.57
Indoor Conditions	Heat Flux Above Radiator	(W/m^2)	Normal	−53.6	56.07	15.54	31.48
	Internal Radiator Heat Flux	(W/m^2)	Normal	−29.73	28.31	−9.25	15.49
	Internal Heat Flux	(W/m^2)	Normal	−11.76	15.02	−1.14	5.165
	Outlet Temperature from Building	(C°)	Normal	9.8	35.01	20.5	5.93

Table A3
Statical Description of Test (November)

Category	Parameter	Unit	Distribution	Minimum	Maximum	Mean	Standard Deviation
Outdoor Variables	Ambient Temperature	(C°)	Normal	−0.96	24.82	9.42	8.32
	Wind Speed	(m/s)	Normal	0.11	2.03	0.5	0.37
	Instantaneous Wind Speed	(m/s)	Normal	0.2	4.79	0.92	0.76
	Net Radiation	(W/m^2)	Normal	−198.93	−57.17	−108.41	40.8
	Atmospheric Radiation	(W/m^2)	Normal	226.9	295.4	255.27	12.2
	Effective Sky Temperature	(C°)	Normal	−21.61	−4.47	−14.16	3.05
Indoor Variables	Thermal Storage Surface Temperature	(C°)	Normal	0.28	10.83	5.83	2.01
	Internal Wall Surface Temperature	(W/m^2)	Normal	−12.12	1.24	−5.9	3.013
	Indoor Temperature at 0.2m Distance	(C°)	Normal	−0.39	13.56	5.92	3.61
	Outlet Temperature from Building	(C°)	Normal	0.08	14.8	6.71	3.51

Table A4
Hyperparameter Configurations of Base Learners for Models in Reference and Test Buildings (September)

Algorithm	Model 1 (Reference)	Model 2 (Test)
KNN	n_neighbors = 5, p = 2, weight = 'uniform'	n_neighbors = 5, p = 2, weight = 'uniform'
RF	n_estimators = 200, random_state = 42	n_estimators = 350, random_state = 42
XGB	n_estimators = 200, random_state = 42, learning_rate = 1.1	n_estimators = 200, random_state = 42, learning_rate = 1.1
LR	Default	Default
SVR	kernel = 'sigmoid', C = 0.15	kernel = 'sigmoid', C = 1
DT	max_depth = 100	max_depth = 45
MLP	hidden layer = 2, neuron= (75, 25), random_state = 42, max_iter = 500	hidden layer = 1, neuron = (80), random_state = 42, max_iter = 500
Adaboost	n_estimators = 250, learning_rate = 0.9 random_state = 42	n_estimators = 135, learning_rate = 0.85, random_state = 42
GBM	n_estimators = 100, learning_rate = 0.9 random_state = 42	n_estimators = 150, learning_rate = 1 random_state = 42
LightGBM	n_estimators = 300, learning_rate = 0.6, random_state = 42	n_estimators = 250, learning_rate = 0.78, random_state = 42
Lasso	alpha = 1.0, random_state = 42	alpha = 2.0, random_state = 42

Table A5
Hyperparameter Configurations of Base Learners for Models in Reference and Test Buildings (November)

Algorithm	Model 3 (Reference)	Model 4 (Test)
KNN	n_neighbors = 4, p = 1, weight = 'uniform'	n_neighbors = 5, p = 2, weight = 'uniform'
RF	n_estimators = 150, random_state = 42	n_estimators = 300, random_state = 42
XGB	n_estimators = 200, random_state = 42, learning_rate = 0.1	n_estimators = 200, random_state = 42, learning_rate = 0.25
LR	Default	Default
SVR	kernel = 'sigmoid',	kernel = 'sigmoid', C = 0.85
DT	max_depth = 120	max_depth = 60
MLP	hidden layer = 2, neuron= (25, 25), random_state = 42, max_iter = 300	hidden layer = 2, neuron= (110, 10), random_state = 42, max_iter = 300
Adaboost	n_estimators = 250, learning_rate = 1 random_state = 42	n_estimators = 135, learning_rate = 1.1, random_state = 42
GBM	learning_rate = 0.7, random_state = 42	n_estimators = 150, learning_rate = 1 random_state = 42
LightGBM	n_estimators = 250, learning_rate = 1, random_state = 42	n_estimators = 100, learning_rate = 0.8, random_state = 42
Lasso	alpha = 1.0, random_state = 42	random_state = 42

