frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 06 March 2025
DOI 10.3389/fenrg.2025.1473946

Check for updates

OPEN ACCESS

EDITED BY
Maria Cristina Piccirilli,
University of Florence, Italy

REVIEWED BY
Rudrodip Majumdar,
National Institute of Advanced Studies, India
Ali Fguiri,
Gabes University, Tunisia

*CORRESPONDENCE
Fayaz Hussain,
fayaz@tdtu.edu.vn

RECEIVED 30 August 2024
ACCEPTED 14 February 2025
PUBLISHED 06 March 2025

CITATION
Mujtaba MA, Munir MA, Akhtar M, Mahmood B, Ansar T, Khawar Z, Khalid S, Basit A, Jamil S, Kalam MA, Hussain F and Bhowmik C (2025) Leveraging machine learning to optimize cooling tower efficiency for sustainable power generation. *Front. Energy Res.* 13:1473946. doi: 10.3389/fenrg.2025.1473946

Leveraging machine learning to optimize cooling tower efficiency for sustainable power generation

M. A. Mujtaba[1], Muhammad Adeel Munir[1], Muhammad Akhtar[2], Bilal Mahmood[3], Talha Ansar[1], Zeeshan Khawar[1], Shayan Khalid[1], Abdul Basit[1], Saud Jamil[1], M. A. Kalam[4], Fayaz Hussain[5]* and Chiranjib Bhowmik[6]

[1]Department of Mechanical, Mechatronics and Manufacturing Engineering (New Campus), University of Engineering and Technology (UET), Lahore, Pakistan, [2]Quaid e Azam Thermal Power (Pvt) Limited, 1180MW CCPP Bhikki, Lahore, Pakistan, [3]Harbin Electric Company Limited, 1180MW CCPP Bhikki, O&M, Lahore, Pakistan, [4]School of Civil and Environmental Engineering, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW, Australia, [5]Modeling Evolutionary Algorithms Simulation and Artificial Intelligence, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam, [6]Department of Mechanical Engineering, Techno India University, West Bengal, India

United Nations Sustainable Development Goal 7 is about ensuring access to clean and affordable energy, which is a key factor in the development of society. The power generation sector majorly consists of thermal power plants. Cooling towers are a significant part of any power plant to cool steam to be reused again. Hence, the efficiency of power plants can be increased by optimizing the performance of cooling towers. This research paper aims to increase the efficiency of cooling towers by investigating the effect of ambient parameters (changing with climate) on the efficiency of cooling towers for the best site selection. Ambient parameters cannot be controlled after the installation of power plants. Therefore, proper site selection, keeping ambient parameters and their expected change before the installation of power plants, effectively increases the efficiency of the cooling tower and, ultimately, the power plant. For this purpose, data is collected from the 1140 MW combined cycle power plant in Sheikhupura, Pakistan district. A machine learning (Ada boost regressor) model has been used to quantify the effect of ambient parameters on cooling tower efficiency. After tuning the hyperparameters, an R-square score of 0.983 and a root mean squared error of 0.57 are achieved. Afterwards, a sensitivity analysis of relative humidity (%), turned out to be the most important feature, with a contribution of 12%. The novelty of this research lies in its mathematical model for power plant site selection, which optimizes cooling tower efficiency, reduces pollution, and promotes environmental sustainability.

KEYWORDS

power plants, cooling towers, machine learning, ambient parameters, site selecting

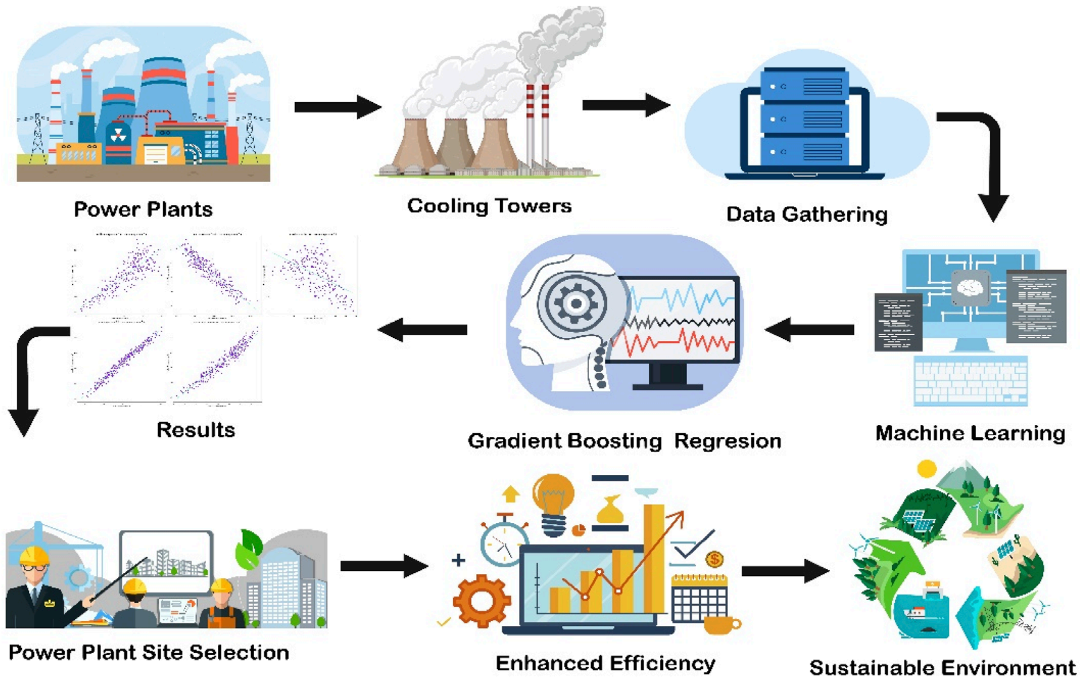
1 Introduction

As industrialization progresses worldwide, there is an increasing demand for energy, resulting in the issue of air pollution. This hazard increases risks to both people's health and the natural world. The International Energy Agency (IEA) predicts a 25% rise in energy demand by 2040, largely driven by the growth of emerging economies. The developing countries make a major contribution towards this increase in energy demand. The escalating use of energy coupled with the use of fuels is exacerbating air pollution. Emissions of pollutants, like Volatile organic compounds and nitrogen oxides (NOx) interact with atmospheric oxidants resulting in the creation of smog (Raza et al., 2021). The United Nations introduced Sustainable Development Goals (SDGs), which are intended to guide towards a sustainable world. This research emphasizes SDG 6 which is about "clean water and sanitation". There is a need for continuous water supply in cooling towers which is sourced from rivers, ground water or lakes and a major portion of the water is wasted due to evaporation, drift, and blowdown which is against all the efforts to conserve the water. This research contributes to reducing the wastage of water by improving the water efficiency of cooling towers in the power sector. In addition, SDG 7 seeks to promote "Clean Energy" in addition to SDG 13, which highlights the issue of "Climate Action" requiring a reduction in the rate of climate change and advocating for clean energy choices. The journey to progress and protection together has led to an increased demand for collaboration in diverse fields. Air pollution not only affects health but also damages the ecosystems, resulting in habitat loss and reduction of biodiversity. The financial losses due to environmental damage highlights the importance for transitioning to clean energy sources and adoption of sustainable methods (Arora and Mishra, 2019). Governments, companies, and other stakeholders are increasing attention towards renewable energy initiatives by introducing stricter environmental regulations to reduce emissions (Rajah et al., 2024).

