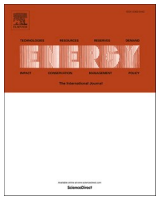
Energy 347 (2026) 140459

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy



Smart HVAC optimization using machine learning and self-adaptive NSGA-II for energy-efficient thermal comfort

Sampath Suranjan Salins [a], Shiva Kumar [b,*], Subraya Krishna Bhat [b]

[a] Manipal Academy of Higher Education, Dubai Campus, PO 345050, United Arab Emirates
[b] Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India

ARTICLE INFO

Keywords:
Air handling unit
Thermal comfort
Predicted mean vote
Energy consumption
Optimization

ABSTRACT

A Building Management System (BMS) aims to maintain optimal thermal comfort within an air-conditioned space, even as external conditions fluctuate. Integrating an optimization model with the Air Handling Unit (AHU) enhances the unit's overall performance while minimizing power consumption. This work focuses on designing and constructing an Air Handling Unit that evaluates performance parameters under varying climatic conditions. Air passes through a helical-coil dehumidifier, an ultrasonic humidifier, and a damper into the room to regulate the required conditions. This study proposes a machine learning-based optimization framework to regulate thermal comfort while minimizing energy consumption in Air Handling Units (AHUs). A predictive model was developed using Random Forest Regressor, XGBoost Regressor, and Artificial Neural Network (ANN), trained on experimental data to estimate PMV, PPD, and energy consumption based on input air conditions. A self-adaptive Non-dominated Sorting Genetic Algorithm II (NSGA-II) was employed to predict optimal inlet air parameters—including temperature, velocity, and relative humidity—within defined thermal comfort constraints. The optimization results were validated experimentally using a test configured with input conditions derived from adaptive NSGA-II predictions, and the resulting thermal comfort indices and energy usage were measured. The prediction errors were minimal—0.8% for energy consumption, 1.2% for Predicted Mean Vote (PMV) and 2.7% for Predicted Percentage of Dissatisfaction (PPD)—demonstrating the accuracy and robustness of the approach. Experimental validation under optimized inlet conditions confirmed the model's reliability, with minimal prediction errors of 0.8% in energy consumption, 1.2% in PMV, and 2.7% in PPD relative to measured values. This work confirms the viability of using ML-based multi-objective optimization for clean, energy-efficient, and comfort-focused HVAC control in smart building environments.

1. Introduction

HVAC systems account for nearly 40% of global energy consumption and contribute to 36% of greenhouse gas emissions. Efficient energy utilization in HVAC systems and maintaining thermal comfort within a space are key factors to consider [1]. Conventional air conditioning (AC) systems often encounter issues such as excessive energy consumption, inconsistent temperature control, frequent maintenance, and inadequate air quality management. These challenges are addressed by air handling units (AHUs), which feature advanced control systems for managing air quality, temperature, and humidity. AHUs are integrated with building management systems to provide more efficient and effective climate control.

It typically includes components such as a dehumidifier, a humidifier, a damper, filters, and a blower. Conventional AHU units were not designed to handle varying climatic conditions and often operated at a constant energy consumption [2]. Adaptive AHU units, however, are crucial for maintaining optimal conditions by adjusting to changing thermal load requirements [3]. One way to reduce energy consumption and enhance occupant satisfaction in air-conditioned spaces is by employing an optimization tool. The optimization method aims to address inefficiencies in the automated operation of AHUs by incorporating real-world data and predictive models [4,5]. Optimization methods offer several benefits, including extending equipment lifespan by reducing wear and tear, providing cost savings for building operators, maintaining comfortable indoor conditions, and minimizing energy loss [6]. This study aimed to develop an optimization model to identify the ideal thermal comfort parameters, such as Predicted Mean Vote (PMV)

* Corresponding author.
E-mail address: shiva.kumar@manipal.edu (S. Kumar).

https://doi.org/10.1016/j.energy.2026.140459
Received 14 October 2025; Received in revised form 5 February 2026; Accepted 12 February 2026
Available online 14 February 2026

and Predicted Percentage of Dissatisfaction (PPD), for different climatic conditions. Energy saving was a key performance indicator, and the work also emphasized achieving optimal power consumption to maximize overall performance.

Some alternative cooling methods are used to maintain thermal comfort in the room. Stasi et al. [7] examined the effectiveness of natural ventilation using computational fluid dynamics (CFD) analysis, focusing on how cross-ventilation could enhance cooling efficiency. Their study revealed that cross ventilation was particularly adequate on the hottest days. Ferdyn-Grygierek et al. [8] explored passive and energy-efficient methods for maintaining optimal operating conditions in Polish homes. Their research focused on an automatically controlled air circulation system, which effectively ensured proper ventilation and thermal comfort within the space. Suranjan Salins et al. [9] utilized a reciprocating evaporative cooling test rig to address cooling loads under varying climatic conditions. This unit proved effective in hot and arid climates, achieving an air temperature drop of 8.2 °C.

Many researchers have investigated the optimization of performance parameters in their studies, and this work seeks to align with and build upon their techniques in the current context. Kumar et al. [10] used Air Handling Units (AHUs) to regulate temperature, humidity, and airflow within a data center. The challenge addressed was optimizing airflow and AHU fan design for effective cooling. The proposed machine learning technique successfully predicted the optimal fan speed, ensuring efficient cooling and maintaining the data center's temperature. Wani et al. [11] introduced a resilient control strategy that leverages meta-heuristic algorithms to enhance both energy efficiency and indoor thermal comfort. The proposed framework was evaluated using weather data from Auckland, New Zealand. The results showed that the optimization model effectively reduced annual energy consumption by 49.13%. Lin et al. [12] employed three primary techniques to regulate conditions within a specified space. The first technique involved an algorithm for the Fan Coil Unit (FCU) to adjust the temperature based on varying climatic conditions. The second technique focused on calculating the work required by the refrigeration unit to maintain thermal comfort. The third technique used a genetic algorithm to reduce HVAC system energy consumption. The results showed that the system achieved a 39.71% reduction in energy usage. Similar work by Lee et al. [13] used a deterministic policy gradient algorithm to control cooling coils and auto-tune PID controllers in a simulator. Machine Learning techniques were employed to maintain thermal comfort. The study found a daily average energy savings of 13.71%. Razban et al. [14] employed algorithms to model the energy consumption of heating and cooling coils, fans, and other components. They developed two control systems based on occupancy and carbon dioxide (CO_2) data. An on/off fuzzy logic controller, implemented in MATLAB/Simulink, was used to reduce energy consumption and improve system performance. The results demonstrated that the fuzzy logic controller achieved a 62% reduction in fan energy consumption every week, while maintaining occupant thermal comfort. Shang et al. [15] utilized a multivariable nonlinear dynamic model to optimize temperature, humidity, and carbon dioxide concentration within a room. The results showed that this model operated effectively with minimal uncertainties. Hosamo et al. [16] developed a Digital Twin framework integrated with a plugin to gather sensor data on thermal comfort for the Norwegian region and applied optimization processes using MATLAB. The overall HVAC system energy consumption was modeled using an artificial neural network (ANN) within a Simulink model and a multi-objective genetic algorithm (MOGA). The results indicated an energy savings of 13.2% during the summer season. Casillas et al. [17] investigated a fault-detection model for Air Handling Units (AHUs), focusing on duct models. The simulations incorporated sensors and mechanical components to replicate real-life scenarios closely. The simulation and fault modeling techniques proved effective in predicting operations and evaluating control robustness and fault tolerance. Zheng et al. [18] employed an automatic AHU technique to test the system under three different climatic conditions, achieving a maximum energy savings of 55.2%. Ambroziak and Chojecki [19] integrated a nonlinear autoregressive model with moving-average and exogenous inputs with a fuzzy self-tuning swarm optimization algorithm. This model effectively optimized both old and new AHUs, improving their reliability and efficiency. Li et al. [20] investigated variable-air-volume (VAV) regulation in AHUs, focusing on the supply fan and other components. They identified fuzzy PI regulation as a crucial parameter, resulting in a 2.7% reduction in energy consumption for the air supply fan. Yao and Huang [21] used a multi-objective optimization method to manage the thermal conditions in a data centre. They optimized rack intake temperature and energy consumption using the Non-Dominated Sorting Genetic Algorithm II (NSGA-II). Their study concluded that the desired thermal conditions were effectively maintained within the data centre. Similar work was carried out by Talib et al. [22], in which the data centre was considered, resulting in energy savings of 9.04%. The impact of indoor humidity on cooling requirements was assessed by Kostyák et al. [23] by using an optimization tool. The study concluded that the AHU effectively managed humidity control.Bareschino et al. [24] employed the Hooke–Jeeves algorithm to optimize the cooling coil. They introduced an auto-tuned PID control system enhanced with machine learning to maintain a steady thermal environment at (23.5 ± 0.5 °C) with 97% accuracy during occupied periods. Their approach resulted in a 13.71% reduction in cooling energy consumption. Gao et al. [25] used a reinforcement learning (RL) algorithm to optimize thermal parameters, achieving an 88.4% improvement in thermal comfort. Samadi and Shahbakhti [26] introduced a nonlinear predictive control (NMPC) tool to evaluate both energy requirements and indoor infection risks related to COVID-19. The study focused on a classroom equipped with sensors to measure temperature, relative humidity, and carbon dioxide levels. The results demonstrated that the model could predict temperature and carbon dioxide concentration with minimal errors of 0.8% and 2.4%, respectively. Kim et al. [27] developed an Artificial Neural Network (ANN) predictive model for use in building air conditioning systems to forecast energy consumption. After validating the forecasting models, the AHU prediction model demonstrated a mean bias error of 5% during training and 4.5% during testing. The results showed that the ANN effectively predicted energy consumption and significantly reduced it.

Recent advancements in multi-objective optimization have demonstrated that traditional NSGA-II, while widely used, often suffers from premature convergence and limited diversity control when addressing highly nonlinear HVAC optimization problems. Contemporary studies show a clear shift toward adaptive and self-tuning evolutionary algorithms that dynamically adjust crossover, mutation, and selection parameters to improve search efficiency. Hao et al. [28] introduced an *improved NSGA-II* featuring adaptive weighting and Lévy-based search operators, achieving significantly better convergence and avoidance of local optima compared to the classical algorithm. Similarly, Zhao [29] proposed an *adaptive strategy-enhanced NSGA-II* that demonstrated stronger performance on benchmark ZDT/DTLZ functions, validating the importance of dynamic operator control for maintaining population diversity and improving Pareto-front quality. A comprehensive survey by Ma et al. [30] further highlights that algorithmic self-adaptation is now considered a core requirement for solving high-complexity engineering problems, particularly where objective landscapes exhibit strong nonlinearity and interactions.

In the HVAC domain specifically, recent optimization frameworks increasingly incorporate adaptive GA, metaheuristics, and multi-objective reinforcement learning (RL). Soleimani and Sa'adati [31] reported that NSGA-II remains one of the most stable algorithms for HVAC comfort–energy trade-off optimization, outperforming GA and PSO variants in Pareto stability for smart buildings. Likewise, Digital Twin–integrated adaptive GA approaches have shown superior performance, achieving 3–5% improvements over NSGA-II and MOPSO by dynamically tuning search operators based on surrogate-model feedback [32]. These recent findings collectively reinforce that self-adaptive

NSGA-II offers clear advantages over static evolutionary strategies, particularly for HVAC optimization tasks with competing objectives (thermal comfort vs. energy use), highly coupled variables, and diverse operating conditions.

Building on these developments, the present study introduces a novel integration of hyperparameter-optimized machine learning models with a self-adaptive NSGA-II framework that automatically tunes evolutionary parameters to enhance optimization robustness. Unlike prior ML-GA-based HVAC optimization studies, our method combines (i) Optuna-driven tuning of Random Forest, XGBoost, and ANN models, (ii) a fully self-adaptive evolutionary optimization strategy, and (iii) experimental validation on a custom-built AHU test rig using inlet conditions predicted by the algorithm. In addition, the optimization is validated experimentally on a fully constructed AHU test rig using inlet conditions generated by the algorithm—an aspect rarely demonstrated in prior ML-based optimization research. This combination of self-adaptation, experimental validation, and multi-objective ML-driven optimization constitutes the core novelty of this work.