References

[1] S. Vashisht, D. Rakshit, Recent advances and sustainable solutions in automobile air conditioning systems, J. Clean. Prod. 329 (Dec. 2021) 129754, https://doi.org/10.1016/j.jclepro.2021.129754.

[2] K. Sundaram, K.R. Sri Preethaa, Y. Natarajan, A. Muthuramalingam, A.A.Y. Ali, Advancing building energy efficiency: a deep learning approach to early-stage prediction of residential electric consumption, Energy Rep. 12 (Dec. 2024) 1281–1292, https://doi.org/10.1016/j.egyr.2024.07.034.

[3] H. Wang, X. Chen, N. Vital, E. Duffy, A. Razi, Energy optimization for HVAC systems in multi-VAV open offices: a deep reinforcement learning approach, Appl. Energy 356 (Feb. 2024) 122354, https://doi.org/10.1016/j.apenergy.2023.122354.
[4] IEA, Global air conditioner stock, 1990-2050, IEA, Paris https://www.iea.org/data-and-statistics/charts/global-air-conditioner-stock-1990-2050, Licence: CC BY 4.0.
[5] S. Liu, W. Ge, X. Meng, Influence of the shading nets on indoor thermal environment and air-conditioning energy consumption in lightweight buildings, Energy Rep. 11 (Jun. 2024) 4515–4521, https://doi.org/10.1016/j.egyr.2024.04.032.
[6] W.S.-H. Kaghembega, S. Chen, A. Tchewafei, K.B.F. Lionel, Paper: modeling and scenario analysis of residential building energy conservation in cities of different weather, Energy Rep. 11 (Jun. 2024) 2670–2684, https://doi.org/10.1016/j.egyr.2024.02.025.
[7] M. Hernandez-Chavela, F.R. Ceja-Soto, Á. Marroquín de Jesús, El tapanco en viviendas vernáculas de Querétaro, México. Uso y eficiencia higrotérmica, Rev. Hábitat Sustentable 13 (1) (Jun. 2023) 108–119, https://doi.org/10.22320/07190700.2023.13.01.09.
[8] S. Dwivedi, Z. Parveen, S. Tiwari, U. Porov, Peltier-based air conditioning system, Int. J. Res. Appl. Sci. Eng. Technol. 11 (5) (May 2023) 5963–5966, https://doi.org/10.22214/ijraset.2023.53061.
[9] S. Lan, Deprizon, N. Razmjooy, Enhancing the performance of zero energy buildings with boosted coyote optimization and elman neural networks, Energy Rep. 11 (Jun. 2024) 5214–5226, https://doi.org/10.1016/j.egyr.2024.05.001.
[10] A. Myat, Application of artificial intelligence in air conditioning systems, in: Recent Updates in HVAC Systems, IntechOpen, 2023, https://doi.org/10.5772/intechopen.107379.
[11] Z. Sheng, Development of Climate Friendly Cooling Technologies: Trends and Driving Factors, 1990–2019, Apr. 2023, https://doi.org/10.56506/SRVA2054.
[12] M. Zafaranchi, H. Sozer, Enhancing energy efficiency through hourly assessments of passive interventions in educational-office buildings: a case study in a Mediterranean climate, Energy Rep. 11 (Jun. 2024) 423–441, https://doi.org/10.1016/j.egyr.2023.12.016.
[13] Y. Song, K.S. Darani, A.I. Khdair, G. Abu-Rumman, R. Kalbasi, A review on conventional passive cooling methods applicable to arid and warm climates considering economic cost and efficiency analysis in resource-based cities, Energy Rep. 7 (Nov. 2021) 2784–2820, https://doi.org/10.1016/j.egyr.2021.04.056.
[14] E. Manyumbu, V. Martin, J.N. Chiu, Prospective PCM–desiccant combination with solar-assisted regeneration for the indoor comfort control of an office in a warm and humid climate—a numerical study, Energies 16 (14) (Jul. 2023) 5391, https://doi.org/10.3390/en16145391.