Cooling towers, essential in many industrial processes, are considered critical components in energy consumption and environmental impact. Regarding power plants, manufacturing facilities, and commercial buildings, cooling towers are large structures which can play a prominent role in maintaining optimal temperatures of machines and in various industrial processes (Schulze et al., 2019). Cooling towers have evolved through novel and interesting technical development and now have become an integral part of infrastructure. Their extensive utilization has sparked concerns regarding their footprint in terms of energy and water usage (Suamir et al., 2018). The operation of cooling towers is based on circulation of water to increase the heat exchange. A cooling tower is a component which needs a source of energy input primarily in the form of electricity. It is approximated that cooling towers account for 20% of the worldwide electricity production (Kılkış and Krajačić et al., 2019). When hot water is released from condenser to back into the environment, it can disrupt the ecosystem by raising temperatures in rivers, lakes, and other reservoirs of water. This can affect the habitats of animals and contribute to thermal pollution. Due to these hazards, organizations are working to improve the performance of their cooling systems. Cooling mechanisms, such as water evaporation cooling, can improve energy efficiency while reducing environmental impacts. To further minimize harm, companies can adopt water conservation and recycling practices in cooling tower operations (Barbosa et al., 2019).

This research is based on the major research gap about the influence of ambient parameters on cooling tower efficiency. This research is useful for the best site selection for power plants considering optimal parameters for ambient conditions. The primary objective is to enhance cooling tower efficiency by leveraging conditions, thereby lowering energy consumption and environmental impact. One more novel aspect of this research is the use of advanced machine learning models including Gradient Boosting, Cat Boost, and AdaBoost. The main reason for the selection of these algorithms is their effectiveness in handling the data with non-linear relationships and analysis of feature importance in energy-related studies (Tyralisand Papacharalampous, 2021). These models have shown high predictive accuracy in previous studies for modelling operational parameters in the powerplants. They effectively map complex functions in high-dimensional input spaces and process large datasets while avoiding overfitting. AdaBoost is a very effective ensemble learning model that can combine and improve weak learners to form a strong predictive model. In the study of (Shah et al., 2024), AdaBoost has shown its ability to improve its predictive accuracy in complex environmental systems. Gradient Boosting is based on optimizing the loss function, which makes it suitable to capture complex data patterns. It is very flexible model which has shown its effectiveness in prior studies (Bassi et al., 2021). Cat Boost is particularly used to handle categorical variables, and it is effective for mitigation of overfitting in large datasets.

The "water-energy nexus" is the term referred to the interdependence of water resources and energy production, as thermal power plants require large amounts of water for cooling (Das et al., 2024). Water scarcity is turning out to be the greatest concern for power generation as global warming due to climate change is increasing. Cooling towers play a very significant role in dissipating heat for power plants. Dissipating a considerable amount of heat requires enough renewable water resources. Extreme climate conditions like extensive droughts and heat waves severely affect the ability of cooling towers to operate without any problem to water supply access by power plant facilities. Areas where water scarcity takes place can be a factor in limiting the operation of power plants, thus reducing their potential capacity for electricity generation. For that reason, many countries today use techniques in using less water while generating power. Using efficient towers in cooling would tremendously minimize water consumption, thereby increasing the power facility's resistance to water shortages. A cooling tower site selection should consider ambient humidity, temperature, and water resource availability for long-term sustainability. This research provides useful insight to all stakeholders and decision-makers using machine-learning models and thoroughly analyzing each contributing input factor to enhance the effectiveness of cooling towers and performance of combined cycle power plants. This study is, therefore, an advancement toward establishing the correlation between cooling tower outlet water temperature and ambient parameters. The knowledge of this study can help technical experts to select suitable areas for the installation of power plants that can contribute to sustainability by saving energy and reducing emissions. This research offers the key to unlocking the door of energy conservation by helping



GRAPHICAL ABSTRACT

the energy sector to find ways to improve the effectiveness of cooling towers. The goals of this study are in line with international initiatives for a cleaner and more eco-friendly world balanced with justice for all.

1.1 Literature review

There are different cooling towers that are classified based on the nature of their heat transfer mechanism, the type of airflow pattern, and their application (Dhorat et al., 2018). There are three major categories that include wet cooling towers, dry cooling towers, and hybrid cooling towers with different operational characteristics that determine efficiency. Wet cooling towers rely on evaporation of water to remove heat and thus are very efficient but sensitive to ambient wet bulb temperature (Ali et al., 2022). These have the following primary parameters: wet bulb temperature, relative humidity, airflow rate, water flow rate, and efficiency of the packing material. As wet bulb temperature is predominant in evaporative cooling, it ranks among the factors of utmost importance for performance prediction. Their analysis of sensitivity indicates how sensitive the ambient humidity and airflow rates are to cooling efficiency. Dry cooling towers are air-to-air heat exchangers, making them suitable for areas where water is scarce. Their key parameters are dry bulb temperature, ambient air velocity, and heat exchanger efficiency. Their Performance mainly depends upon the ambient air temperature and material of the heat exchanger. Sensitivity Analysis determines the amount to which the rate of airflow and efficiency of heat exchanger impact the drop in temperature. Hybrid Cooling Towers hybridize wet and dry cooling for maximum performance and minimum water consumption. Hybrid towers dynamically switch between wet and dry cooling modes and thus complex control mechanisms must be designed for modeling (Taimoor et al., 2022). Their sensitivity analysis assesses how thresholds for switching between wet and dry operation impact efficiency under varying ambient conditions.

The faults due to the complexity of air conditioner systems (ACS) are hard to detect since the system is very complicated. The paper (Sulaiman et al., 2020) focused on the influence of many faults in the coefficient of performance (COP) of the ACSs, and the machine learning algorithm was used to categorize these flaws. The machine learning models used to classify defects are multi-layer perceptron (MLP), deep learning, and support vector machine (SVM). This research concluded that 99.4% accuracy was observed by MLP as compared to deep learning and support vector machines. An accuracy of 97% was observed by SVM after MLP, which was the second-highest accuracy (Sulaiman et al., 2020). Cooling towers are affected by annual climate change, particularly by seasonal temperature variations. Mathematical models combined with mass and energy balance equations are critical for improving and forecasting a cooling tower's efficiency. The hot regions have higher values of fan slack; thus, using variable frequency drives (VFDs) is encouraged to decrease energy consumption (Pontes et al., 2019). A comparative analysis has been performed of this research with current research in which it has been observed that more than five machine learning regression models have been used, and among these models, the most suitable model was selected that achieved the highest R^2 score of 0.984 and mean square error (MSE) of 0.24. In the testing phase, it was observed that the predicted points were very close to actual points. As revealed by previous studies, the efficiency of cooling towers can be predicted using machine learning and mathematical models. Furthermore, it has also been concluded from the above research that using VFDs in hot climate areas reduces energy consumption.