An extensive literature review indicates that researchers have explored a range of cooling techniques, including both active and passive evaporative cooling methods. Additionally, some studies have focused on reducing indoor heat through effective ventilation strategies. Optimization techniques have been employed to predict indoor conditions and parameters related to air-handling units, using a variety of methods. These include genetic algorithms, deterministic policy gradient algorithms, fuzzy logic, multivariable nonlinear dynamic models, Digital Twin frameworks, artificial neural networks (ANNs), non-linear autoregressive models, Hooke–Jeeves algorithms, and reinforcement learning (RL) algorithms. These methods have been applied to predict cooling effects, thermal comfort, fan speed regulation, energy requirements, carbon dioxide levels, and unit faults. Such research has predominantly focused on data centers, building environments, and industrial workspaces.

1.1. Research gaps and objectives

Despite the extensive research on predictive modeling for air handling units (AHUs), there is a notable gap in the literature concerning the optimization of various input conditions, such as inlet relative humidity, temperature, and air velocity, to maintain thermal comfort within the ideal and acceptable limits of PMV and PPD. While these parameters significantly influence the system's energy consumption, optimizing them to achieve thermal comfort within the intended space has not been adequately explored.

Machine Learning (ML) offers significant advantages for modeling the complex interactions among input variables, such as air velocity, inlet temperature, and humidity, and output parameters, such as energy consumption and thermal comfort indices (PMV and PPD). Unlike classical models, ML algorithms can effectively capture the nonlinear relationships inherent in AHU systems, allowing for more accurate predictions and control decisions. The lack of studies on the optimized ranges of these input conditions underscores a critical need for further research, as ML-based prediction and optimization are crucial to enhancing energy efficiency and ensuring occupant comfort.

To address the identified gap, a sustainable energy-based air handling unit (AHU) has been developed, incorporating a vapor-compression dehumidifier. The air is cooled below its dew point to induce condensation, effectively managing humidity levels. This dehumidified air is then connected to a humidifier, allowing precise regulation of both relative humidity and temperature before it enters the room. The study investigates air velocity in the range of 1.5–6 m/s, ambient air temperature in the range of 25–45 °C, and inlet relative humidity in the range of 50–90%. These ranges were selected to represent typical indoor and near-extreme operating conditions relevant to thermal comfort and dehumidification applications. To optimize AHU performance, operating parameters have been fine-tuned using machine learning approaches. In this study, OPTUNA was employed to fine-tune three machine learning models: Random Forest Regressor, XGBoost Regressor, and Artificial Neural Network (ANN), each with distinct hyperparameter search spaces NSGA-II is used to optimize the input parameters related to AV, inlet temperature, and RH to have optimum PMV and PPD values while minimizing the energy demand.

2. Methodology

This study aims to achieve optimal thermal comfort in air-conditioned spaces while minimizing energy consumption. Air, conditioned by varying climatic factors, passes through both dehumidifiers and humidifiers before being circulated within the room at a controlled velocity, managed by a damper. Data on Predicted Mean Vote (PMV), Predicted Percentage of Dissatisfied (PPD), and energy consumption are collected for different inlet temperatures, relative humidity (RH) levels, and air velocities. This data is then optimized to determine the ideal conditions for maximizing experimental results. Methodology is shown in Fig. 1.

Optimizing air handling unit parameters involves adjusting various settings to enhance system efficiency and performance. The primary goal of this optimization is to create optimal indoor conditions for occupants by fine-tuning airflow rate, temperature, fan speed, and humidity. Additionally, the process aims to reduce overall power consumption. Predicted Mean Vote (PMV) is an index used to assess thermal comfort and to evaluate occupants' thermal sensation in a given environment. It takes into account factors such as temperature, humidity, clothing, and metabolic rate. The PMV scale ranges from −3 (cold) to +3 (hot), with 0 indicating neutral thermal comfort. A PMV value between −0.7 and + 0.7 is generally considered to be a reasonable and satisfactory range.

$$PMV = \left(0.303e^{2.1*M} + 0.028\right)*[(M-W) - H - E_c - C_{res} - E_{res}] \quad (1)$$

Where M being the metabolic rate, W/m^2, H is the sensitive heat loses, W - effective mechanical power, W/m^{2}W/m^2, E_c is the heat exchanged by the evaporation on the skin, W/m^2, C_{res} is the heat exchange by convection, W/m^2, E_{res} is the evaporation rate exchange in breathing, W/m^2.

Sensitive heat loss (H) is given by equation (2).

$$H = 3.96x10^{-8} f_{cl}\left[(t_{cl}+273)^4 - (t_r+273)^4\right] - f_{cl}\, h_c\, (t_{cl}-t_a) \quad (2)$$

Where, f_{cl} is factor for clothing surface, t_{cl} is the clothing surface temperature, t_r is the mean radiant temperature, t_a is the ambient temperature. Equations (3)–(5) gives the heat exchanged through evaporation of skin, convection and breathing, h_c is the heat transfer coefficient and P_a is the vapor partial pressure.

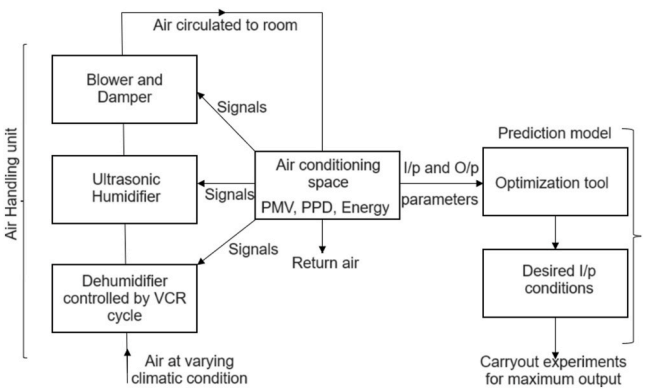


Fig. 1. Optimization of Air handling unit parameters.

$$E_c = 3.05\ x\ 10^{-3}\ [5773 - 6.99\ (M - W) - P_a] - 0.42\ [\ (M - W) - 58.15] \quad (3)$$

$$C_{res} = 0.0014\ M\ (34 - t_a) \quad (4)$$

$$E_{res} = 1.7\ x\ \ 10^{-5}\ M\ (5867 - P_a) \quad (5)$$

PPD is a thermal comfort parameter that complements the PMV. Expressed as a percentage, PPD indicates the proportion of occupants likely to experience discomfort, with higher values reflecting greater levels of dissatisfaction. A PPD value less than 22.5 is generally accepted as the most reasonable and optimal range for satisfactory performance. It is shown using equation (6).

$$PPD = 100 - 95e^{-\left(0.03353*PMV^4 + 0.2179*PMV^2\right)} \quad (6)$$

E is the total energy consumed by the combination of blower, compressor and humidifier. It is given by equation (7).

$$E = E_{Blower} + E_{Compressor} + E_{Humidifier} \quad (7)$$

3. Construction and working of the AHU unit

Construction: The automatic AHU unit controls the flow of conditioned air within a specified space for HVAC applications. It manages the intake of fresh air and adjusts thermal parameters—such as temperature, humidity, and air velocity—based on signals received from the room. The AHU system includes a variable control damper (VCR) that regulates air flow and is powered by a DC 12V 100 rpm gear motor. The blower, which provides the air, operates at 370W, 200V, and speeds between 3000 and 3600 rpm. An ultrasonic humidifier introduces humid air and has a power rating of 38W and 22V. The dehumidification coils circulate cold water, with a half-ton VCR unit cooling the water at 220 to 240V. The return air exhaust fan operates at 230V and has a volume flow rate of 3.5 m^3/min. The adiabatic cooling space measures 1x1x1 m^3, and the AHU stand is constructed from ½" PVC tubing. Fig. 2 illustrates the top-view schematic of the air handling unit.

Five sensors are positioned within the room, transmitting data to the Arduino, which then controls the AHU components. In the AHU system, Arduino acts as the central controller, gathering data from various sensors. It uses predefined algorithms to analyze this data and make decisions based on the room's requirements. The temperature and humidity sensors (DHT11) measure moisture levels and temperature, adjusting the variable control dampers and humidifier accordingly. The dust sensor (GP2Y1010AU0F) manages the intake of fresh air and reduces pollution. The CO and CO2 sensor (MQ135) monitors air quality and controls the exhaust fan. The presence sensor (GY-906 MLX90614ESF) detects heat-emitting objects and controls the variable control dampers, while the motion sensor (HC-SR501) detects movement and adjusts the supply VCD accordingly.

Working: The air-conditioned space is equipped with sensors linked to an Arduino system that controls the blower, damper, humidifier, and dehumidifier. The air quality entering the room depends on the room's thermal settings. The blower draws in outside air, which then interacts with cold coils in the air handling unit to dehumidify it, powered by the VCR cycle. After dehumidification, the air passes through an ultrasonic humidifier that adjusts humidity to the room's requirements. The motor-controlled damper regulates the air flow rate, and the air is filtered to remove dust particles before it enters the room. Fig. 3 shows the sectional model of the AHU unit.

3.1. Instruments used and experimental conditions

Various instruments are used to measure input and output parameters. A Digital Thermometer measures temperatures from −20 to 100 °C with an accuracy of ±0.1 °C, a resolution of 0.1 °C, and features a stainless-steel probe. Digital Hygrometer measures relative humidity in the air-conditioned space with a range of 0 to 99%, an accuracy of ±1%, a power supply of 1.5V, and a resolution of 0.1%.Anemometer measures air velocity with a range of 0-30 m/s, an accuracy of ±0.5 m/s, a resolution of 0.1 m/s, and a threshold sensitivity of 1 m/s. A clamp meter measures instantaneous voltage and current, with a range of 400 A for AC current, 600 V for both AC and DC voltage, and 40 MΩ for resistance. Table 1 provides details on the experimental conditions and performance parameters associated with the AHU.

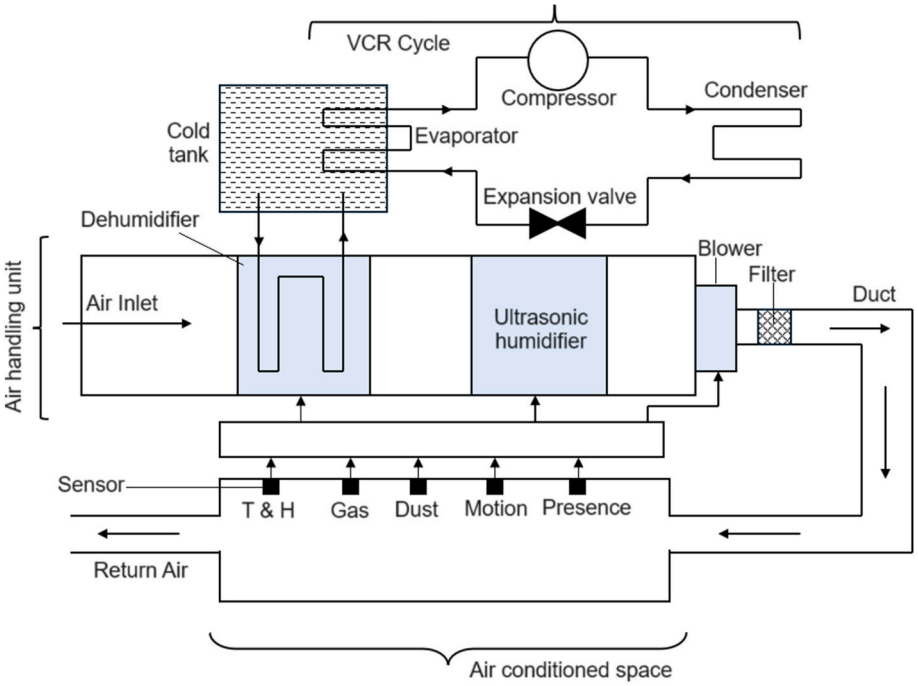


Fig. 2. Schematic sketch of an AHU unit.