[15] M.Q. Oleiwi, M.K.A.M. Sulaiman, M.F. Mohamed, Passive Cooling Strategies in the hot climate: a review study, Arid Int. J. Sci. Technol. (Jun. 2023) 76–106, https://doi.org/10.36772/arid.aijst.2023.6115.
[16] A.S. Farooq, K. Alkaabi, S.B. Hdhaiba, Exploring radiative sky cooling resource map and the impact of meteorological conditions on radiative emitters. A perspective of GCC countries, Energy Rep. 10 (Nov. 2023) 473–483, https://doi.org/10.1016/j.egyr.2023.06.054.
[17] M. Zeyghami, D.Y. Goswami, E. Stefanakos, A review of clear sky radiative cooling developments and applications in renewable power systems and passive building cooling, Sol. Energy Mater. Sol. Cells 178 (May 2018) 115–128, https://doi.org/10.1016/j.solmat.2018.01.015.
[18] Y. Xu, et al., A transparent radiative cooling emitter with multi-band spectral regulation for building energy saving, Compos. Commun. 43 (Nov. 2023) 101717, https://doi.org/10.1016/j.coco.2023.101717.
[19] Y. Tang, Q. Tao, Y. Chen, J. Zheng, Y. Min, Building envelopes with radiative cooling materials: a model for indoor thermal environment assessment based on climate adaptation, J. Build. Eng. 74 (Sep. 2023) 106869, https://doi.org/10.1016/j.jobe.2023.106869.
[20] Q. Gong, L. Lu, J. Chen, Design and performance investigation of a novel self-adaptive radiative cooling module for thermal regulation in buildings, Appl. Energy 352 (Dec. 2023) 121928, https://doi.org/10.1016/j.apenergy.2023.121928.
[21] C.-L. Luo, et al., Enhanced passive radiative cooling coating with Y2O3 for thermal management of building, Opt. Mater. 138 (Apr. 2023) 113710, https://doi.org/10.1016/j.optmat.2023.113710.
[22] H. Tang, et al., Radiative cooling performance and life-cycle assessment of a scalable MgO paint for building applications, J. Clean. Prod. 380 (Dec. 2022) 135035, https://doi.org/10.1016/j.jclepro.2022.135035.
[23] Y. Zhang, et al., Sub-ambient cooling effect and net energy efficiency of a super-amphiphobic self-cleaning passive sub-ambient daytime radiative cooling coating applied to various buildings, Energy Build. 284 (Apr. 2023) 112702, https://doi.org/10.1016/j.enbuild.2022.112702.
[24] S. Feng, et al., Superamphiphobic interface-enhanced inorganic coating for multi-environments tolerant and zero-energy building radiative cooling, Chem. Eng. J. 475 (Nov. 2023) 146191, https://doi.org/10.1016/j.cej.2023.146191.
[25] H. Sun, C. Hou, T. Ji, X. Zhou, Z. Ren, Y. Song, Processing bulk wood into a light-permeable passive radiative cooling material for energy-efficient building, Compos. Part B Eng. 250 (Feb. 2023) 110426, https://doi.org/10.1016/j.compositesb.2022.110426.
[26] T. Li, et al., A radiative cooling structural material, Science 364 (6442) (May 2019) 760–763, https://doi.org/10.1126/science.aau9101.
[27] D. Han, et al., Sub-ambient radiative cooling under tropical climate using highly reflective polymeric coating, Sol. Energy Mater. Sol. Cells 240 (Jun. 2022) 111723, https://doi.org/10.1016/j.solmat.2022.111723.
[28] J. Mandal, et al., Hierarchically porous polymer coatings for highly efficient passive daytime radiative cooling, Science 362 (6412) (Oct. 2018) 315–319, https://doi.org/10.1126/science.aat9513.