The performance of cooling towers is affected by one of the most critical ambient parameters: wet bulb temperature. Several other factors also contribute to influence the efficiency of cooling towers, such as dry bulb temperature, ambient pressure, mass flow rates of water and air (Laković and Laković et al., 2012). These factors include both ambient parameters and internal parameters of the cooling tower. The present study identified a research gap, as previous studies considered only ambient parameters and analyzed their contribution towards efficient cooling towers. To address this, artificial intelligence tools, such as machine learning, have been used to conduct sensitivity analysis of ambient features. First, relations between the features have been identified to show the strong ties between these ambient features. Second, a machine learning model is used to perform the analysis, and the importance of each feature is identified. Third, with well-trained machine learning algorithms, the cooling performance is predicted under different ambient conditions, which can be very valuable for installing cooling towers and optimizing their efficiency. Finally, this machine learning model is used to analyze ambient parameters' sensitivity to quantify their effects on cooling tower performance. Overall, the paper fills the research gap by applying machine learning to provide a deeper understanding of how other ambient parameters and wet bulb temperatures affect cooling tower performance.

For real-time optimization of multicell cooling towers, applications of machine learning techniques represent a leap forward. Training an artificial neural network (ANN) using steady-state load-following data yielded an impressive R^2 score of 0.966. This method yielded an annual energy savings of 6.7% in a year-round simulation, especially during the colder months and periods of low electrical load (Blackburn et al., 2020). The use of machine learning algorithms plays an important role in the efficiency of a computerized data center. Machine learning algorithms were used to model the COP of data centers. This paper (Shoukourian et al., 2017) demonstrates the role of the COP model in making data centers more efficient. Predictions of COP were performed, and they were also validated with the real-time data of data centers (Shoukourian et al., 2017). In current research, machine learning models make cooling towers more energy efficient. This research also shows the critical ambient features from which the outlet water temperature of cooling is highly affected. This study presents the importance of machine learning algorithms in different industrial sectors to make them more energy efficient.

Wind power values have been predicted with the help of machine learning using the data of wind speed (Demolli et al., 2019). Predictive maintenance has been performed using machine learning. In this study, previous maintenance data of hydroelectric power plants is used for training and classification of data. High-level decision trees and SVM have been used as the most effective algorithms to predict failure possibilities, highlighting the feasibility and practicality of using predictive machine learning in practical industrial applications (Xayyasith et al., 2018). In parallel, another study uses predictive maintenance (PDM) to reduce the vibration intensity of cooling tower fans in a process plant. The results reveal a significant reduction in vibration levels, which implies improved fan reliability (Rupinder, 2009). These studies have demonstrated the versatility of machine learning and its applicability in many industries. It is used for forecasting and predictive maintenance, which ultimately leads to cost and time efficiency. At the same time, machine learning has some limitations, such as training data must have adequate quantity and quality for accurate estimation. In the current research, machine learning has also shown significant cost- and time-efficient advantages. Rather than practically observing the effect of ambient parameters on the outlet temperature of a cooling tower, which could be time and money-consuming, machine learning has been used to observe the effects of ambient parameters on the cooling tower's performance.

The heat transfer model for the cooling tower enables the real-time optimization of the water-cooling system, independent of operating condition variations, and identifies the main features of heat transport material processing directly linked to the packing's heat transfer performance. Subsequently, a hybrid programming particle swarm optimization (HP-PSO) algorithm was developed to handle discrete and continuous changes in that system (Ma et al., 2021). The study demonstrates the algorithm's effectiveness in finding optimal conditions beyond other methods. Interestingly, this study shows that higher wet-bulb temperatures require higher cooling tower outlet temperatures, up to 32°C sometimes. Furthermore, the optimal return water temperature difference is highly dependent upon ambient conditions, with a difference of 5°C that is more energy-efficient for higher temperature values, and for lower ones, a 7°C difference is preferred. In energy savings, the most compelling evidence lies: the HP-PSO system achieves a remarkable reduction of up to 15.3% as compared to traditional methods, showing its potential for sustainable energy production (Ma et al., 2021). It is evident from past literature that previous research focused on single ambient parameters, whereas the present study considers multiple ambient parameters. Furthermore, previous research emphasized energy efficiency only, but the current study emphasizes both energy efficiency and sustainability. Notably, the present study addresses these issues and contributes to the selection of power plant sites, setting it apart from its predecessors.

As a transition to digital twins, data-driven models play an important role in optimizing cooling towers' performance. Combined with the Cross Industry Standard Process for Data Mining (CRISP-DM) approach, this framework simplifies performance forecasting, system analysis, and intensive energy management (Pontes et al., 2019). The integration of Building Information Modeling (BIM) and the Internet of Things (IoT) into the framework was used for predictive maintenance in terms of mechanical, electrical, and plumbing components (MEP). The function to predict MEP component conditions has been accomplished by the use of machine learning algorithms such as support vector machines (SVM) and artificial neural networks (ANN) (Cheng J. C. et al., 2020). The application of artificial intelligence in both computing and production equipment is leveraged to predict the RUL which is the remaining useful life. A hybrid neural network, which is formed by the combination of a fully convolutional neural network (FCN), Long Short-Term Memory (LSTM), and Multilayer Perceptron (MLP), is used to estimate the remaining useful life (RUL) of supercomputer cooling systems (Lima et al., 2023). In this study, the effectiveness of air conditioning systems (ACSs) is increased through a machine learning-based Model Predictive Control (MLB-MPC) algorithm (Chen et al., 2023). Previous studies have shown the effectiveness of machine learning models like CRISP-DM methodology, BIM, and IoT for predictive maintenance. By taking significant ambient parameters into consideration, current study highlights a practical and effective approach to enhance reliability of cooling tower. Furthermore, this study emphasizes that obtaining nearly 99% accuracy represents a remarkable advancement in the domain, demonstrating the potential of machine learning to enhance cooling tower performance.

The current research is focused on improving cooling towers' performance using machine learning, which significantly contributes to the field by addressing several gaps identified in existing literature. First, in the literature, cooling parameters like mass flow rates of air and water have been used to calculate the effectiveness of cooling towers, but the effect of only ambient parameters on a cooling tower's performance has not yet been considered. This major research gap has been covered in this study. Second, ambient parameters can help with the site selection of a power plant, but they must be considered before installing a power plant. The best site selection will help improve a cooling tower's efficiency. Comparing ambient parameters among the sites will help in choosing the best site. Third, this study also plays an important role in improving sustainability. Sustainable Development Goals (SDG) 6, 7, and 13 have been targeted in this research, which will lead towards a sustainable environment.

2 Methodology

2.1 Data collection

The research process map is presented in Figure 1. The selection of variables used in the input section was made systematically concerning the target variables. All data regarding the input features and the target variable of the efficiency of cooling tower is obtained from a combined cycle power plant situated at Sheikhupura, Pakistan. Several quantitative study reports are used in the analysis to ensure the validity of the data collected. The dataset was then examined for missing values, and any records with unavailable data are excluded.