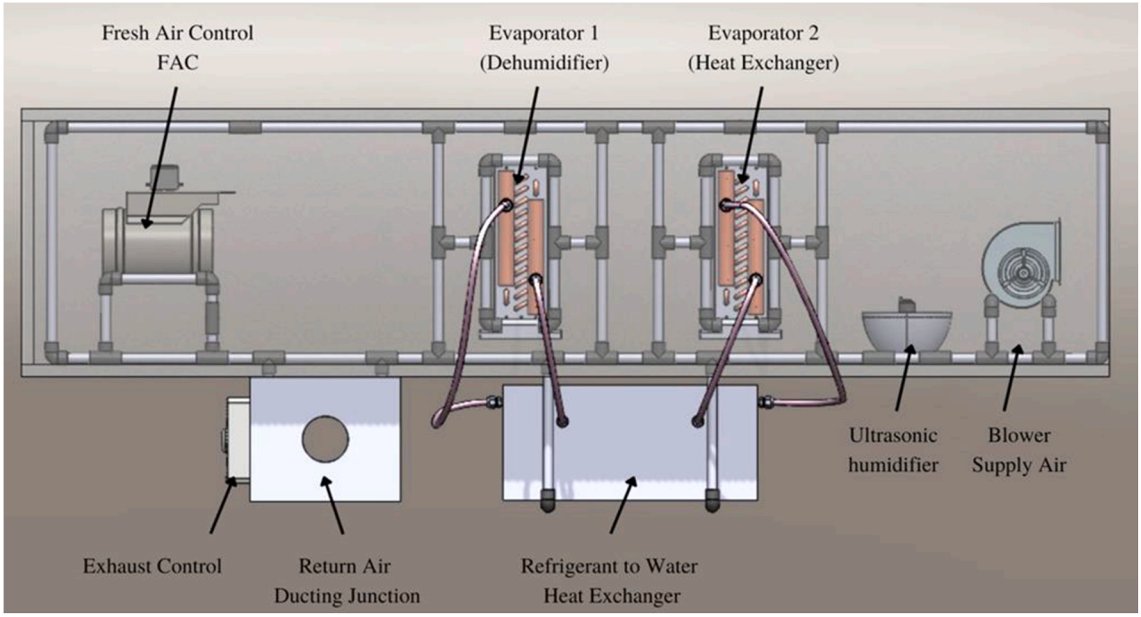


Fig. 3. Sectional model view of the AHU unit.

Table 1
Experimental conditions.

Varying conditions	Measuring points	Performance terms
Air velocity: 1.5, 3, 4.5 and 6 m/s Values of temperature: 25, 30, 35, 40 and 45 °C Inlet RH = 50, 60, 70, 80 and 90% Sensible heat source inside the room: 200 W.	Humidity, temperature and air velocity with its respective power consumption values.	Performance predicted mean vote, predicted percentage of dissatisfaction, and energy consumption.

4. Results and discussion

To examine the impact of various parameters on the AHU performance, experiments were conducted using the conditions as mentioned in Table 1 on the AHU test facility. To ensure the reliability and repeatability of the experimental data, three independent trials were conducted for each combination of operating conditions. For each experimental condition, the three measured values for each parameter were averaged, and the resulting mean values were used for further analysis. These averaged data were then employed to evaluate the performance parameters, namely Predicted Mean Vote (PMV), Predicted Percentage of Dissatisfied (PPD), and energy consumption. Conducting multiple trials and averaging the results helped reduce random experimental errors, improve data consistency, and enhance the overall statistical reliability of the reported results.

The influence of inlet air velocity and inlet relative humidity (RH) on thermal comfort was examined using PMV and PPD as shown in Fig. 4. The results demonstrate clear trends in both parameters as functions of air velocity and RH. It was observed that PMV values decreased with increasing air velocity across all RH levels. This trend suggests that higher air velocity enhances convective and evaporative heat loss from the human body, thereby increasing the sensation of coolness. As a result, occupants tend to feel cooler at higher air velocities, as reflected in the more negative PMV values. Similar trends are observed in the literature Shakya et al. [33].

Moreover, at any constant air velocity, PMV was found to be lower (more negative) with higher inlet RH which is similar to the research findings of Dyvia and Arif [34]. While high RH generally impedes evaporative cooling and is expected to make occupants feel warmer, the observed decrease in PMV with RH may be attributed to the combined cooling effect of airflow overpowering the moisture-induced discomfort. Alternatively, this could indicate that the cooling impact of air velocity becomes more pronounced under high humidity conditions due to changes in convective heat transfer dynamics. The PPD index increased gradually with increasing air velocity across all RH levels. This indicates that although higher air velocities provide enhanced cooling, they may also lead to overcooling or cause discomfort due to drafts, especially in cooler environments. This is consistent with the PMV trend moving further into the negative (cooler) range.

Additionally, an increase in RH led to higher PPD values across all air velocities. Elevated RH levels reduce evaporative heat loss, leading to greater thermal discomfort among occupants. The combined effect of high RH and high air velocity likely contributes to a thermal environment that deviates further from the thermal neutrality preferred by most individuals, thus increasing the dissatisfaction rate. A similar trend has been reported in the literature by Ruivo et al. [35] and Madhwesh et al. [36], which is in good agreement with the findings of the present study.

The combined influence of air velocity and RH indicates a trade-off in achieving thermal comfort. While increased air velocity can mitigate thermal stress in warm environments, excessive air movement in humid conditions may lead to discomfort rather than relief. High RH consistently impairs thermal comfort, as evidenced by its negative impact on both PMV and PPD. These findings highlight the need to maintain optimal RH levels while carefully adjusting air velocity to avoid overcooling and ensure occupant comfort.

As shown in Fig. 4, energy consumption increases consistently with rising air velocity across all levels of RH. In addition, at any given air velocity, higher RH leads to greater energy demand. The relationship between these variables is not merely linear; it reveals a compounding interaction where increases in both air velocity and RH intensify the total energy consumption. As air velocity increases, the power required to move air through ducts or ventilation systems also rises. Therefore, even modest increases in velocity can result in disproportionately higher energy usage. Additionally, as airflow increases, the system faces greater frictional and dynamic resistance within the ducts and air passages, further demanding higher mechanical effort and electrical input. Relative humidity plays a crucial role in influencing energy needs,

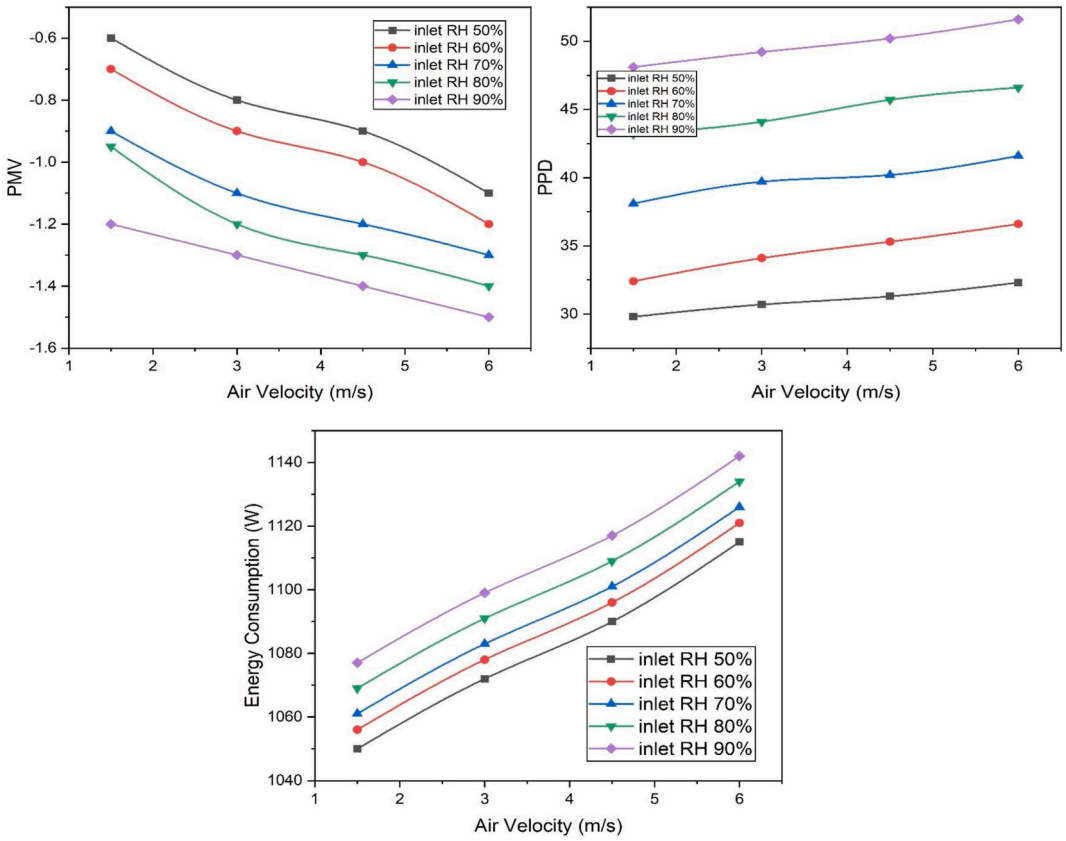


Fig. 4. Variation of PMV, PPD and Energy consumption with air velocity for different inlet RH values.

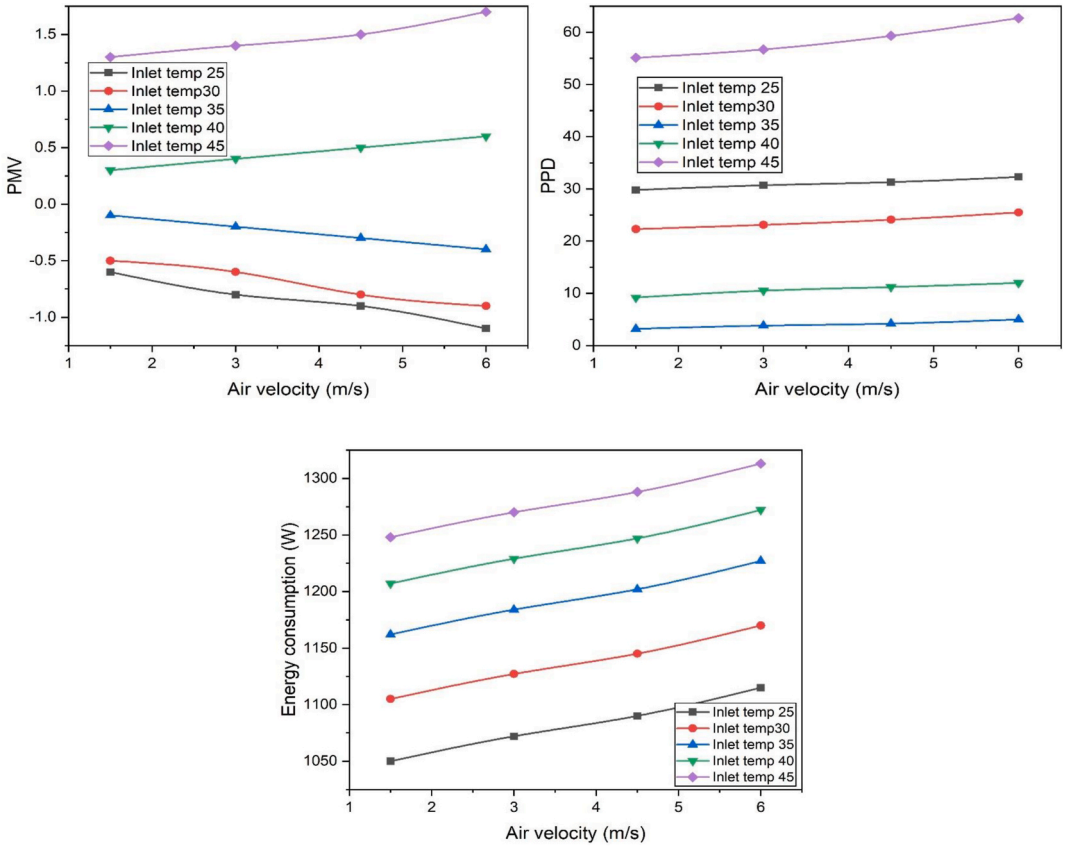


Fig. 5. Variation of PMV, PPD and Energy consumption with air velocity for different inlet temperature values.

particularly through its effect on thermal conditioning. Moist air contains more latent heat, which requires additional energy to remove during cooling and dehumidification processes. Higher RH also means the HVAC system must perform more condensation and often operate longer to achieve the desired indoor humidity levels, increasing the energy demands. The nature of RH's influence becomes even more pronounced at higher air velocities. The graph shows that energy consumption at 90% RH is significantly higher than at 50% RH for the same velocity, and this difference widens as air velocity increases. This indicates a compounding effect: high RH amplifies the energy required to condition air, and when combined with fast airflow, it escalates energy consumption even further.

A comparison of the present study with the works of Michalak [37] and Kim et al. [27] indicates that the power consumption of the proposed system is substantially lower than the values reported in the literature. The proposed unit achieves power savings of more than 50% when compared with these studies, demonstrating markedly improved energy performance. Consequently, the system exhibits superior energy efficiency and can be considered a sustainable alternative to conventional approaches.