[29] A.P. Raman, M.A. Anoma, L. Zhu, E. Rephaeli, S. Fan, Passive radiative cooling below ambient air temperature under direct sunlight, Nature 515 (7528) (Nov. 2014) 540–544, https://doi.org/10.1038/nature13883.
[30] Y. Zhai, et al., Scalable-manufactured randomized glass-polymer hybrid metamaterial for daytime radiative cooling, Science 355 (6329) (Mar. 2017) 1062–1066, https://doi.org/10.1126/science.aai7899.
[31] X. Huang, J. Mandal, A.P. Raman, Do-it-yourself radiative cooler as a radiative cooling standard and cooling component for device design, J. Photon. Energy 12 (1) (Nov. 2021), https://doi.org/10.1117/1.JPE.12.012112.
[32] Y. Chen, et al., Colored and paintable bilayer coatings with high solar-infrared reflectance for efficient cooling, Sci. Adv. 6 (17) (Apr. 2020), https://doi.org/10.1126/sciadv.aaz5413.
[33] E.A. Goldstein, A.P. Raman, S. Fan, Sub-ambient non-evaporative fluid cooling with the sky, Nat. Energy 2 (9) (Sep. 2017) 17143, https://doi.org/10.1038/nenergy.2017.143.
[34] A.R. Gentle, G.B. Smith, A subambient open roof surface under the mid-summer sun, Adv. Sci. 2 (9) (Sep. 2015), https://doi.org/10.1002/advs.201500119.
[35] Y. Wu, H. Zhao, H. Sun, M. Duan, B. Lin, S. Wu, A review of the application of radiative sky cooling in buildings: challenges and optimization, Energy Convers. Manag. 265 (Aug. 2022) 115768, https://doi.org/10.1016/j.enconman.2022.115768.
[36] H. Fang, et al., Performance evaluation of a metamaterial-based new cool roof using improved Roof Thermal Transfer Value model, Appl. Energy 248 (Aug. 2019) 589–599, https://doi.org/10.1016/j.apenergy.2019.04.116.
[37] D. Zhao, et al., Subambient cooling of water: toward real-world applications of daytime radiative cooling, Joule 3 (1) (Jan. 2019) 111–123, https://doi.org/10.1016/j.joule.2018.10.006.
[38] M. Santamouris, Cooling the cities – a review of reflective and green roof mitigation technologies to fight heat island and improve comfort in urban environments, Sol. Energy 103 (May 2014) 682–703, https://doi.org/10.1016/j.solener.2012.07.003.
[39] H. Pieper, T. Kirs, I. Krupenski, A. Ledvanov, K. Lepiksaar, A. Volkova, Efficient use of heat from CHP distributed by district heating system in district cooling networks, Energy Rep. 7 (Oct. 2021) 47–54, https://doi.org/10.1016/j.egyr.2021.09.041.
[40] R. Fortin, J. Mandal, A.P. Raman, S. Craig, Passive radiative cooling to sub-ambient temperatures inside naturally ventilated buildings, Cell Reports Phys. Sci. 4 (9) (Sep. 2023) 101570, https://doi.org/10.1016/j.xcrp.2023.101570.
[41] W. Mucha, A. Mainka, E. Brągoszewska, Impact of ventilation system retrofitting on indoor air quality in a single-family building, Build. Environ. 262 (Aug. 2024) 111830, https://doi.org/10.1016/j.buildenv.2024.111830.
[42] H. Chen, H. Zhang, S. Wu, Y. Liu, H. Zhong, Numerical simulation and optimisation design for ventilation and heat dissipation in high-temperature and high-load indoor substations, Case Stud. Therm. Eng. 59 (Jul. 2024) 104502, https://doi.org/10.1016/j.csite.2024.104502.
[43] J.P. Bijarniya, J. Sarkar, P. Maiti, Environmental effect on the performance of passive daytime photonic radiative cooling and building energy-saving potential, J. Clean. Prod. 274 (Nov. 2020) 123119, https://doi.org/10.1016/j.jclepro.2020.123119.