Table 1 presents a summary of the utilized data. Data sourced from the 1140 MW combined cycle power plant in Sheikhupura, Pakistan. A total of 264 data points were left after removing any missing values. To simulate the cooling tower and predict the efficiency of cooling tower, 6 features are considered which can be classified in two groups: (i) ambient features (Ambient Temperature, Ambient. Pressure, Relative. Humidity and (ii) working parameters (Inlet Temperature, wet. bulb Temperature, outlet. Temperature These factors are selected based on the effect on the cooling tower's efficiency.

2.2 Data visualization and pre-processing

Data visualization is very important in the development and training of ML models, as it offers insight into data density and distribution in terms of its inputs and outputs features. An ideal feature set ensures comprehensive data representation across the full range of variable values within the system's operational scope, ensuring that the model possesses complete knowledge about the system which is under investigation. For this purpose, box plots are generated from the data to obtain necessary graphical representation of the variables while the heat maps are generated to assess the correlations in the input variables.

It is important to note that there are many methods in literature to characterize linear interdependence of two variables and one such method is the Pearson correlation coefficient (PCC) (Shoukourian et al., 2017). Computing the Pearson Correlation Coefficient (PCC) for selected variables in an input-output model is essential to understand their linear relationships. This is a critical step in conducting machine learning tasks. Also, the possibility of analyzing nonlinear or interactive effects of the variables can be analyzed using ML models. This research calculated the Pearson Correlation Coefficient (PCC) between the input and target variables using the cooling tower data set to evaluate their linear relationship. The PCC is mathematically expressed as shown in Equation 1:

$$R_{xy} = \frac{\sum_{i}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i}^{N}(y_i - \bar{y})^2}} \tag{1}$$

where R_{xy} denotes the Pearson Correlation Coefficient (PCC) describing the relationship between the independent variable, X, and the dependent variable, Y. This therefore means that PCC takes the value from −1 up to 1. In other words, if $R_{xy} = 1$ in this model, it means that the variables underlying the model are perfectly linearly related to each other, and it also If $R_{xy} > 0$, then the relationship is positive If $R_{xy} < 0$, then, the relationship is negative. However, it is possible to have the case of $R_{xy} = 0$ which ascertains that there is no correlation among all variables.

2.3 Training and development of machine learning models

To estimate the performance of a cooling tower under various simulated conditions and environmental factors, three tree-based predictive modeling approaches are utilized: AdaBoost, Gradient Boost, and Cat Boost. These models are widely recognized in machine learning for their ability to interpret complex nonlinear relationships and interactions between numerous input variables. They adeptly map complex functions within high-dimensional input spaces and handle sizable datasets without falling prey to overfitting, unlike the challenges sometimes seen with multilayer perceptron. Notably, these methods have demonstrated strong results when applied to datasets with 200–1,000 data points and dimensions ranging from five to fifteen in the input space (Demolli et al., 2019). Thus, these algorithmic features benefit in building the process model for the cooling tower efficiency with accommodation of the dataset containing robust conditions of the input variables.

AdaBoost functions as an adaptive boosting technique utilized with decision trees for classification and regression tasks. This method tends to prioritize those trees that exhibit

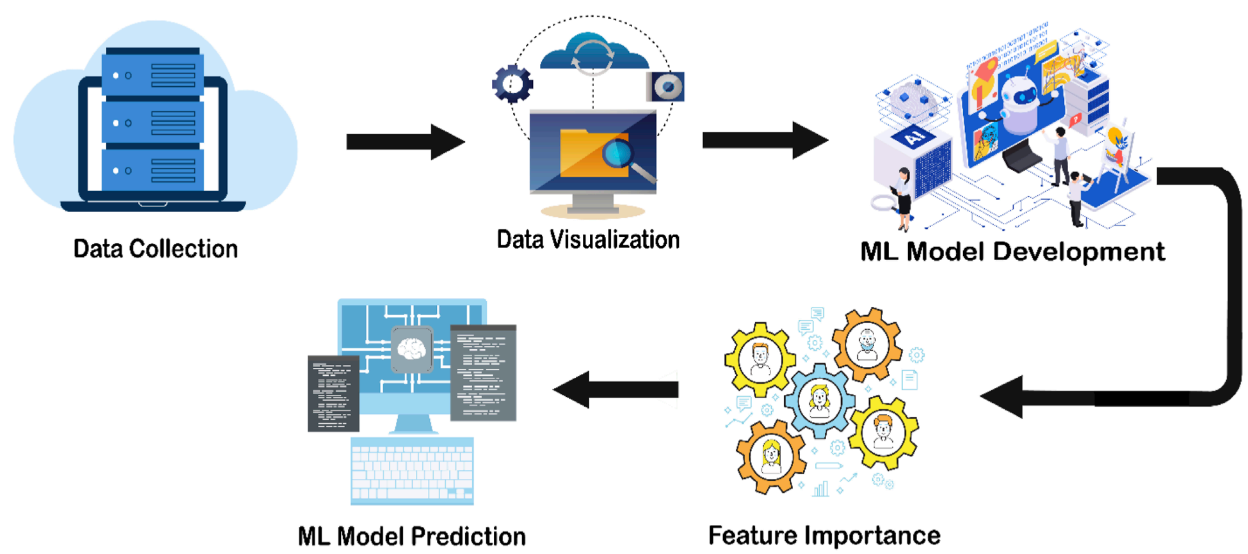


FIGURE 1
Machine learning project life cycle for modelling.

TABLE 1 Description of input features along with ranges.

	Input feature	Unit	Data range	No of samples
1	Ambient Temperature	°C	10.9–43.3	264
2	Ambient Pressure	kPa	973.3–1,026.9	264
3	Relative Humidity	%	18.5–80.1	264
4	Inlet Temperature	°C	30.4–41.5	264
5	Wet Bulb Temperature	°C	10.2–31	264
6	Outlet Temperature	°C	21.8–34.6	264

greater predictive errors, utilize them for modeling (referred to as decision stumps), and iteratively adjusts these errors throughout the modeling phase to optimize prediction accuracy for the given dataset.

For accurate and reliable forecasts with a high capacity to generalize across various scenarios using machine learning algorithms, it is important to carefully select several factors. Each algorithm data flow has a distinct parameter space, and the literature offers a variety of methods to ascertain the optimal values for hyper-parameters. Different techniques such as greedy optimization, grid search, random search, and particularly Bayesian optimization play roles in tuning hyperparameters (Xayyasith et al., 2018). Of these, grid search is well-regarded for its systematic exploration of parameter spaces to pinpoint the optimal combination of hyperparameters yielding superior performance in each ML model. In this study, by considering the importance of hyperparameter tuning, the grid search technique has been utilized. In ML, overfitting is a big problem when the models are not well trained in the approximation of the function space. To address this error, we use k-fold cross-validation which helps reduce overfitting. By evaluating the results of the trained machine learning model against models trained with k-fold cross-validation, we can identify any overfitting issues.

