Wang Z, Zeng C, Zhu Z, Li Y, Ma X, Zhao X. Time-series machine learning for predictive optimisation of a highly efficient evaporative cooling system. Building Services Engineering Research and Technology. Copyright © 2025 The authors. DOI: https://doi.org/10.1177/01436244251315047.

Time-Series Machine Learning for Predictive Optimisation of a Highly Efficient Evaporative Cooling System

ZHICHU WANG[1], CHENG ZENG[2], ZISHANG ZHU[3], YUNHAI LI[1], XIAOLI MA[1], XUDONG ZHAO[1]
[1] School of Engineering, University of Hull
[2] Energy and Environment Institute, University of Hull
[3] Department of Civil Engineering, Surveying and Construction Management, Kingston University London, London, UK
* Corresponding author: C.Zeng@hull.ac.uk

Abstract

As data centres become integral to modern infrastructure, their energy consumption, particularly in cooling systems, presents a critical challenge for sustainability. This paper addresses this issue by applying time-series machine learning models to forecast the performance of a highly efficient 100 kW evaporative cooling system applied in a real-world data centre. Using a dataset spanning four months, we developed and optimised two predictive models based on XGBoost and Random Forest, to estimate cooling capacity and Coefficient of Performance (COP). Initial results showed suboptimal performance, with the XGBoost model achieving a Mean Absolute Error (MAE) of 1.34 for cooling capacity and 6.50 for COP, alongside a negative R-squared, indicating poor fit. However, after hyperparameter tuning, the Random Forest model significantly improved the predictions, achieving an MAE of 0.39 and an R-squared of 0.85 for cooling capacity, and an MAE of 2.21 and an R-squared of 0.54 for COP. These findings underscore the potential of these models to optimise cooling efficiency, offering valuable insights for reducing energy consumption and operational costs in data centre operations. This research paves the way for more sustainable data centre designs and operations across diverse climatic conditions.

Practical Application

The predictive models developed in this study enable building environment professionals to optimise data centre cooling systems. By accurately forecasting cooling capacity and Coefficient of Performance (COP) under varying environmental conditions, these models allow for proactive adjustments to cooling strategies, ensuring efficient operation and minimising energy waste. This research provides a practical tool for enhancing the sustainability of data centres, directly supporting industry efforts to meet stringent energy efficiency targets and reduce the carbon footprint of critical infrastructure.

Keywords: Data centre, evaporative cooling, machine learning, time-series forecasting

1.0 Introduction

1.1 The Challenge of Energy-Efficient Cooling Systems in Data Centres

Data centres have become integral infrastructures for worldwide organisations, serving as dedicated spaces to host computing applications and store vast amount of data.[1] Over the past decade, data centres have further evolved into critical components, facilitating a myriad of daily activities, such as remote data services, cloud computing, and social media interactions. The compact design of data centres allows for efficient management of large-scale data processing and storage, making them pivotal for a substantial user base. However, the rapid increase in data centre usage has raised significant concerns about their power consumption, highlighting the pressing need for improved energy efficiency and environmental sustainability.

In 2022, data centres accounted for approximately 1% of global electricity consumption, with this figure reaching as high as 2.5% in the United Kingdom.[2] While the IT systems contribute substantially to a data centre's total power consumption, cooling systems, depending on the specific cooling techniques utilised, can account for up to 40% of the total energy demand.[3] A highly efficient evaporative cooling system, demonstrate a Coefficient of Performance (COP) of up to 52.5, which significantly surpasses the typical COP range of 2 to 4 in conventional vapor compression cooling systems.[4] Integrating such advanced evaporative cooling systems within data centres offers considerable potential for energy savings and carbon emissions reduction.

Different data centres may exhibit unique design characteristics, albeit adhering to common principles centred around regulating temperature and humidity levels to ensure optimal server functionality, striking a delicate balance between safety and performance. Consequently, the necessity of preliminary modelling during the design phase is universally recognised, underscoring the importance of predicting and optimising environmental conditions tailored to the specific requirements of each data centre.[5] However, physical modelling without device-specific calibration often fails to accurately capture the dynamic nature of data centre cooling systems.[6] In contrast, machine learning techniques have shown greater efficacy by directly interacting with systems or by learning from extensive datasets collected from operational systems to optimise and forecast system performance.[7,8,9] Integrating machine learning with data centre cooling systems, especially using efficient evaporative cooling technologies, thus emerges as a promising research direction to achieve energy-efficient solutions.