[44] S.W.Y. Wu, H. Zhao, H. Sun, M. Duan, B. Lin, A Review of the Application of Radiative Sky Cooling in Buildings: Challenges and Optimization, 2023 arXiv: 2307.05862.

[45] H.L. Gough, et al., Field measurement of natural ventilation rate in an idealised full-scale building located in a staggered urban array: comparison between tracer gas and pressure-based methods, Build. Environ. 137 (Jun. 2018) 246–256, https://doi.org/10.1016/j.buildenv.2018.03.055.

[46] Y. Tong, et al., Field measurements on thermal stratification and cooling potential of natural ventilation for large space buildings, Int. J. Vent. 19 (1) (Jan. 2020) 49–62, https://doi.org/10.1080/14733315.2018.1544730.

[47] S. Fan, M.S. Davies Wykes, W.E. Lin, R.L. Jones, A.G. Robins, P.F. Linden, A full-scale field study for evaluation of simple analytical models of cross ventilation and single-sided ventilation, Build. Environ. 187 (Jan. 2021) 107386, https://doi.org/10.1016/j.buildenv.2020.107386.

[48] H. Hu, H. Kikumoto, R. Ooka, C. Lin, B. Zhang, Comprehensive validation of experimental and numerical natural ventilation predictions based on field measurement with experimental house, Build. Environ. 207 (Jan. 2022) 108433, https://doi.org/10.1016/j.buildenv.2021.108433.

[49] W. Liu, et al., Machine learning applications for photovoltaic system optimization in zero green energy buildings, Energy Rep. 9 (Dec. 2023) 2787–2796, https://doi.org/10.1016/j.egyr.2023.01.114.

[50] H.-J. Wang, T. Jin, H. Wang, D. Su, Application of IEHO–BP neural network in forecasting building cooling and heating load, Energy Rep. 8 (Jul. 2022) 455–465, https://doi.org/10.1016/j.egyr.2022.01.216.

[51] T.C. Quevedo, M.S. Geraldi, A.P. Melo, Applying machine learning to develop energy benchmarking for university buildings in Brazil, J. Build. Eng. 63 (Jan. 2023) 105468, https://doi.org/10.1016/j.jobe.2022.105468.

[52] Y. Hu, et al., Adaptive corrected parameters algorithm applied in cooling load prediction based on black-box model: a case study for subway station, Energy Build. 297 (Oct. 2023) 113429, https://doi.org/10.1016/j.enbuild.2023.113429.

[53] X. Deng, et al., Bagging–XGBoost algorithm based extreme weather identification and short-term load forecasting model, Energy Rep. 8 (Nov. 2022) 8661–8674, https://doi.org/10.1016/j.egyr.2022.06.072.

[54] C. Cakiroglu, Y. Aydın, G. Bekdaş, U. Isikdag, A.N. Sadeghifam, L. Abualigah, 'Corrigendum to "Cooling load prediction of a double-story terrace house using ensemble learning techniques and genetic programming with SHAP approach", Energy Build. 313 (2024) 114254 https://doi.org/10.1016/j.enbuild.2024.114329. Energy Build., vol. 315, p. 114329, Jul. 2024.

[55] V. Vajrobol, et al., Enhancing aviation control security through ADS-B injection detection using ensemble meta-learning models with Explainable AI, Alex. Eng. J. 112 (Jan. 2025) 63–73, https://doi.org/10.1016/j.aej.2024.10.042.

[56] A. Naghibi, Exploring explainable ensemble machine learning methods for long-term performance prediction of industrial gas turbines: a comparative analysis, Eng. Appl. Artif. Intell. 138 (Dec. 2024) 109318, https://doi.org/10.1016/j.engappai.2024.109318.