2.4 Error metrics

Performance criteria are established to evaluate the efficacy of the suggested machine-learning algorithm. It is essential to include both the coefficient of determination (R^2) and mean-square-error (MSE) in the list of performance metrics. Both performance metrics are shown in Equations 2, 3.

$$R^2 = 1 - \frac{\sum_{i}^{N} (y_i - \hat{y}_i)^2}{\sum_{i}^{N} (y_i - \bar{y}_i)^2} \tag{2}$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \tag{3}$$

where y_i represents the actual target variable values and $\hat{y}_i$ represents the modeled target variable values apart from y_i = actual target variable values, y_i = average of target variable values and i Total number of observation = N, constituted by i = 1, 2, 3, 4, …. R^2 is a metric for evaluating the accuracy of a machine learning model's predictions, with zero indicating a low capability to match input features with target variables, and one signifying perfect correlation between inputs and targets. Meanwhile, Mean Squared Error (MSE) measures the variance between the actual data points in a dataset and those predicted by a model.

2.5 SHAP evaluation of the developed machine learning model

The next and implicitly consecutive step in building a high-performing ML model is to determine the importance of the input variables relative to the target variable. Hence, different approaches have been described in the previous section in undertaking feature importance analysis (Ma et al., 2021). When calculating SHAP (SHapley Additive Explanations) values for input variables, a model-agnostic methodology is applied that constructs cooperative games with the input variables to evaluate their impact on the output variable (Cheng J. C. et al., 2020). In SHAP methods analysis, sensitivity can be evaluated for individual data points or the whole dataset. This yields SHAP values for input variables, ranking their importance. Understanding key variables' impact on the outcome is crucial for designing lab experiments and enhancing industrial procedures.

3 Results and discussion

3.1 Data description

After formatting the data for the target variables, box plots are generated to show their distribution. Box plots are another method that effectively depicts the variation within a data set (Lima et al., 2023). The second illustration presents how the input variables are distributed within the input space concerning the target variables, Out, as depicted in Figure 2. Most of the variables are concentrated in the 25 percent – 75 percent range, or inter-quartile range (IQR). Some variables have more than one observation, that is, there are data points whose values are greater than 1 for some of the variables. Outliers may be calculated as 1.5*IQR times the distance from the higher quartiles or the lower quartiles. The distribution profiles derived from literature demonstrate a broad operational range for both input and target variables typically used in the design and function of cooling tower simulations. Owing to the great operating range of the various variables, techniques of ML can be implemented to predict efficiency of the fluid when certain inputs are applied.

PCC is employed to establish the extent of the linear relationship between the input and target features. Figure 3 displays heat maps generated from the PCC analysis, illustrating in its columns the correlation between the input variables and the output. Temp (represented in blue and green as shown in the figure). The scores that are less than PCC reveal the lack of a direct relationship between two factors with a supposition of the presence of a nonlinear and intricate relationship between two variables. Consequently, the current techniques for developing and training ML models can recognize complex relationships and patterns within a dataset to precisely construct a functional representation of the connection between the input and target variables in the system under investigation.

The Pearson correlation coefficient (PCC) map is created for the input variables, as well as between the input variables and the target variable. Some variables show high PCC values. However, most variables display low PCC values, indicating a nonlinear relationship among the variables within the collected dataset.

3.2 Model performance

Three tree-based machine learning algorithms—Cat Boost, Gradient Boosting, and AdaBoost—are developed to estimate the outdoor temperature from the dataset identified during data collection, visualization, and processing stages of the research. The division of data into training and testing sets followed a consistent split-ratio of 80/20. Overall, the fine-tuning of hyperparameters across these machine learning models is key to achieving reliable predictive performance. For the Cat Boost model, tuned hyperparameters include learning rate, loss function, depth, number of iterations, L2 regularization, and maximum bin number; for the Gradient Boosting model, adjustments are made to learning rate, maximum depth, sub-sample, col sample by tree, and number of estimators; in the AdaBoost models, learning rate, loss function, and number of estimators are the hyperparameters that are tweaked.

In summary, the side-by-side scatter plots for actual versus predicted responses from the training and testing data sets corresponding to Cat Boost, AdaBoost, and Gradient Boost algorithms are illustrated in Figure 4. Statistical analysis revealed that AdaBoost exhibited superior performance compared to Cat Boost and Gradient Boost. Specifically, for the training set, the R-squared values, which reflect the model's accuracy, were reported as 0.991, 0.997, and 0.999 for AdaBoost, Cat Boost, and Gradient Boost, respectively. For the testing set, these values are 0.991, 0.988, and 0.880 from three used algorithms, AdaBoost emerged as the most effective, also demonstrating the smallest root mean square error (RMSE) with a value of 0.11.

The AdaBoost algorithm was notably effective in both training and testing stages, showing superior R^2 values of 0.991, as compared to Gradient Boost and LG Boost.

Given the strong performance of the models during training and on test data, there could have been an issue with overfitting. Overfitting occurs when models fit too closely to a specific dataset and pick up noise, along with random fluctuations that negatively impact their predictive ability on new data. This reduces the models' capacity to generalize from the underlying process. The current study used k-fold cross-validation (CV) to assess potential overfitting in the machine learning (ML) models. With k-fold CV, one divides the dataset into k parts (k = 5 was chosen for this study), then checks the model's effectiveness on 1 part after training on the remaining k-1 parts. By averaging results across the k iterations, more generalized training is achieved, and it addresses bias-variance concerns. As presented in Table 2, the overfitting issue for the AdaBoost, Gradient Boost, and Cat Boost models appears well-managed based on the closeness of the R^2 values from the k-fold method the R^2 values from testing data, as shown in Figure 4. This illustrates that the AdaBoost models demonstrate strong predictive capabilities and robust performance outside of the sample set. The AdaBoost loose curve performance is shown in Figure 5.

3.3 The impact of identified input factors on output temperature

The machine learning model developed from the provided dataset acts as a practical surrogate of the system being analyzed. The choice hinges on selecting a thoroughly trained model with strong