Fig. 5 collectively illustrates the impact of air velocity and inlet temperature on thermal comfort indices—PMV, PPD and energy consumption. As observed, energy consumption consistently increases with both rising inlet temperature and air velocity, indicating a direct relationship where higher thermal loads and greater airflow demand more energy input. The PMV values rise with increasing inlet temperature, shifting from cooler to warmer sensations, and exhibit a variable trend with air velocity. At lower inlet temperatures, increasing air velocity reduces PMV and enhances thermal comfort. However, at higher inlet temperatures, PMV rises with air velocity, likely due to reduced cooling effectiveness. Similarly, PPD values increase with both air velocity and inlet temperature, especially at the extremes, suggesting that thermal dissatisfaction increases under hotter, faster airflow conditions. This indicates a trade-off between maintaining comfort and conserving energy, where optimizing inlet conditions is essential to balancing energy use with acceptable comfort levels. Optimization is essential in this context to balance the competing objectives of energy efficiency and thermal comfort. Without optimization, there is a risk of either excessive energy usage with marginal comfort gains or insufficient cooling performance leading to occupant dissatisfaction. An optimized operating point would identify the ideal combination of inlet temperature and air velocity that minimizes energy consumption while maintaining acceptable PMV and PPD values, ensuring occupant comfort is not sacrificed.

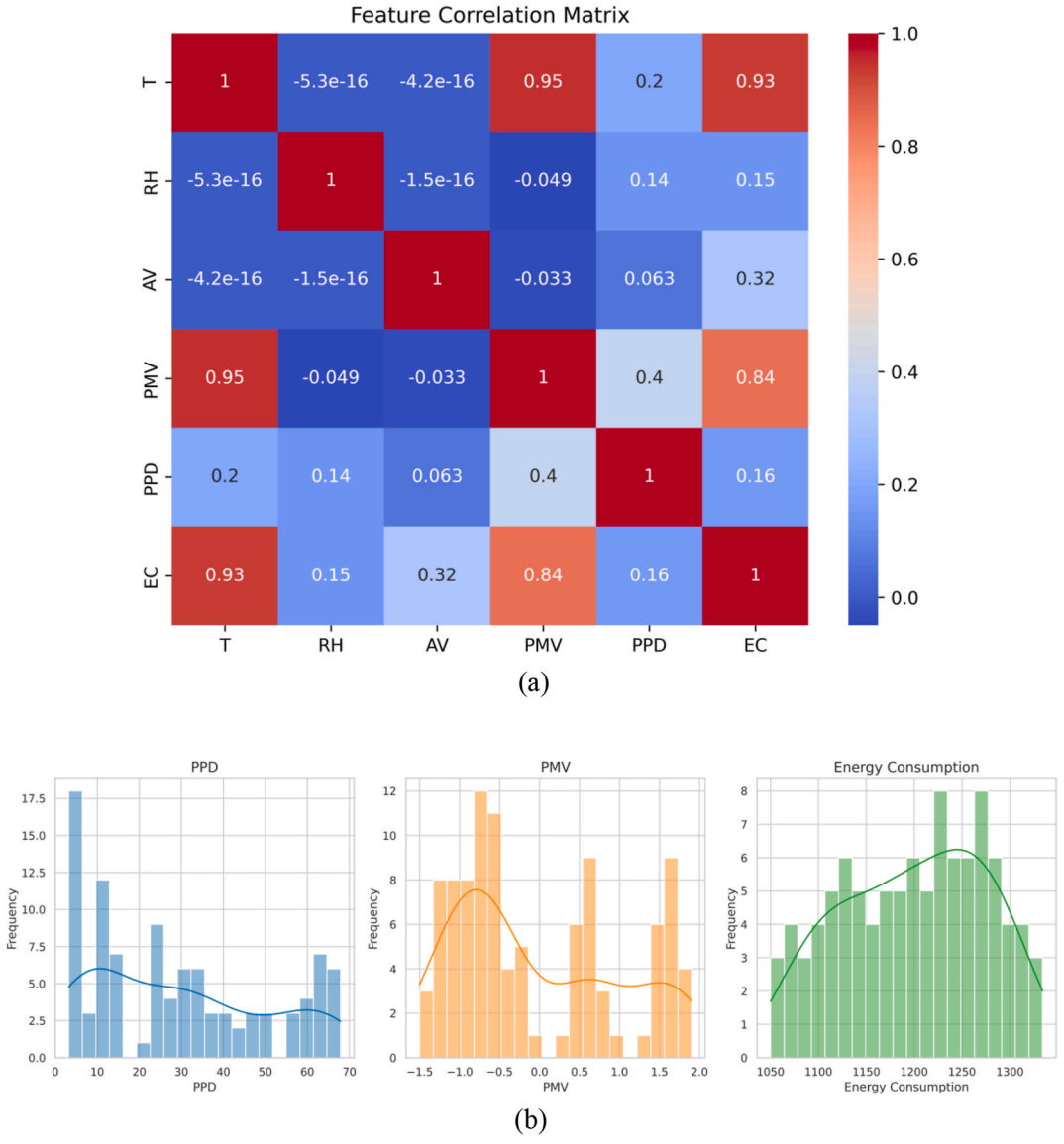


Fig. 6. Descriptive analytics of the data: (a) correlation heat map, (b) histogram of responses.

4.1. Exploratory data analysis

Exploratory data analysis (EDA) plays a pivotal role in understanding the relationships between input and output variables, identifying patterns, and assessing the overall distribution of data prior to model development. As part of this process, a feature correlation matrix and distribution plots of the response variables were constructed to gain insights into variable interactions and distributional characteristics. These visualizations provide a foundational understanding of the dataset, guiding subsequent modeling steps and helping to identify influential variables in optimizing thermal comfort and energy efficiency.

The correlation heatmap (Fig. 6) reveals strong positive correlations between Temperature (T) and both PMV (Predicted Mean Vote) and Energy Consumption (EC), with correlation coefficients of 0.95 and 0.93, respectively. This indicates that as indoor temperature increases, thermal comfort perception (PMV) and energy demand also tend to rise. Similarly, PMV shows a significant positive correlation with EC ($R = 0.84$), consistent with the expected increase in energy use to achieve higher comfort levels. In contrast, Relative Humidity (RH) and Air Velocity (AV) exhibit minimal correlation with other variables, suggesting weaker direct effects on the output responses. The moderate positive correlation between PMV and PPD (Predicted Percentage of Dissatisfied) ($R = 0.40$) is also noteworthy, as it reflects their interdependence in evaluating thermal comfort.

The histograms of the response variables Fig. 6 further enhance this understanding. PPD shows a skewed distribution with a peak at lower dissatisfaction values, indicating that most conditions yielded relatively comfortable environments. In contrast, PMV is spread across both negative and positive values, reflecting a mix of slightly cool to slightly warm thermal sensations across the dataset. The energy consumption histogram displays near-normal distribution with a slight right skew, suggesting that while most data points lie around a central energy demand, higher energy usage scenarios also occur with moderate frequency.

4.2. Machine learning (ML) model

All machine learning modeling and optimization tasks in this study were carried out using Google Colaboratory (Colab), a cloud-based Python development environment that provides free access to GPU/TPU-accelerated computation. Colab enabled seamless integration of data analysis, model development, and optimization in a single workflow, without the need for local hardware resources. The following relevant Python libraries were utilized:

- NumPy and Pandas for numerical operations and data manipulation.
- Scikit-learn for traditional machine learning models (Random Forest) and Artificial Neural Network (ANN), data preprocessing, and performance evaluation.
- XGBoost library for implementing Extreme Gradient Boosting (XGBoost) regressor.
- Matplotlib and Seaborn for visualizations such as correlation heatmaps and histograms.
- Optuna for automated hyperparameter optimization.
- Joblib for saving trained models for future inference.

The dataset consisted of 100 experimental samples, obtained from all combinations of inlet temperature, relative humidity, and air velocity specified in Table 1. For model development and validation, the dataset was partitioned into training and testing subsets using an 80:20 split ratio. A fixed random seed was applied to ensure reproducibility of the results across all machine learning models. Cross-validation was not employed because the dataset consisted of fixed experimental measurements rather than stochastic samples; instead, robustness was ensured through Optuna-based hyperparameter optimization and independent evaluation on the test set.

4.2.1. The OPTUNA framework

Hyperparameter tuning is a vital step in improving the performance and generalization of ML models. Rather than relying on manual trial-and-error or exhaustive grid search approaches, this study adopts OPTUNA, an advanced, open-source, hyperparameter optimization framework [38]. OPTUNA automates the search for optimal hyperparameters, leveraging efficient sampling strategies and pruning techniques to accelerate the discovery of high-performing configurations. OPTUNA allows dynamic search space definitions and leverages intelligent exploration to minimize computational costs while maximizing model performance. In this study, OPTUNA was employed to fine-tune three ML models: RandomForest Regressor, XGBoost Regressor, and Artificial Neural Network (ANN), each with distinct hyperparameter search spaces.

The workflow for OPTUNA-based hyperparameter optimization is illustrated in Fig. 7. The process begins with importing the thermal comfort dataset, which is then split into training and test subsets to facilitate model evaluation. Prior to optimization, Min-Max normalization is applied to ensure that input features are scaled within a common range, enhancing model convergence.

Subsequently, the OPTUNA optimization routine is invoked, which iteratively evaluates different hyperparameter combinations using the selected performance metric—in this case, the coefficient of determination (R^2 score). For each suggested trial, the corresponding machine learning model is trained and evaluated. The hyperparameter set yielding the highest R^2 score on the validation set is retained. Finally, the best-performing model configuration is selected and retrained using the full training data. This optimized model serves as the final model used for subsequent prediction and optimization tasks in the study.

4.2.2. Random Forest Regression

Random Forest Regression is a powerful ensemble learning method introduced by Breiman (2001) [39], designed to improve the predictive performance of individual decision trees by reducing overfitting and variance through the aggregation of multiple trees. This model is particularly effective for regression problems involving complex, nonlinear relationships and high-dimensional feature spaces.

The working principle of the Random Forest algorithm is illustrated in Fig. 8, where multiple decision trees are trained independently on bootstrap samples drawn from the original dataset. For a given input, each tree provides an independent prediction, and the final regression

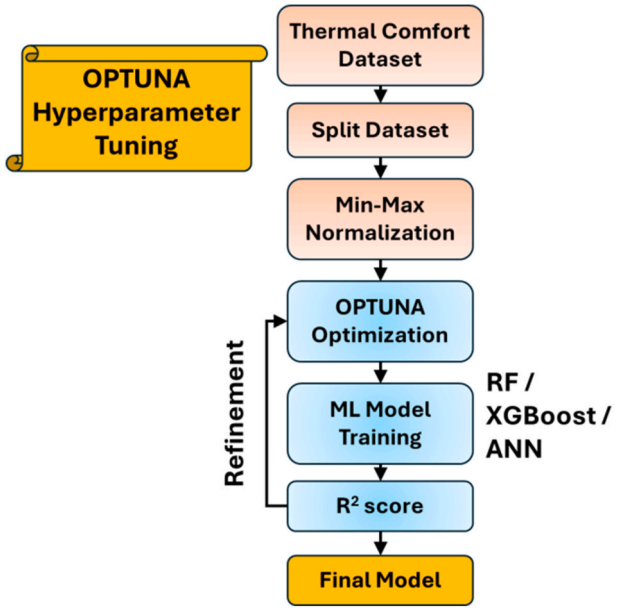


Fig. 7. Optimization of ML models using OPTUNA framework.