[57] S.S. Ansari, A. Azeem, M. Asad, K. Zafar, S.M. Ibrahim, Comparative analysis of conventional and ensemble machine learning models for predicting split tensile strength in thermal stressed SCM-blended lightweight concrete, Mater. Today Proc. (Apr. 2024), https://doi.org/10.1016/j.matpr.2024.04.081.

[58] D. Liang, Z. Rui, F. Yuguang, A robust evaluating strategy of tunnel deterioration using ensemble machine learning algorithms, Eng. Appl. Artif. Intell. 133 (Jul. 2024) 108364, https://doi.org/10.1016/j.engappai.2024.108364.

[59] B. Sadaghat, S. Afzal, A.J. Khiavi, Residential building energy consumption estimation: a novel ensemble and hybrid machine learning approach, Expert Syst. Appl. 251 (Oct. 2024) 123934, https://doi.org/10.1016/j.eswa.2024.123934.

[60] Z. Wang, Z. Liang, R. Zeng, H. Yuan, R.S. Srinivasan, Identifying the optimal heterogeneous ensemble learning model for building energy prediction using the exhaustive search method, Energy Build. 281 (Feb. 2023) 112763, https://doi.org/10.1016/j.enbuild.2022.112763.

[61] Y. Shen, L. Wu, S. Liang, Explainable machine learning-based model for failure mode identification of RC flat slabs without transverse reinforcement, Eng. Fail. Anal. 141 (Nov. 2022) 106647, https://doi.org/10.1016/j.engfailanal.2022.106647.

[62] S. Mangalathu, H. Shin, E. Choi, J.-S. Jeon, Explainable machine learning models for punching shear strength estimation of flat slabs without transverse reinforcement, J. Build. Eng. 39 (Jul. 2021) 102300, https://doi.org/10.1016/j.jobe.2021.102300.

[63] M.Z. Naser, V.K. Kodur, Explainable machine learning using real, synthetic and augmented fire tests to predict fire resistance and spalling of RC columns, Eng. Struct. 253 (Feb. 2022) 113824, https://doi.org/10.1016/j.engstruct.2021.113824.

[64] A. Sharma, M. Singh, Batch reinforcement learning approach using recursive feature elimination for network intrusion detection, Eng. Appl. Artif. Intell. 136 (Oct. 2024) 109013, https://doi.org/10.1016/j.engappai.2024.109013.

[65] X. Ding, F. Yang, F. Ma, An efficient model selection for linear discriminant function-based recursive feature elimination, J. Biomed. Inf. 129 (May 2022) 104070, https://doi.org/10.1016/j.jbi.2022.104070.

[66] Y. Wang, Y. Li, Mapping the ratoon rice suitability region in China using random forest and recursive feature elimination modeling, F. Crop. Res. 301 (Oct. 2023) 109016, https://doi.org/10.1016/j.fcr.2023.109016.

[67] N. Rtayli, N. Enneya, Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization, J. Inf. Secur. Appl. 55 (Dec. 2020) 102596, https://doi.org/10.1016/j.jisa.2020.102596.

[68] 'sklearn.feature_selection.RFE'. https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html (accessed July 25, 2024).

[69] Z. Wang, Y. Wang, R.S. Srinivasan, A novel ensemble learning approach to support building energy use prediction, Energy Build. 159 (Jan. 2018) 109–122, https://doi.org/10.1016/j.enbuild.2017.10.085.

[70] K. Li, J. Zhang, X. Chen, W. Xue, Building's hourly electrical load prediction based on data clustering and ensemble learning strategy, Energy Build. 261 (Apr. 2022) 111943, https://doi.org/10.1016/j.enbuild.2022.111943.

[71] Y. Bai, K. Liu, Y. Wang, Comparative analysis of thermal preference prediction performance in different conditions using ensemble learning models based on ASHRAE Comfort Database II, Build. Environ. 223 (Sep. 2022) 109462, https://doi.org/10.1016/j.buildenv.2022.109462.