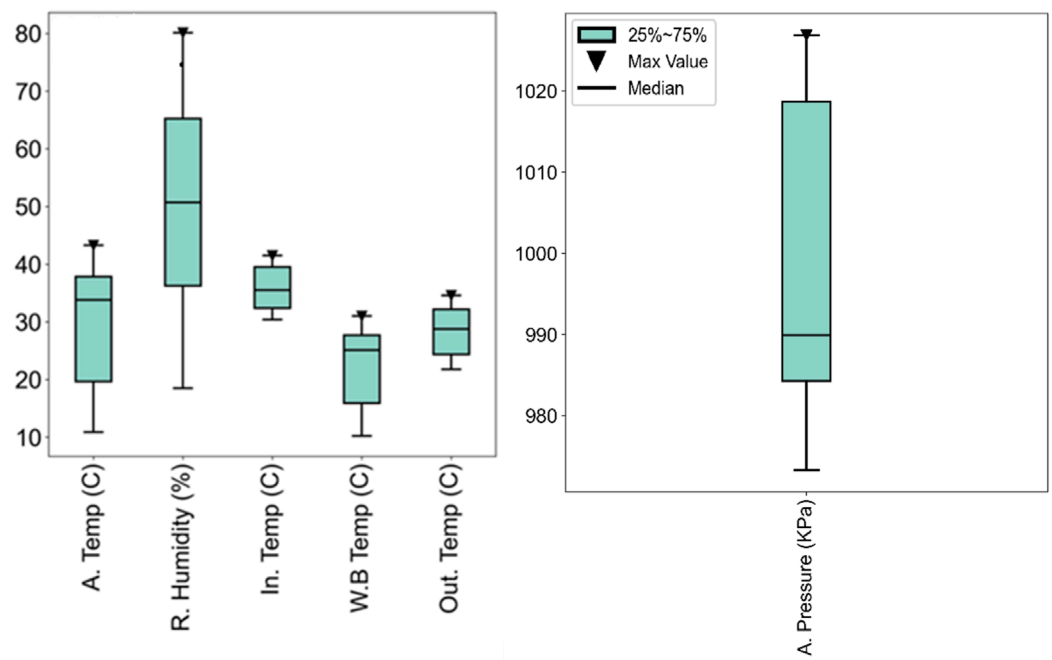


FIGURE 2
The data visualization employs box plots for the input variables, revealing a well-distributed spread across their operating ranges.

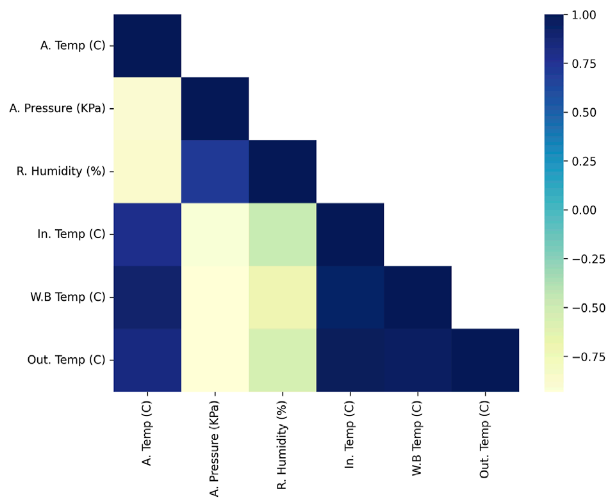


FIGURE 3
Pearson correlation coefficient map.

predictive power to uncover the system's inherent mechanics and to determine which input variables are more important. To attain this, a feature-important analysis based on the SHAP analysis is conducted. The results obtained from the AdaBoost model indicate that the model made precise predictions for the output Temperature when considering the input variables within the framework of the SHAP analytical method. This analysis highlighted the key input features of the process under investigation. Figure 6 illustrates the ranked importance of input variables as determined by SHAP. Remarkably, these findings constitute the existing knowledge about the mechanism of the cooling tower.

SHAP analysis reveals that Relative Humidity is the most vital factor in the cooling tower's performance amid changes in ambient temperature. The waterfall SHAP plot in Figure 7 shows how individual characteristics of a model affect the predictions for a given instance. Starting with a base value, it gradually adds or removes SHAP values to illustrate the impact of each attribute on the final prediction, highlighting how they increase or decrease the predicted value and the extent of their effect.

4 Research contribution

This research is focused on addressing key gaps by examining how well ambient parameters can improve cooling tower efficiency. This study investigates the impact these parameters have on a cooling tower's outlet temperature, as illustrated in Figure 6. By analyzing the relationship between dependent and independent variables, this study highlights the importance of considering ambient factors when designing power plants to maximize efficiency. Additionally, this research contributes to optimal site selection for cooling towers and power plants by comparing ambient conditions across potential locations. Figure 4 shows that the cooling outlet temperature drops as the wet bulb temperature decreases and *vice versa*. Thus, selecting a cooling tower site often involves choosing locations with lower wet bulb temperatures. Similarly, understanding how the outlet temperature correlates with other ambient factors like pressure and relative humidity can help in best site selection. This research underscores sustainability as a key contribution, aiming at Sustainable Development Goals 6, 7 and 13 to enhance human

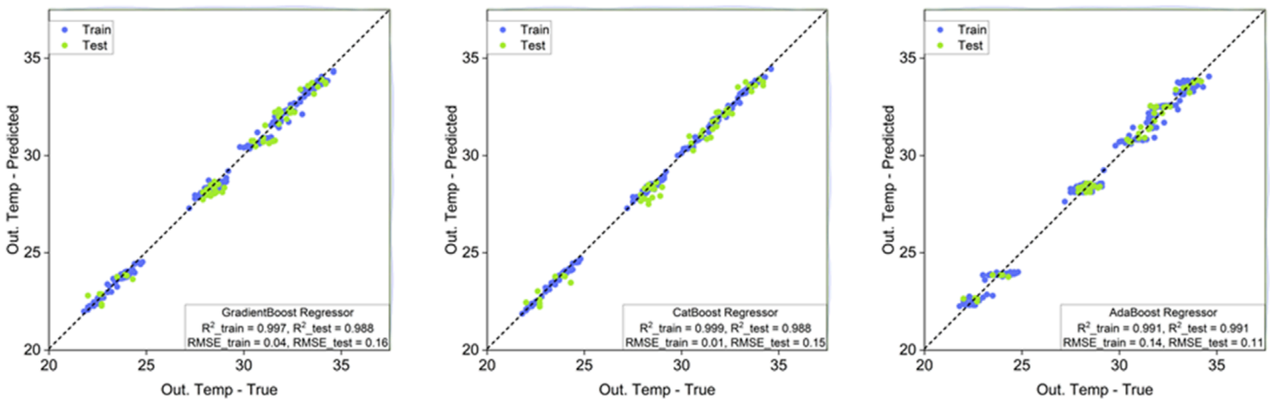


FIGURE 4
Comparative scatter plots showing the actual vs predicted values using Gradient Boost, Cat Boost, and AdaBoost models.

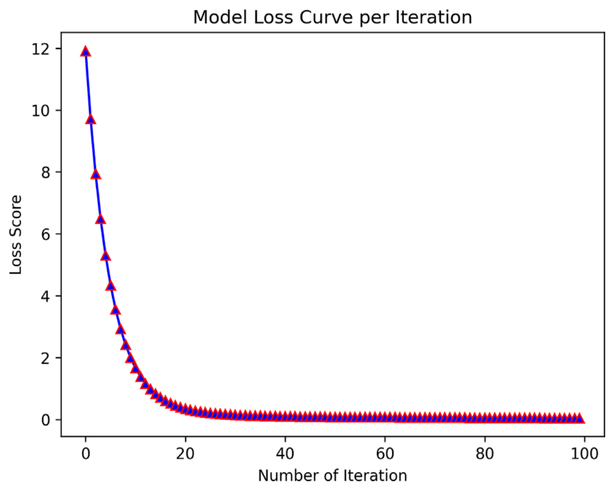


FIGURE 5
Adaboost loss curve performance. Curve indicates no overfitting.