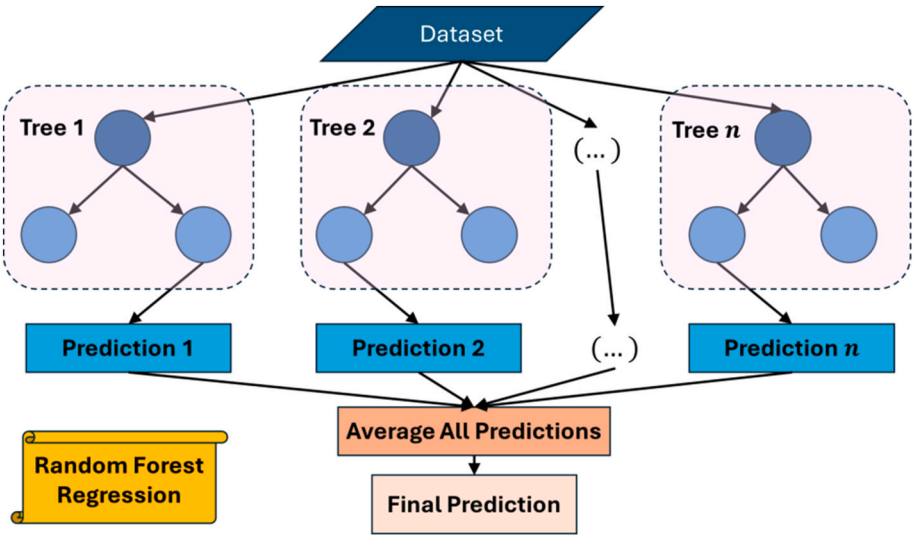


Fig. 8. Working principle of Random Forest Regression.

output is obtained by averaging the predictions from all trees in the ensemble.

Let the training dataset be represented as:

$$\mathscr{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}, \mathbf{x}_i \in \mathbb{R}^d, y_i \in \mathbb{R} \tag{8}$$

where $\mathbf{x}_i$ is the input vector of d features and y_i is the corresponding target value. The Random Forest model constructs B regression trees $\{T_b\}_{b=1}^{B}$, each trained on a bootstrap sample of the data. At each split in a tree, a random subset of features is selected to determine the best split, enhancing model diversity and reducing correlation between trees. The prediction for a new input $\mathbf{x}$ is given by [40]

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} T_b(\mathbf{x}) \tag{9}$$

This ensemble averaging reduces the model variance while maintaining low bias, resulting in robust generalization performance. In this study, key hyperparameters of the Random Forest model—namely the number of estimators ($n_{estimators}$) and the maximum depth of the trees (max_depth)—were optimized using the Optuna framework. The search ranges considered were: $n_{estimators} \in [50, 200]$ and $max\_depth \in [3, 20]$.

4.2.3. *Extreme Gradient Boosting (XGBoost) regression*

Extreme Gradient Boosting (XGBoost) is a highly efficient and scalable implementation of gradient boosting proposed by Chen and Guestrin (2016) [41]. It is designed to optimize both the computational speed and model performance, making it a preferred choice in many structured data problems. In XGBoost, a strong predictive model is built sequentially by adding new decision trees that correct the errors made by the ensemble of previous trees. Unlike Random Forests, where trees are built independently and averaged, in boosting, trees are added iteratively, with each new tree trained to predict the residuals (errors) of the existing ensemble (refer Fig. 9).

The objective function minimized by XGBoost is given by

$$\mathscr{L}(\phi) = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{i=1}^{N} \Omega(f_k) \tag{10}$$

where l is a differentiable convex loss function, namely, squared error. $\hat{y}_i$ is the prediction for instance i, f_k denotes an individual tree, $\Omega(f_k)$ is a regularization term controlling model complexity given by [42]

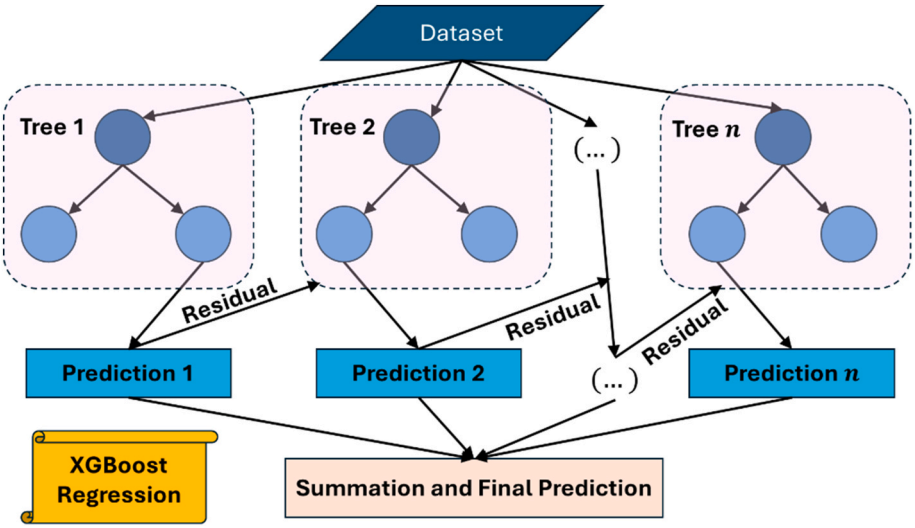


Fig. 9. Working principle of XGBoost Regression.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{11}$$

with T being the number of leaves and ω_j the score on each leaf. XGBoost employs second-order Taylor approximation of the loss function, leading to fast convergence and improved accuracy. Additionally, it includes techniques such as shrinkage (learning rate), column subsampling, and sparsity-aware split finding to further boost performance. In this study, key hyperparameters of XGBoost—namely the number of estimators ($n_{estimators}$), the maximum depth of each tree (*max_depth*), and the learning rate (*learning_rate*)—were tuned using the Optuna framework over the ranges: $n_{estimators} \in [50,\ 200]$, $max\_depth \in [3, 20]$ and $learning\_rate \in [0.01, 0.3]$.

4.2.4. Artificial neural network (ANN) regression

Artificial Neural Networks (ANNs) are a class of flexible function approximators inspired by the architecture of biological neural systems. They are particularly effective for capturing complex, nonlinear relationships between inputs and outputs in regression tasks. A feedforward artificial neural network (ANN) is generally structured with an input layer, multiple hidden layers, and an output layer. Within the network, each neuron processes information by calculating a weighted combination of inputs and applying a nonlinear activation function, allowing it to capture and represent complex relationships in the data. For an ANN with L layers, the output of layer l is given by [43]

$$\mathbf{a}^l = \sigma\left(\mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^l\right) \tag{12}$$

where $\mathbf{a}^{(l-1)}$ is the activation vector from the previous layer, $\mathbf{W}^{(l)}$ and $\mathbf{b}^l$ are weight matrix and bias vector of layer l, and σ is the activation function (e.g., ReLU, tanh) [44,45]. The network is trained to minimize a loss function, usually mean-squared error, for regression using the Adam optimizer.

In this work, a fully connected ANN architecture comprising two hidden layers was used. Neuron numbers and the initial learning rate were optimized using Optuna. Air velocity, inlet temperature and inlet relative humidity were the parameters in the input layer and PMV, PPD and EC were the parameters in the output layer. The search spaces were number of neurons in the first hidden layer: 32 to 128, number of neurons in the second hidden layer: 16 to 64, initial learning rate (*learning_rate_init*): 0.001 to 0.1. The ANN model is shown in Fig. 10. The hidden layers utilized the Rectified Linear Unit (ReLU) as their activation function, whereas the output layer implemented a linear activation function, which is appropriate for predicting continuous regression outputs.

4.2.5. Performance evaluation of ML models

The scatter plot in Fig. 11(a) compares predicted versus actual values for all outputs across the three models. The closeness of the points to the 45° line indicates strong predictive performance. The R^2 values, 0.99 for Random Forest, 1.00 for XGBoost, and 0.99 for ANN reflect excellent goodness of fit, with XGBoost achieving near-perfect prediction accuracy.

The bar charts in Fig. 11(b) show the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for each model. XGBoost demonstrates the lowest MAE and RMSE, confirming its superior predictive accuracy. Random Forest and ANN also achieve low errors but are slightly outperformed by XGBoost in both metrics.

The error histograms (refer Fig. 11(c)) for Random Forest, XGBoost, and ANN reveal that most prediction errors are centered around zero, indicating unbiased predictions. XGBoost exhibits the narrowest error spread, consistent with its lower MAE and RMSE, whereas Random Forest and ANN show slightly wider distributions, corresponding to their marginally higher error values. Overall, XGBoost emerges as the best-performing model among the three, with Random Forest and ANN providing competitive, yet slightly less accurate, results.

4.3. Multi-response optimization using NSGA-II algorithm

To address the multi-objective optimization problem of minimizing energy consumption (EC) while satisfying thermal comfort constraints on PMV and PPD, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) was employed. NSGA-II, introduced by Deb et al. (2002) [46], is a widely adopted evolutionary optimization algorithm capable of effectively handling problems involving conflicting objectives without requiring any user-defined weight assignments.

A schematic representation of the adopted NSGA-II framework tailored to the present study is illustrated in Fig. 12. The optimization process commences by defining the input parameters — air inlet temperature (T), inlet relative humidity (RH), and air velocity (AV), and initializing a random population of candidate solutions. The corresponding output parameters (PPD, PMV, and EC) are predicted using the trained XGBoost surrogate model. To ensure uniformity in objective scaling, EC values are normalized using min-max scaling

$$EC_{norm} = \frac{EC - EC_{min}}{EC_{max} - EC_{min}} \tag{13}$$

where EC_{min} and EC_{max} denote the minimum and maximum energy

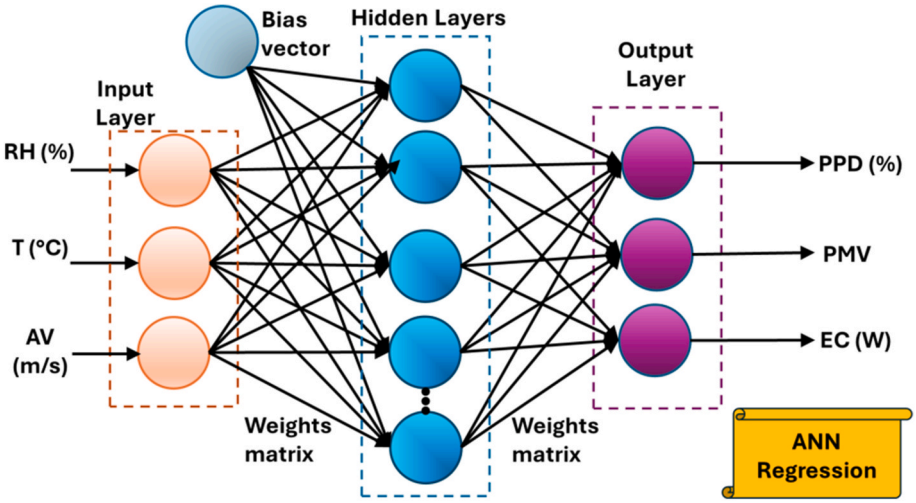


Fig. 10. General architecture of ANN Regression.