[72] Z. Li, H. Lei, E. Ma, J. Lai, J. Qiu, Ensemble technique to predict post-earthquake damage of buildings integrating tree-based models and tabular neural networks, Comput. Struct. 287 (Oct. 2023) 107114, https://doi.org/10.1016/j.compstruc.2023.107114.

[73] K. Li, J. Tian, W. Xue, G. Tan, Short-term electricity consumption prediction for buildings using data-driven swarm intelligence based ensemble model, Energy Build. 231 (Jan. 2021) 110558, https://doi.org/10.1016/j.enbuild.2020.110558.

[74] K. Tang, X. Zhao, Z. Xu, H. Sun, A stacking ensemble model for predicting soil organic carbon content based on visible and near-infrared spectroscopy, Infrared Phys. Technol. 140 (Aug. 2024) 105404, https://doi.org/10.1016/j.infrared.2024.105404.

[75] Y. He, H. Zhang, Y. Dong, C. Wang, P. Ma, Residential net load interval prediction based on stacking ensemble learning, Energy 296 (Jun. 2024) 131134, https://doi.org/10.1016/j.energy.2024.131134.

[76] R. Du, et al., Estimation of winter canola growth parameter from UAV multi-angular spectral-texture information using stacking-based ensemble learning model, Comput. Electron. Agric. 222 (Jul. 2024) 109074, https://doi.org/10.1016/j.compag.2024.109074.

[77] H. Zhu, G. Xie, A.S. Berrouk, Enhancing performance of multi-pressure evaporation organic Rankine Cycle/Supercritical Carbon Dioxide Brayton cycle through genetic algorithm and Machine learning optimization, Energy Convers. Manag. 301 (Feb. 2024) 118037, https://doi.org/10.1016/j.enconman.2023.118037.

[78] H. Yang, P. Chen, H. Zhou, J. Tan, An efficient position optimization method based on improved genetic algorithm and machine learning for sparse array, AEU - Int. J. Electron. Commun. 179 (May 2024) 155312, https://doi.org/10.1016/j.aeue.2024.155312.

[79] J. Fan, et al., Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: a review and case study in China, Renew. Sustain. Energy Rev. 100 (Feb. 2019) 186–212, https://doi.org/10.1016/j.rser.2018.10.018.

[80] M.M. Rajabi, B. Ataie-Ashtiani, H. Janssen, Efficiency enhancement of optimized Latin hypercube sampling strategies: application to Monte Carlo uncertainty analysis and meta-modeling, Adv. Water Resour. 76 (Feb. 2015) 127–139, https://doi.org/10.1016/j.advwatres.2014.12.008.

[81] C. Song, R. Kawai, Monte Carlo and variance reduction methods for structural reliability analysis: a comprehensive review, Probabilistic Eng. Mech. 73 (Jul. 2023) 103479, https://doi.org/10.1016/j.probengmech.2023.103479.

[82] C. Joo, H. Park, J. Lim, H. Cho, J. Kim, Machine learning-based heat deflection temperature prediction and effect analysis in polypropylene composites using catboost and shapley additive explanations, Eng. Appl. Artif. Intell. 126 (Nov. 2023) 106873, https://doi.org/10.1016/j.engappai.2023.106873.
[83] G. Zheng, Y. Zhang, X. Yue, K. Li, Interpretable prediction of thermal sensation for elderly people based on data sampling, machine learning and SHapley Additive exPlanations (SHAP), Build. Environ. 242 (Aug. 2023) 110602, https://doi.org/10.1016/j.buildenv.2023.110602.
[84] H.A.H. Al-Najjar, B. Pradhan, G. Beydoun, R. Sarkar, H.-J. Park, A. Alamri, A novel method using explainable artificial intelligence (XAI)-based Shapley Additive Explanations for spatial landslide prediction using Time-Series SAR dataset, Gondwana Res. 123 (Nov. 2023) 107–124, https://doi.org/10.1016/j.gr.2022.08.004.