1.2 Black-box Model

As one of the fundamental concepts in machine learning, black-box model has been widely applied to engineering,[10] data science,[11] and financial sector,[112] focusing on the characterisation of systems through their inputs and outputs without delving into the intricacies of internal mechanisms.[13] A schematic diagram of the working theory is shown in Figure 1. The primary advantage of black-box models lies in their ability to handle complex systems whose underlying mechanics are either too intricate for explicit modelling or irrelevant to the target objectives.[14] In this study, the authors applied this concept to build predictive models for the performance of data centre cooling system by only introducing air condition datasets excluding physical system measurement, e.g., the size of the heat-and-mass exchanger, to enhance the practicality of the models.

[insert Figure 1.]

Figure 1 – Black-box model

2.0 Time Series Machine Learning Techniques

2.1 Introduction to Time-Series Forecasting

Time-series data represents a sequence of observations or measurements collected over time, providing critical insights into temporal dynamics. In the context of data centre operations, time-series data often encompasses historical records of environmental variables, such as temperature, humidity, and energy consumption, each with an associated timestamp, for instance, in dd/mm/yyyy format. Unlike cross-sectional data, which provides a snapshot of variables at a single point in time, time-series data reveals patterns across different temporal dimensions, making it highly relevant for forecasting tasks.

Time-series data exhibits two primary features, trend and seasonality, which are fundamental to effective predictive modelling. Trend reflects the long-term movement or trajectory of the data, which may indicate an overall growth or decline in a variable's behaviour over an extended period. This characteristic helps identify the direction of the system's evolution, thereby aiding strategic planning and optimisation efforts. On the other hand, seasonality refers to recurring patterns or cyclical behaviours observed at regular intervals, often driven by exogenous influences, such as climatic conditions or specific operational schedules.

2.2 Time Series Machine Learning Applications in HVAC systems

Time series machine learning techniques have been employed in many studies to optimise HVAC systems, achieving improvements in energy efficiency and prediction accuracy. For instance, a Long Short-Term Memory (LSTM) model has been applied to predict cooling loads, effectively capturing temporal dependencies in building HVAC systems.[15] Similarly, a hybrid model combining Convolutional Neural Networks (CNN) and LSTM have demonstrated superior accuracy and efficiency in predicting building HVAC system performance compared to traditional methods.[16] Moreover, Nonlinear Autoregressive Exogenous (NARX) models have been used to optimise and forecast building cooling loads, demonstrating enhanced forecasting accuracy through optimised parameter selection.[17]

Other than neural networks, there are other machine learning models have also proven effective for HVAC system performance forecasting and optimisation. For example, autoregressive integrated moving average (ARIMA) has been used to predict chillers performance in commercial buildings and investigate the most effective variables in improving predictability.[18] Support vector regression (SVR), has also shown high predictive accuracy in optimising and forecasting heating and cooling load in residential buildings.[19] These applications highlight the versatility of time series machine learning models in predicting and optimising performance of HVAC systems.

2.3 Machine Learning Model Selection for the Cooling System

Besides the abovementioned Machine Learning (ML) applications, in terms of algorithms, there are two tree-based ML algorithms suitable for this time series problem, Extreme Gradient Boosting (XGBoost) and Random Forest. The two algorithms were selected for this research because of their capability in handling complex data, managing non-linearity, and their robustness against overfitting.

XGBoost is a boosting algorithm that sequentially builds decision trees, where each tree corrects the errors through iterative corrections of the ensemble built thus far. A schematic diagram of XGBoost structure is shown in Figure 2. The main innovations of XGBoost over traditional gradient boosting include the addition of L1 and L2 regularisation terms to control model complexity and mitigate overfitting, parallelisation for enhanced computational efficiency, and an advanced algorithm to handle sparse data effectively.[20] XGBoost is particularly suited for time series prediction tasks due to its ability to model feature interactions and temporal dependencies effectively. Moreover, its inherent flexibility in feature engineering, including the integration of lagged features, makes it ideal for applications requiring the capture of seasonality and trends.

[insert Figure 2.]