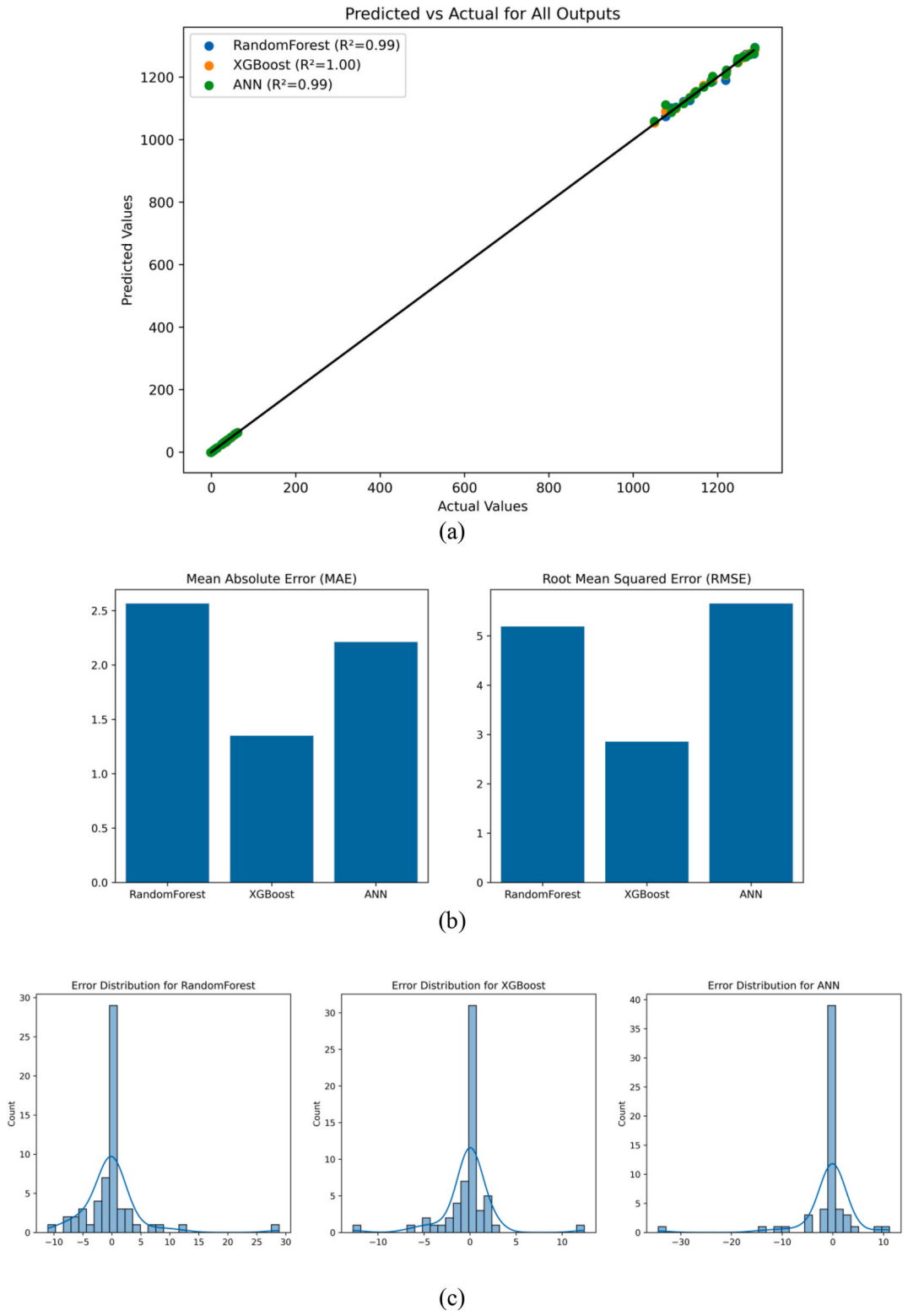


Fig. 11. (a) Correlation plot with R^2 scores, (b) comparison of MAE and RMSE, and (c) error distribution plot of the trained models.

consumption values predicted across the population, respectively.

Subsequently, hard constraints are applied: any solution with a PMV value outside the thermal comfort band ($-0.7 \leq PMV \leq +0.7$) or PPD greater than 25% is discarded. The feasible solutions undergo non-dominated sorting based on Pareto dominance principles. A solution A is said to dominate another solution B if: A is no worse that B in all objectives, and A is strictly better than B in at least one objective, which can be mathematically expressed as

$$A \prec B \quad if \quad \forall i : f_i(A) \leq f_i(A) \quad and \quad \exists j : f_j(A) < f_j(B) \tag{14}$$

where f_i and f_j represent the objective functions under consideration. The non-dominated fronts thus obtained are used to rank individuals, and a crowded-comparison operator ensures diversity by preferring individuals located in less crowded regions of the Pareto front. The

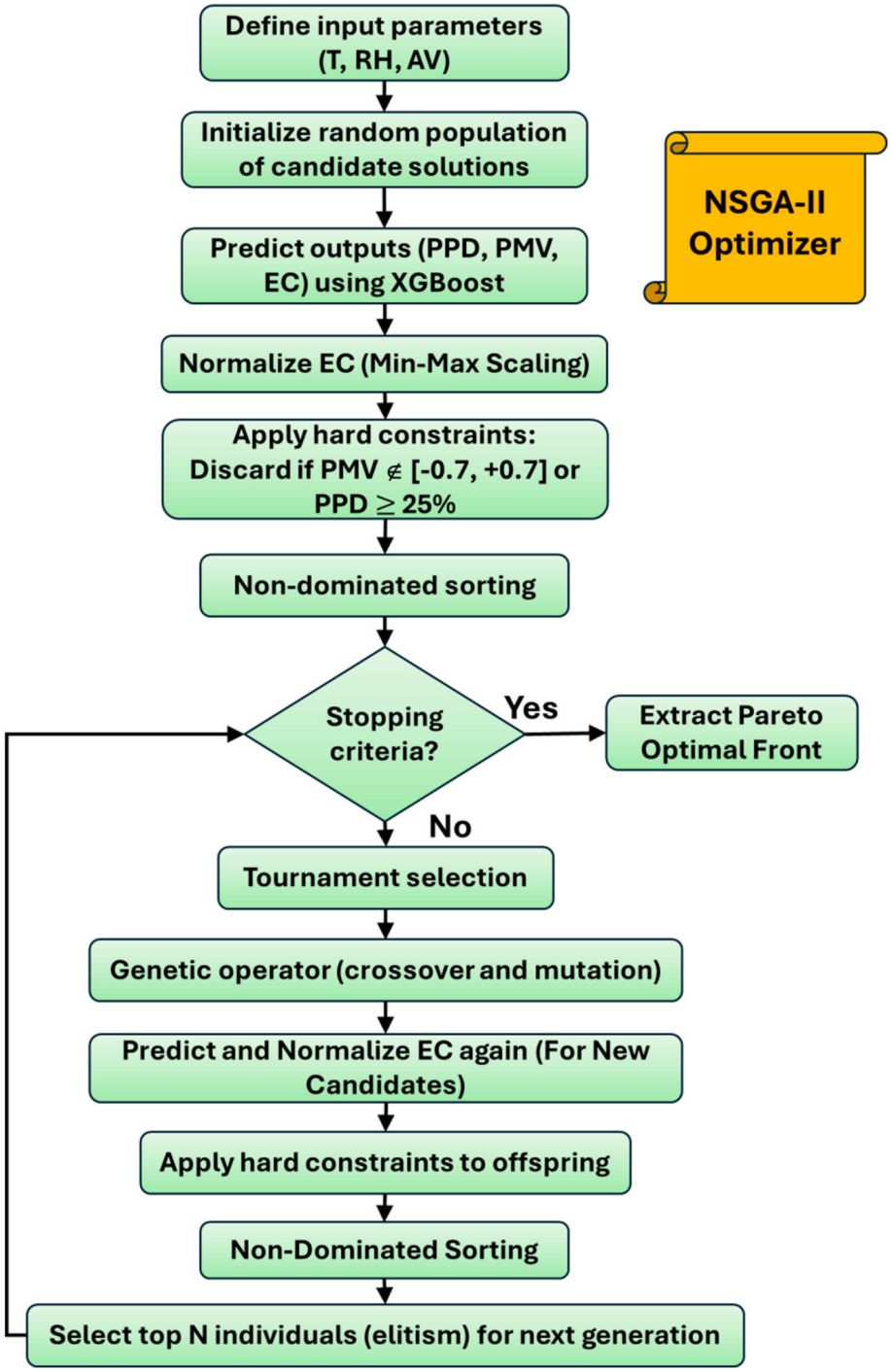


Fig. 12. Flowchart illustrating the NSGA-II based multi-objective optimization framework employed in this study.

crowded distance (d_i) for each solution is computed to quantify the density of solutions surrounding it

$$d_i = \sum_{m=1}^{M} \left(\frac{f_m^{i+1} - f_m^{i-1}}{f_m^{max} - f_m^{min}} \right) \tag{15}$$

where M is the number of objectives, and f_m^{i+1} and f_m^{i-1} denote the neighboring solutions in the sorted list according to the m-th objective.

Selection of individuals for reproduction is carried out using binary tournament selection based on rank and crowded distance. Genetic operators — simulated binary crossover (SBX) and polynomial mutation — are applied to generate offspring. The offspring population is then evaluated by predicting outputs and reapplying hard constraints. The combined parent and offspring populations undergo another round of non-dominated sorting, and elitism is ensured by selecting the best individuals to form the next generation. The optimization process iterates until a specified stopping criterion (e.g., maximum number of generations) is satisfied, after which the Pareto-optimal solutions are extracted.

Through the adoption of NSGA-II, the optimization process successfully navigates the conflicting objectives of maintaining thermal comfort and minimizing energy consumption without requiring subjective trade-offs. This approach aims to ensure a well-distributed Pareto front, offering decision-makers a spectrum of optimal solutions depending on their preferred balance between energy efficiency and

occupant comfort.

4.4. Self-adaptive NSGA-II algorithm

To enhance optimization performance, a self-adaptive version of NSGA-II was developed in this study. As shown in Fig. 13, the approach integrates hyperparameter optimization into the evolutionary process. Before the standard NSGA-II steps commence, critical parameters such as population size, crossover probability, and mutation probability are tuned automatically using the Optuna library based on the NSGA-II technique, ensuring that the algorithm is better tailored to the problem landscape.

After defining the input parameters, a random initial population is generated, and the trained XGBoost model predicts the output responses (PPD, PMV, and energy consumption). Energy consumption values are normalized using a min-max approach to maintain balance across ob-

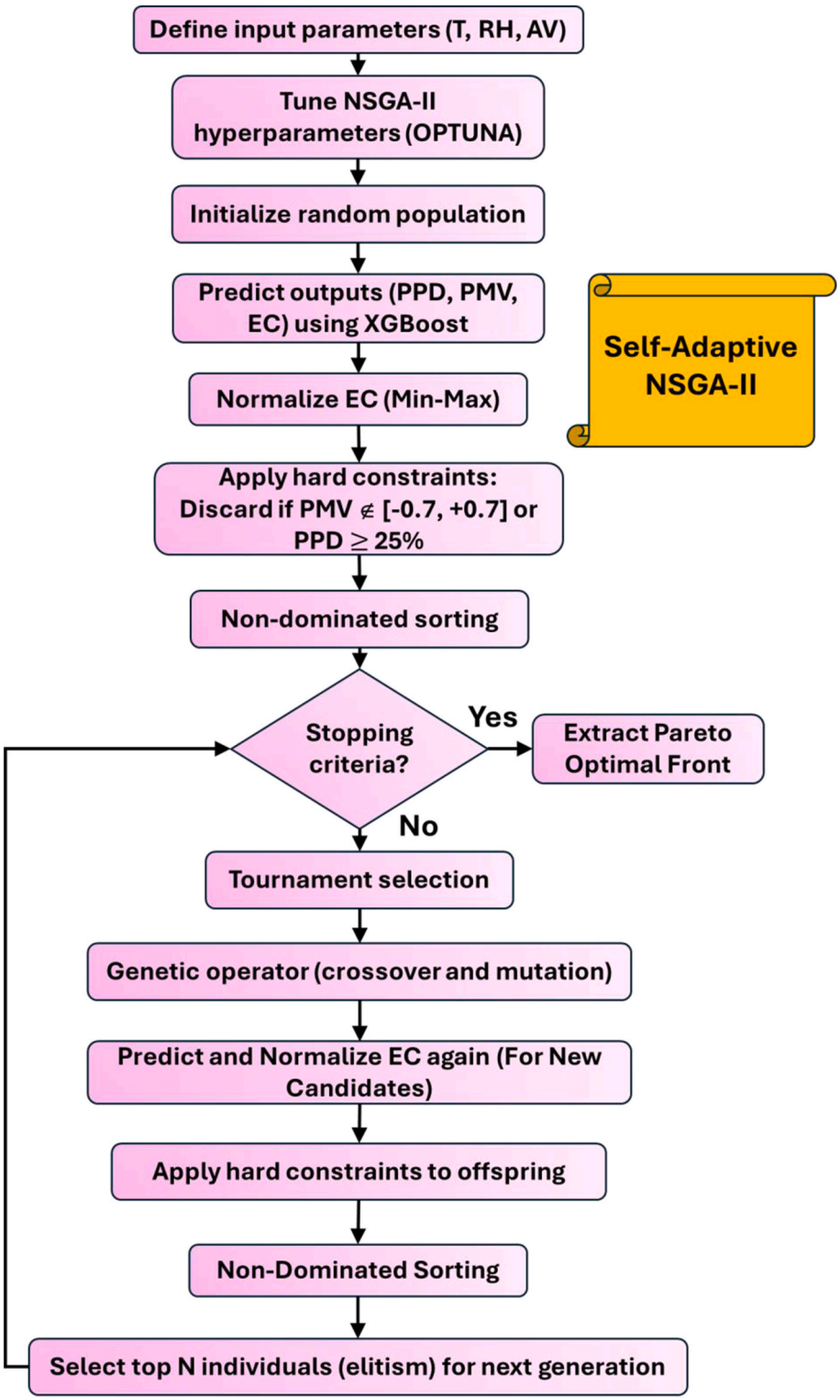


Fig. 13. Flowchart illustrating the self-adaptive NSGA-II technique employed in this study.

jectives. Hard constraints are enforced by discarding candidates with PMV values outside the acceptable range $[-0.7, +0.7]$ and PPD exceeding 25%. These thresholds represent the acceptable bounds for maintaining optimal indoor thermal comfort.