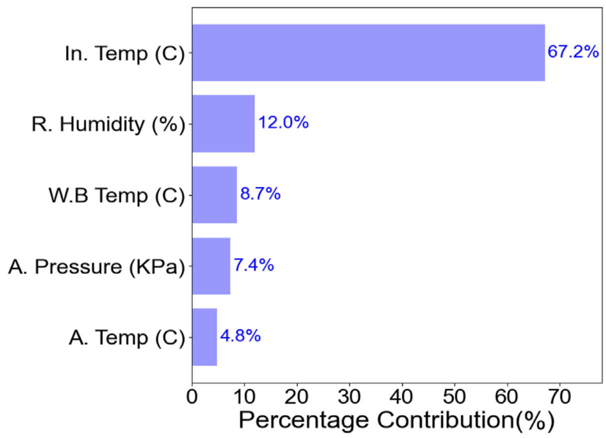


FIGURE 6
SHAP analysis.

TABLE 2 Machine learning models' effectiveness with CV.

Parameters	AdaBoost	CatBoost	GradientBoost
Training R^2	0.991	0.999	0.999
Testing R^2	0.991	0.988	0.988
CV-R^2	0.983	0.983	0.983
MSE	0.11	0.15	0.16

and environmental health. Increasing the efficiency of cooling towers improves power plant performance, reduces emissions, and promotes environmental sustainability. This work not only fills knowledge gaps regarding ambient factors but also offers actionable guidance for cooling tower and power plant site planning, support achieving the Sustainable Development Goals.

Furthermore, this research provides a quantitative assessment of different input parameters to impact the performance of cooling towers, providing optimization strategies based on data-driven solutions. By combining machine learning algorithms, this study provides proactive techniques in the operations of cooling towers based on real-time data for environmental conditions. Moreover, the findings suggest that smarter AI-driven cooling systems can be developed which can be self-regulated according to fluctuating environmental conditions. The research is useful for industry professionals and policymakers to formulate guidelines and policies to develop infrastructure of energy-efficient systems. Also, by incorporating the metrics for sustainability, this research presents not only the approach to improve energy efficiency but also environmental sustainability to ensure the long terms benefits for industries as well as climate resilience. Another contribution is the use of SHAP based analysis of feature importance which helps to understand the influence of parameters. This important insight can be used in other thermal management systems in industrial units to optimize their performance. This research adds valuable

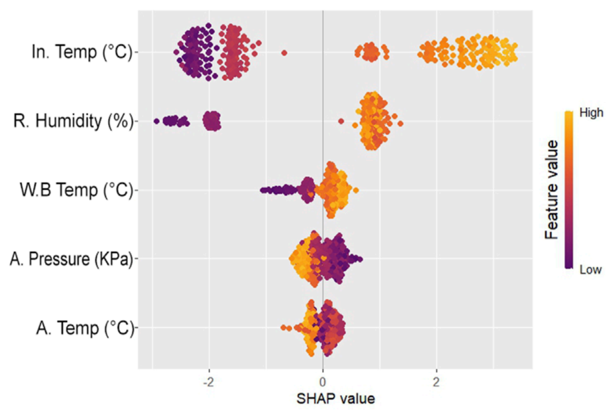


FIGURE 7
Waterfall SHAP plot.

knowledge by highlighting the economic, technical and policy related implications.

5 Conclusion

- The study presents a method to assess how environmental factors influence the efficacy of cooling towers (CT) for optimal site selection. An AdaBoost regression model, chosen for its superior R^2 score and minimal MSE compared to other ML models, was employed for analysis. The method's robustness was confirmed using data from an 1140 MW combined cycle power plant in Sheikhupura, Pakistan. The findings outline methods for enhancing CT efficiency by considering ambient conditions prior to the installation of power plants. The research also contributes to Sustainable Development Goals such as SDG 6 which is about clean water and sanitation for all, SDG 7, which is about ensuring access to "Affordable and Clean Energy," and SDG 13, which addresses "Climate Action," thereby supporting a sustainable environment.
- The accuracy of the AdaBoost regression is tested using Mean Square Error (MSE) and R^2 score, with the AdaBoost model achieving an impressive R^2 score of 0.984 on the testing dataset and an MSE of 0.11. Heatmaps illustrate various correlations among variables with minimal discrepancies between original and noise-added data. The SHAP value graph demonstrates how each feature affects the dependent variable. The actual versus predicted outlet temperatures of a CT are compared, indicating that the predicted values closely match the test dataset.
- The efficiency of power plants hinges on their CTs performance. Higher efficiency in CTs translates to higher overall plant efficiency. This is largely influenced by the CT's outlet temperature, which according to machine learning (ML) analyses, directly correlates with the CT's inlet temperature, the ambience temperature, and the wet bulb temperature, but inversely with relative humidity and ambient pressure. The data suggests that considering these ambient factors when installing CTs can boost their efficiency, resulting in reduced energy use and emissions, thereby promoting sustainability.
- The AdaBoost model might not entirely sustain its performance on power plants if it is deployed in climatic regions with quite different climatic conditions (such as arid, humid, or polar). This inconsistency results from various aspects, which are data dependency, feature sensitivity and adaptability. The training data can greatly influence the performance of AdaBoost. So, if trained on temperate climate data, it may fail to generalize well to any other climate that has very different ambient conditions unless fine-tuned with new data or retrained. Climatic variables are also nonlinear in interaction and may dramatically differ across regions. Hence, transfer learning techniques or including diversified climate data during training may be helpful for consistency in model performance.
- The study is confined to pre-installation factors for power plants since ambient conditions cannot be adjusted post-installation. The aim is to identify optimal sites for cooling towers (CTs) to ensure efficiency. Additionally, forecasting the outlet temperature by considering external ambient factors and internal factors such as water and air flow rates, alongside variable frequency drive speeds, can aid in enhancing cooling efficiency.
- Internal factors including water quality, variability in airflow, and practices regarding maintenance do impact the cooling tower's efficiency significantly. Thus, it might enhance the prediction ability of the model if such internal parameters are considered. Future studies can incorporate these internal factors by expanding the dataset to include operational and maintenance-related variables

It is also important to analyze the combined effect of these external and internal parameters using high-end machine learning techniques. More importantly, multivariate sensitivity analysis will be important to understand more about the nature of interrelations between these factors, which then leads to making more robust, generalizable, and reliable prediction models for estimating CT efficiency.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

MM: Conceptualization, Formal Analysis, Supervision, Writing–review and editing, Methodology. MAM: Software, Writing–review and editing, Supervision, Validation. MA: Data curation, Writing–review and editing, Resources. BM: Data curation, Resources, Writing–review and editing. TA: Investigation, Software, Writing–original draft. ZK: Visualization, Writing–original draft. SK: Software, Writing–original draft. AB: Software, Writing–original draft. SJ: Software, Writing–original draft. MK: Writing–review and editing. FH: Funding acquisition, Writing–review and editing. CB: Funding acquisition, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