The evolutionary process follows the standard NSGA-II structure: non-dominated sorting, tournament selection, crossover, mutation, and elitism. However, by starting with tuned hyperparameters, the search process achieves better convergence and maintains higher diversity across generations. The final output is a refined Pareto-optimal front balancing thermal comfort and energy efficiency. This self-adaptive strategy eliminates manual hyperparameter selection and improves the robustness of the optimization, making it highly suited for complex multi-objective problems.

To comprehensively assess the effectiveness of the proposed self-adaptive NSGA-II framework, a comparison was performed with a manually tuned NSGA-II approach. In the manual tuning, multiple combinations of hyperparameters were explored by varying the population size (100, 200), crossover probability (0.1, 0.5, 0.9), and mutation probability (0.1, 0.5, 0.9). For each combination, the optimization process was carried out, and the resulting Pareto fronts were analyzed. The optimal solutions from each case were identified based on two criteria: (i) the solution corresponding to the minimal energy consumption (EC) while satisfying the imposed thermal comfort constraints, and (ii) the solution nearest to the ideal point in the objective space, determined using the Euclidean distance method. Subsequently, the performance of the manual approach was compared with that of the self-adaptive NSGA-II, highlighting the superior robustness and efficiency of the latter in converging to high-quality solutions.

To objectively identify the optimal solution from the set of Pareto-optimal candidates, the Euclidean distance (E_d) metric was utilized. This metric measures the geometric distance of a candidate solution from an ideal reference point in the multi-objective space. The Euclidean distance is calculated as

$$E_d = \sqrt{\left(\frac{PPD - PPD_{ideal}}{PPD_{range}}\right)^2 + \left(\frac{PMV - PMV_{ideal}}{PMV_{range}}\right)^2 + \left(\frac{EC - EC_{ideal}}{EC_{range}}\right)^2} \tag{16}$$

where PPD_{ideal}, PMV_{ideal}, and EC_{ideal} represent the ideal values of the respective objective (here $PPD_{ideal} = 0\%$, $PMV_{ideal} = 0$, and $EC_{ideal} =$ minimum value from the experiments). The denominators PPD_{range}, PMV_{range}, and EC_{range} represent the normalization factors based on the observed ranges of the respective objectives to ensure scale invariance. A lower E_d value indicates a solution that is closer to the ideal conditions, thus providing a systematic and quantitative criterion for selecting the most balanced trade-off solution among multiple objectives.

The optimization results obtained using both the manual and self-adaptive variants of the NSGA-II algorithm are summarized in Table 2. In the manual approach, optimization was performed by specifying fixed hyperparameter combinations, namely population sizes of 100 and 200, crossover probabilities of 0.1, 0.5, and 0.9, and mutation probabilities of 0.1, 0.5, and 0.9. Each combination generated a different set of Pareto-optimal solutions, subsequently evaluated based on thermal comfort metrics, PPD, PMV, and EC.

It was observed that several manually tuned settings, such as (Population size = 100, Crossover = 0.1, Mutation = 0.1) and (Population size = 200, Crossover = 0.9, Mutation = 0.1), resulted in excellent thermal comfort (PPD~3.80–4.20%) while maintaining reasonable energy consumption (~1160–1184 W). However, despite these promising outcomes, significant variations were seen across different parameter combinations, indicating the strong dependence of the manual approach on specific hyperparameter choices. Furthermore, no single manually defined setting consistently minimized both energy consumption and thermal discomfort simultaneously.

In contrast, the self-adaptive NSGA-II approach demonstrated superior performance by dynamically adjusting the crossover and mutation probabilities during the optimization process. The optimal solution identified through the self-adaptive method exhibited a PPD of 22.29%, a PMV of −0.50, and a significantly lower energy consumption of 1104.99 W compared to most manually optimized results. A comparison of the present study with earlier works reported in the literature, such as Michalak [48] and Kim et al. [49], indicates that the power consumption of the proposed system is considerably lower than that of conventional systems. The observed reduction corresponds to substantial energy savings, reflecting a clear improvement in overall energy performance and highlighting the efficiency advantage of the proposed approach. More importantly, this solution achieved the minimum Euclidean distance (E_d) value among all candidates, reflecting the best trade-off between thermal comfort and energy efficiency. Thus, the self-adaptive approach not only reduced the manual effort involved in hyperparameter selection but also consistently converged towards better quality solutions. These findings affirm that incorporating self-adaptive mechanisms within evolutionary algorithms can greatly enhance the robustness and effectiveness of optimization in complex multi-objective problems, particularly when compared to fixed manual parameter tuning.

4.5. Validation of the optimized results

Experiments were conducted on a test rig using input air conditions predicted by the self-adaptive NSGA-II optimization method. The resulting output parameters—PMV, PPD, and energy consumption—were measured and are presented in Fig. 14. The comparison between the optimized and experimental results highlights the effectiveness of the machine learning-based optimization approach in achieving thermal comfort while reducing energy consumption. The optimized PMV of −0.5 and PPD of 22.3% fall within the acceptable comfort thresholds (PMV ∈ [−0.7, +0.7], PPD ≤25%) and closely match

Table 2
Comparison of optimal solutions derived from manual hyperparameter tuning and self-adaptive NSGA-II based on thermal comfort metrics and operating conditions.

Approach	NSGA-II Hyperparameters			Thermal Comfort Metrics			Operating Parameters			E_d
	Pop. size	Crossover prob.	Mutation prob.	PPD (%)	PMV	EC (W)	T (°C)	RH (%)	AV (m/s)	
Random	100	0.10	0.10	3.78	−0.21	1163.00	37.49	67.86	1.53	58.13
	100	0.50	0.50	23.51	−0.59	1108.99	32.39	62.43	2.44	23.85
	100	0.90	0.90	24.59	−0.69	1131.02	30.59	67.81	3.30	35.81
	100	0.10	0.90	3.81	−0.20	1183.99	35.36	52.16	4.29	79.09
	100	0.90	0.10	4.22	−0.30	1170.98	39.68	70.84	2.94	66.12
	200	0.10	0.10	11.18	0.49	1220.99	44.57	75.98	2.49	116.5
	200	0.50	0.50	23.10	−0.60	1126.99	30.98	57.99	4.02	31.91
	200	0.90	0.90	23.10	−0.60	1126.99	32.80	54.92	4.18	31.91
	200	0.10	0.90	4.67	−0.39	1193.02	38.28	79.22	3.99	88.15
	200	0.90	0.10	3.81	−0.20	1184.02	38.92	57.55	3.58	79.12
Self-Adaptive	150	0.70	0.46	22.29	−0.50	1104.99	30.94	50.44	2.67	22.29

* Pop. size: population size, Crossover prob.: crossover probability, Mutation prob.: mutation probability.

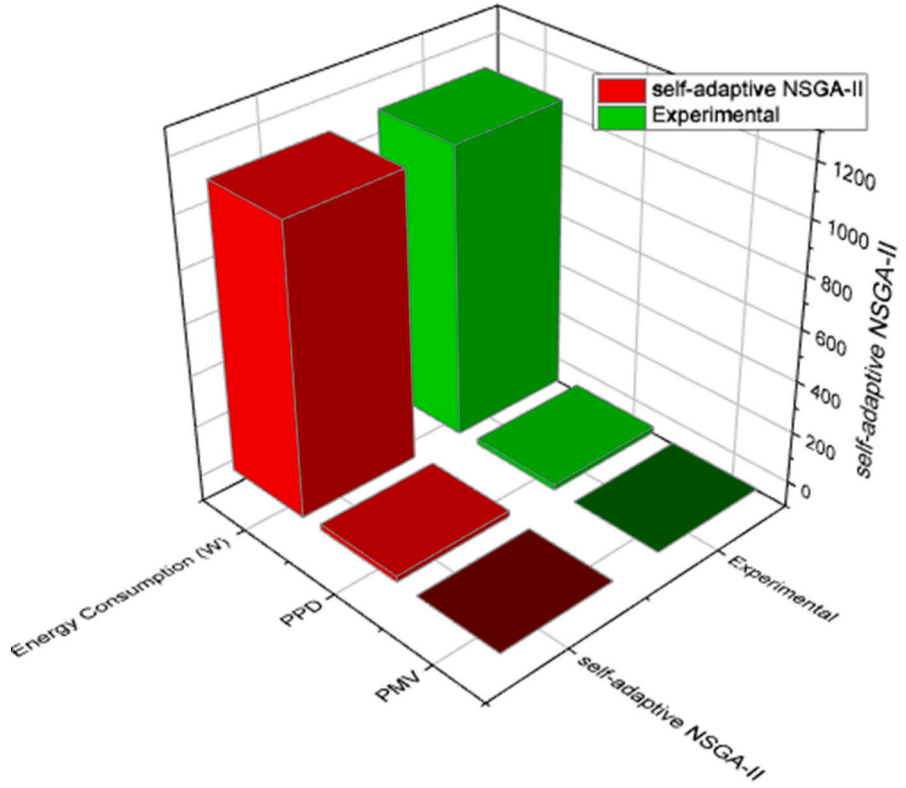


Fig. 14. Comparision of Experimental and NSGA-II optimized Parameters.

the experimental outcomes (PMV = −0.54, PPD = 21.7%). The relative errors between the experimental and predicted values are minimal: 0.8% for energy consumption, 1.2% for PMV, and 2.7% for PPD. These results confirm the reliability and accuracy of the self-adaptive NSGA-II method as a robust tool for optimizing energy-efficient HVAC operations while maintaining thermal comfort.

5. Conclusions

An AHU test rig has been fabricated to investigate the combined effects of air velocity, inlet temperature, and inlet relative humidity on thermal comfort indices—specifically PMV (Predicted Mean Vote) and PPD (Predicted Percentage of Dissatisfied)—as well as on energy consumption. It was observed that PMV decreased with increasing air velocity, regardless of relative humidity (RH). At any given air velocity, higher inlet RH led to lower PMV values. The PPD index, however, gradually increased with rising air velocity across all RH levels. Similarly, an increase in RH resulted in higher PPD values across all air velocities. Energy consumption was found to increase consistently with both higher inlet temperature and air velocity, indicating a direct relationship in which greater thermal loads and higher airflow rates demand greater energy input.

This study demonstrated the effectiveness of a machine-learning-integrated optimization framework for improving the performance of Air Handling Units (AHUs) in both thermal comfort and energy efficiency. Using a self-adaptive NSGA-II algorithm in conjunction with an ML model, Extreme Gradient Boosting (XGBoost), optimal inlet air conditions—temperature, velocity, and humidity—were predicted and optimized under strict comfort constraints.

Experimental validation was carried out using a specially designed test rig configured with the optimized input parameters. The observed output values—PMV of −0.54, PPD of 21.7%, and energy consumption of 1114 W—closely matched the predicted optimized values of PMV -0.5, PPD 22.3%, and energy consumption 1105 W. The relative prediction errors were low: 0.8% for energy consumption, 1.2% for PMV, and 2.7% for PPD, confirming the high accuracy and reliability of the ML-based model.

Overall, the findings validate that the proposed method can effectively maintain desired thermal comfort conditions while achieving a measurable reduction in energy consumption. The self-adaptive NSGA-II framework, combined with predictive machine learning models, provides a robust, scalable solution for smart, energy-efficient HVAC control systems. This approach directly supports Sustainable Development Goals such as SDG 7 (Affordable and Clean Energy) by enhancing energy efficiency, and SDG 13 (Climate Action) by contributing to the reduction of greenhouse gas emissions through optimized energy use.

6. Limitations and scope for future work

Although the proposed self-adaptive NSGA-II optimization framework demonstrates strong predictive accuracy and effective performance enhancement for the laboratory-scale AHU setup, several limitations remain that suggest avenues for continued advancement. First, the current model has been validated only on a controlled test rig, which does not fully capture the complexities of full-scale building HVAC systems involving multi-zone thermal interactions, varying occupancy levels, duct losses, and external environmental influences. As a result, the direct scalability of the model to entire buildings remains uncertain and requires additional system-level studies. The present study focuses on steady-state inlet air conditions; however, real-world buildings experience rapid fluctuations in temperature, humidity, and occupancy, necessitating the development of models capable of handling dynamic, time-varying conditions. Sensor integration is another area with inherent constraints, as the current setup utilizes a fixed and limited array of sensors, while real deployment scenarios would require additional inputs such as CO_2 concentration, volatile organic compounds and indoor air quality indices, and occupant density for more comprehensive optimization.