Author MA was employed by QATPL, 1180MW CCPP Bhikki. Author BM was employed by Harbin Electric Company Limited, 1180MW CCPP Bhikki, O&M.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Ali, A. K., Mohammed, A. Q., and Mahdi, Q. S. (2022). Experimental study of a natural draft hybrid (wet/dry) cooling tower with a splash fill type. *AIMS Energy* 10 (4), 648–664. doi:10.3934/energy.2022031

Arora, N. K., and Mishra, I. (2019). United Nations sustainable development goals 2030 and environmental sustainability: race against time. *Environ. Sustain.* 2 (4), 339–342. doi:10.1007/s42398-019-00092-y

Barbosa, L. F. F., Nascimento, A., Mathias, M. H., and de Carvalho Jr, J. A. (2019). Machine learning methods applied to drilling rate of penetration prediction and optimization-A review. *J. Petroleum Sci. Eng.* 183, 106332. doi:10.1016/j.petrol.2019.106332

Bassi, A., Shenoy, A., Sharma, A., Sigurdson, H., Glossop, C., and Chan, J. H. (2021). "Building energy consumption forecasting: a comparison of gradient boosting models," in *Proceedings of the 12th International Conference on Advances in Information Technology* (Bangkok, Thailand: Association for Computing Machinery). Article 27. doi:10.1145/3468784.3470656

Blackburn, L. D., Tuttle, J. F., and Powell, K. M. (2020). Real-time optimization of multi-cell industrial evaporative cooling towers using machine learning and particle swarm optimization. *J. Clean. Prod.* 271, 122175. doi:10.1016/j.jclepro.2020.122175

Chen, S., Ding, P., Zhou, G., Zhou, X., Li, J., Wang, L. L., et al. (2023). A novel machine learning-based model predictive control framework for improving the energy efficiency of air-conditioning systems. *Energy Build.* 294, 113258. doi:10.1016/j.enbuild.2023.113258

Cheng, J. C., Chen, W., Chen, K., and Wang, Q. (2020). Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. *Automation Constr.* 112, 103087. doi:10.1016/j.autcon.2020.103087

Das, S. S., Majumdar, R., Krishnan, A. V., and Srikanth, R. (2024). Exploring urban water-energy nexus: a case study of thermal power plants in Raichur and Ballari districts in Karnataka. *Nat. Resour. Forum*. doi:10.1111/1477-8947.12507

Demolli, H., Dokuz, A. S., Ecemis, A., and Gokcek, M. (2019). Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Convers. Manag.* 198, 111823. doi:10.1016/j.enconman.2019.111823

Dhorat, A., Al-Obaidi, M. A., and Mujtaba, I. M. (2018). Numerical modelling and sensitivity analysis of natural draft cooling towers. *Chem. Prod. Process Model.* 13 (4). doi:10.1515/cppm-2017-0078

Kılkış, Ş., Krajačić, G., Duić, N., Montorsi, L., Wang, Q., and Rosen, M. A. (2019). *Research frontiers in sustainable development of energy, water and environment systems in a time of climate crisis*. Elsevier.111938

Laković, M. S., Laković, S. V., and Banjac, M. (2012). Analysis of the evaporative towers cooling system of a coal-fired power plant. *Therm. Sci.* 16, S375–S385. doi:10.2298/tsci120426176l

Lima, A. L. d. C. D., Aranha, V. M., and Nascimento, E. G. S. (2023). Predictive maintenance applied to mission critical supercomputing environments: remaining useful life estimation of a hydraulic cooling system using deep learning. *J. Supercomput.* 79 (4), 4660–4684. doi:10.1007/s11227-022-04833-5

Ma, K., Liu, M., and Zhang, J. (2021). Online optimization method of cooling water system based on the heat transfer model for cooling tower. *Energy* 231, 120896. doi:10.1016/j.energy.2021.120896

Pontes, R. F., Yamauchi, W. M., and Silva, E. K. (2019). Analysis of the effect of seasonal climate changes on cooling tower efficiency, and strategies for reducing cooling tower power consumption. *Appl. Therm. Eng.* 161, 114148. doi:10.1016/j.applthermaleng.2019.114148

Rajah, K., Ghriss, O., Fguiri, A., and Jeday, M.-R. (2024). "Validating the techno-economic and environmental sustainability of solar PV technology in agriculture using RETScreen experts to assess its viability," in *Proceedings of the 2nd international conference on green energy conversion system*. Editors A. Mellit, L. Sbita, K. Kemih, and M. Ghanes (Singapore), 795–800.

Raza, W., Saeed, S., Saulat, H., Gul, H., Sarfraz, M., Sonne, C., et al. (2021). A review on the deteriorating situation of smog and its preventive measures in Pakistan. *J. Clean. Prod.* 279, 123676. doi:10.1016/j.jclepro.2020.123676

Rupinder, S. (2009). Application of predictive maintenance program in cooling tower fans: a case study. *IUP J. Mech. Eng.* 2 (3), 18–28.

Schulze, C., Thiede, S., Thiede, B., Kurle, D., Blume, S., and Herrmann, C. (2019). Cooling tower management in manufacturing companies: a cyber-physical system approach. *J. Clean. Prod.* 211, 428–441. doi:10.1016/j.jclepro.2018.11.184

Shah, F. U., Khan, A. U., Khan, A. W., Ullah, B., Khan, M. R., and Javed, I. (2024). Comparative analysis of ensemble learning algorithms in water quality prediction. *J. Hydroinformatics* 26 (12), 3041–3059. doi:10.2166/hydro.2024.071

Shoukourian, H., Wilde, T., Labrenz, D., and Bode, A. (2017). "Using machine learning for data center cooling infrastructure efficiency prediction," in *2017 IEEE international parallel and distributed processing symposium workshops (IPDPSW)*, 954–963.

Suamir, I. N., Ardita, I. N., and Rasta, I. M. (2018). "Effects of cooling tower performance to water cooled chiller energy use: a case study toward energy conservation of office building," in *2018 international conference on applied science and technology (iCAST)*, 712–717.

Sulaiman, N. A., Abdullah, M. P., Abdullah, H., Zainudin, M. N. S., and Yusop, A. M. (2020). Fault detection for air conditioning system using machine learning. *IAES Int. J. Artif. Intell.* 9 (1), 109. doi:10.11591/ijai.v9.i1.pp109-116

Taimoor, A. A., Saeed, U., Rather, S.-u., Al-Shahrani, S., Bamufleh, H. S., Alhumade, H., et al. (2022). Economic and technical analysis of a hybrid dry cooling cycle to replace conventional wet cooling towers for high process cooling loads. *Energies* 15 (21), 7986. doi:10.3390/en15217986

Tyralis, H., and Papacharalampous, G. (2021). Boosting algorithms in energy research: a systematic review. *Neural Comput. Appl.* 33 (21), 14101–14117. doi:10.1007/s00521-021-05995-8

Xayyasith, S., Promwungkwa, A., and Ngamsanroaj, K. (2018). "Application of machine learning for predictive maintenance cooling system in Nam Ngum-1 hydropower plant," in *2018 16th international conference on ICT and knowledge engineering (ICT&KE)*, 1–5.