Future work should therefore focus on developing real-time control integration, where the trained machine learning models and the self-adaptive NSGA-II algorithm can be embedded within control systems or microcontroller platforms to optimize AHU performance under continuously changing conditions. Incorporating Internet of Things (IoT) connectivity represents another promising direction, enabling cloud-based optimization, remote monitoring, and data-driven adaptive control strategies that can respond to multi-zone feedback. Furthermore, expanding the thermal comfort modeling framework to incorporate additional metrics such as adaptive comfort models, draft sensitivity, vertical temperature gradients, or comprehensive indoor air quality indices would support a more holistic optimization strategy that balances energy efficiency with occupants' well-being.

Nomenclature:

Variables/Parameters:	
$\dot{m}_a$	Air flow rate, kg/s
$\dot{m}_c$	Condensation rate or moisture removal rate, g/s
t_{cl}	Clothing surface temperature
t_r	Mean radiant temperature,
t_a	Ambient temperature
η_H	Humidification efficiency, %
C_{res}	Convection heat transfer rate, W/m^2
E_c	Heat released by the evaporation of the moisture from skin, W/m^2
E_{Blower}	Energy consumed by blower, W
$E_{Compressor}$	Energy consumed by compressor, W
$E_{Humidifier}$	Energy consumed by humidifier, W
E	Total energy consumed, W
E_{res}	Heat released due to breathing, W/m^2.
M	Rate of metabolism, W/m^2
H	Sensitive heat loses
w	Total Work done, W
W	Mechanical power, W/m^2
Δh	Change in specific enthalpy, kJ/kg
h_c	Heat transfer coefficient
P_a	Vapor partial pressure
Abbreviations:	
ANN	Artificial neural network
ACS	Air-conditioned space
AHU	Air handling unit
BMS	Building Management System
CFD	Computational fluid dynamics
DAT	Discharge air temperature
DBT	Dry bulb temperature
FAC	Fresh air control
HVAC	Heating, ventilation and airconditioning
MOGA	Multi-objective genetic algorithm
NSGA-II	Non-Dominated Sorting Genetic Algorithm II
NMPC	Nonlinear predictive control
OA	Outdoor Air
PMV	Predicted mean vote
PPD	Predicted percentage dissatisfaction
PID	Proportional–integral–derivative
RH	Relative humidity
RL	Reinforcement learning
VCD	Volume Control Damper
VCR	Vapor compression refrigeration
WBT	Wet Bulb Temperature
Subscripts:	
a	Air
H	Humidification
cl	Clothing
c	Condensation
r	radiant
res	Heat exchange rate
Chemicals:	
CO	Carbon Monoxide
CO_2	Carbon dioxide
Sensors:	
DHT11	Temperature and Humidity sensor
GP2Y1010AU0F	Dust sensor
MQ135	CO and CO2 sensor
GY-906 MLX90614ESF	Presence sensor
HC-SR501	Motion sensor

(*continued on next column*)

(*continued*)

CRediT authorship contribution statement

Sampath Suranjan Salins: Investigation, Data curation. **Shiva Kumar:** Writing – review & editing, Writing – original draft. **Subraya Krishna Bhat:** Validation, Supervision, Software.

Declaration of competing interest

I am here by submitting the manuscript titled "Smart HVAC Optimization Using Machine Learning and Self-Adaptive NSGA-II for Energy-Efficient Thermal Comfort" There is no conflict of interest among the authors.

Acknowledgements

The authors would like to express their gratitude to School of Engineering and Information Technology (SOEIT), Manipal Academy of Higher Education (MAHE) Dubai, for providing the laboratory facilities needed for the experimentation and simulation of the AHU unit.

Data availability

Data will be made available on request.

References

[1] Zulkafli NI, Sukri MF, Tahir MM, Muhajir A, Hanak DP. Performance analysis and optimisation of the chiller-air handling units system with a wide range of ambient temperature. Clean Eng Technol 2023;14:100643.
[2] Chen Y, Xu J, Wang J, Lund PD. Optimization of a weather-based energy system for high cooling and low heating conditions using different types of water-cooled chiller. Energy 2022;252:124094.
[3] Salins SS, Kumar SS, Thommana AJJ, Vincent VC, Tejero-González A, Kumar S. Performance characterization of an adaptive-controlled air handling unit to achieve thermal comfort in Dubai climate. Energy 2023;273:127186.
[4] Lopes J, Silva J, Teixeira S, Teixeira J. Numerical modeling and optimization of an air handling unit. Energies 2020;14(1):68.
[5] Aziz M, Kadir K, Azman HK, Vijyakumar K. Optimization of air handler controllers for comfort level in smart buildings using nature inspired algorithm. Energies 2023;16(24):8064.
[6] Gunay HB, Newsham G, Ashouri A, Wilton I. Deriving sequences of operation for air handling units through building performance optimization. J Buil Perform Simul 2020;13(5):501–15.
[7] Stasi R, Ruggiero F, Berardi U. Influence of cross-ventilation cooling potential on thermal comfort in high-rise buildings in a hot and humid climate. Build Environ 2024;248:111096.
[8] Ferdyn-Grygierek J, Grygierek K, Gumińska A, Krawiec P, Oćwieja A, Poloczek R, Szkarłat J, Zawartka A, Zobczyńska D, Żukowska-Tejsen D. Passive cooling solutions to improve thermal comfort in Polish dwellings. Energies 2021;14(12):3648.
[9] Suranjan Salins S, Kota Reddy SV, Kumar S. Experimental evaluation of a multistage reciprocating evaporative cooler performance for varied climatic conditions. J Therm Sci 2023;32(4):1523–35.
[10] Kumar R, Khatri SK, Diván MJ. Data center air handling unit fan speed optimization using machine learning techniques. In: 2021 9th international conference on reliability, infocom technologies and optimization (trends and future Directions)(ICRITO). IEEE; 2021, September. p. 1–10.
[11] Wani M, Hafiz F, Swain A, Ukil A. A multi-objective approach to robust control of air handling units for optimized energy performance. Electronics 2023;12(3):661.
[12] Lin CM, Liu HY, Tseng KY, Lin SF. Heating, ventilation, and air conditioning system optimization control strategy involving fan coil unit temperature control. Appl Sci 2019;9(11):2391.
[13] Lee D, Jeong J, Chae YT. Application of deep reinforcement learning for proportional–integral–derivative controller tuning on air handling unit system in existing commercial building. Buildings 2023;14(1):66.
[14] Razban A, Edalatnoor A, Goodman D, Chen J. Energy optimization of air handling unit using CO2 data and coil performance. In: ASME international mechanical engineering congress and exposition, 50596. American Society of Mechanical Engineers; 2016, November. V06BT08A007.

[15] Shang F, Ji Y, Peng W, Duan J, Yin Y. Tracking controller design for air handling units with uncertainties. In: Building simulation, 16. Beijing: Tsinghua University Press; 2023, April. p. 547–56. 4.
[16] Hosamo H, Hosamo MH, Nielsen HK, Svennevig PR, Svidt K. Digital Twin of HVAC system (HVACDT) for multiobjective optimization of energy consumption and thermal comfort based on BIM framework with ANN-MOGA. Adv Build Energy Res 2023;17(2):125–71.
[17] Casillas A, Chen Y, Granderson J, Lin G, Chen Z, Wen J, Huang S. Development of high-fidelity air handling unit fault models for FDD innovation: lessons learned and recommendations. J Buil Perform Simul 2024:1–16.
[18] Zheng Y, Kalbasi R, Karimipour A, Liu P, Bach QV. Introducing a novel air handling unit based on focusing on turbulent exhaust air energy-exergy recovery potential. J Energy Resour Technol 2020;142(11):112108.
[19] Ambroziak A, Chojecki A. The PID controller optimisation module using Fuzzy Self-Tuning PSO for Air Handling Unit in continuous operation. Eng Appl Artif Intell 2023;117:105485.
[20] Li D, Ahmat M, Cao H, Di F. Optimization of double-closed-loop control of variable-air-volume air-conditioning system based on dynamic response model. Buildings 2024;14(3):677.
[21] Yao L, Huang JH. Multi-objective optimization of energy saving control for air conditioning system in data center. Energies 2019;12(8):1474.
[22] Talib R, Nabil N, Choi W. Optimization-based data-enabled modeling technique for HVAC systems components. Buildings 2020;10(9):163.
[23] Kostyák A, Szekeres S, Csáky I. Investigation of sensible cooling performance in the case of an air handling unit system with indirect evaporative cooling: indirect evaporative cooling effects for the additional cooling system of buildings. Buildings 2023;13(7):1800.
[24] Bareschino P, Pepe F, Roselli C, Sasso M, Tariello F. Desiccant-based air handling unit alternatively equipped with three hygroscopic materials and driven by solar energy. Energies 2019;12(8):1543.
[25] Gao Y, Shi S, Miyata S, Akashi Y. Successful application of predictive information in deep reinforcement learning control: a case study based on an office building HVAC system. Energy 2024;291:130344.
[26] Samadi N, Shahbakhti M. Energy efficiency and optimization strategies in a building to minimize airborne infection risks. Energies 2023;16(13):4960.
[27] Kim JH, Seong NC, Choi W. Forecasting the energy consumption of an actual air handling unit and absorption chiller using ANN models. Energies 2020;13(17):4361.
[28] Hao J, Yang X, Wang C, Tu R, Zhang T. An improved NSGA-II algorithm based on adaptive weighting and searching strategy. Appl Sci 2022;12(22):11573. https://doi.org/10.3390/app122211573.
[29] Zhao Y. Adaptive strategy-enhanced NSGA-II for multi-objective optimization with improved convergence and diversity control. Informatica 2025;49(23). https://doi.org/10.3390/10.31449/inf.v49i23.8270.
[30] Ma H, Zhang Y, Sun S, Liu T, Shan Y. A comprehensive survey on NSGA-II for multi-objective optimization and applications. Artif Intell Rev 2023;56(12):15217–70. https://doi.org/10.1007/s10462-023-10526-z.
[31] Soleimani M, Sa'adati H. Multi-objective optimization of energy consumption and thermal comfort in smart HVAC systems using metaheuristic algorithms. Sci J Res Stud Fut Mech Eng 2024;2(1):19–24. https://journalhi.com/mec/article/view/287/332.
[32] Latifah A, Supangkat SH, Leksono E, Indraprastha A. Multi-objective optimization for balancing thermal comfort and energy efficiency using genetic algorithms within a digital twin model. J Eur Systèmes Automatisés 2025;58(5). https://doi.org/10.18280/jesa.580506.
[33] Shakya P, Ng G, Zhou X, Wong YW, Dubey S, Qian S. Thermal comfort and energy analysis of a hybrid cooling system by coupling natural ventilation with radiant and indirect evaporative cooling. Energies 2021;14(22):7825.
[34] Dyvia HA, Arif C. Analysis of thermal comfort with predicted mean vote (PMV) index using artificial neural network. In: IOP conference series: earth and environmental science, 622. IOP Publishing; 2021, 012019. 1.
[35] Ruivo CR, da Silva MG, Broday EE. Study on thermal comfort by using an atmospheric pressure dependent predicted mean vote index. Build Environ 2021;206:108370.
[36] Madhwesh N, Salins SS, Kumar S. Impact of climate conditions on the performance of adaptive air handling unit. Int J Thermofluids 2025;27:101220.
[37] Michalak P. Annual energy performance of an air handling unit with a cross-flow heat exchanger. Energies 2021;14(6):1519.
[38] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining; 2019, July. p. 2623–31.
[39] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[40] Hastie T. The elements of statistical learning: data mining, inference, and prediction. 2009.
[41] Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016, August. p. 785–94.
[42] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001:1189–232.
[43] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw 1989;2(5):359–66.
[44] Bhat SK, Deepak GD. Predictive modelling and optimization of double ring electrode based cold plasma using artificial neural network. Int J Eng 2024;37(1):83–93.
[45] Bhat SK, Ranjan A, Upadhyaya YS, Managuli V. Fatigue strength prediction of cobalt alloys using material composition and monotonic properties: ML-based approach. Mater Res Express 2025;12(4):046505.
[46] Deb K, Pratap A, Agarwal S, Meyarivan TAMT. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 2002;6(2):182–97.