Figure 2 – XGBoost structure

Recent studies have highlighted the successful application of XGBoost in building HVAC systems. An XGBoost-based predictive control strategy for HVAC systems was developed to provide day-ahead demand response, showing significant improvements in smart power grids efficiency.[21] A dynamic threshold enhanced XGBoost model was utilised for early detection of faults in HVAC systems, improving fault detection accuracy significant.[22] A hybrid particle swarm optimisation (PSO)-XGBoost model was applied for estimating the heating load of buildings, which proved to be highly effective for smart city planning.[23] Similarly, a data-driven predictive models based on XGBoost for residential building energy consumption was developed, achieving accurate predictions by introducing segregated heating and cooling days.[24]

Random Forest, in contrast, is a bagging method that constructs multiple decision trees using randomized feature subsets independently, each trained on a different subset of the training data and aggregates their predictions through averaging (for regression) or voting (for classification). A schematic diagram of Random Forest structure is shown in Figure 3. The introduction of randomness through feature selection at each split enhances diversity among individual trees, thereby reducing overfitting risk.[25] Random Forest is well-suited for time series forecasting of cooling systems due to its ability to manage high-dimensional data and its robustness against noise, making it effective for capturing complex patterns in operational data of cooling units.

[insert Figure 3.]

Figure 3 – Random Forest structure

Recent research further emphasises the versatility of Random Forest in building HVAC domain. A Random Forest-based method for cooling load disaggregation was utilised in smart buildings, demonstrating its potential for precise energy monitoring.[26] A hybrid model combining Random Forest and Support Vector Machine was developed for HVAC fault detection, which improved system's response time and accuracy.[27] A quantile Random Forest was applied to predict uncertainty in chiller power consumption, showcasing its effectiveness in commercial building scenarios.[28] Another hybrid Random Forest and Bayesian Inference model was investigated for simultaneous detection of sensor fault diagnosis and bias correction in data centre cooling systems, contributing to enhanced system reliability.[29]

The applications of both XGBoost and Random Forest in building environment studies demonstrate their effectiveness for the proposed research context. However, to the best of the authors' knowledge, these methods have not yet been applied to the highly efficient evaporative cooling systems. This research gap highlights the need for predictive models specifically tailored to these systems, with the potential to enhance forecasting accuracy and unlock further applications that optimise cooling efficiency and sustainability in data centres.

This study exploited a four-month dataset of operational data from a real-world 100 kW highly efficient evaporative cooling system in a data centre. Employing two machine learning algorithms, Random Forest and XGBoost, we endeavoured to construct a performance forecasting model to achieve precise hourly performance predictions for the cooling system across diverse climate conditions, elucidating the distinct impact levels of various input features.

3.0 Model Building

3.1 Data Collection

The dataset acquired from the operational data centre cooling system comprises parameters including dry-bulb temperature and relative humidity of supply air, return air, and ambient air. Additionally, it includes data on cooling capacity, Coefficient of Performance (COP), cumulative power consumption, cumulative running time, and corresponding timestamps. Detailed features (inputs) and targets (outputs) were listed in Table 1.

Features (Inputs)	RA Temp	Dry-bulb Temperature of Return Air
	SA Temp	Dry-bulb Temperature of Supply Air
	OA Temp	Dry-bulb Temperature of Ambient Air
	RA RH	Relative Humidity of Return Air
	SA RH	Relative Humidity of Supply Air
	OA RH	Relative Humidity of Ambient Air
Targets (Outputs)	Cooling Capacity	Cooling Capacity
	COP	Coefficient of Performance

Table 1 – Inputs and outputs of the forecasting model

Data was collected via pre-integrated wireless sensors and uploaded to the control system for analysis, detailed information of the sensors is listed in Table 2. To ensure data quality, sensors were calibrated prior to installation, with recalibrations conducted quarterly.

Sensors	Quantity	Location	Purposes	Accuracy
ALTA Wireless Humidity & Temperature Sensor - Coin Cell Powered	4	Inside data centre	Monitoring and recording temperature and relative humidity of hot and cold aisles	+/- 2% humidity accuracy (between 0% – 100% RH) +/- 0.5°C temperature accuracy (between 0°C–100°C)"
ALTA Wireless Humidity & Temperature Sensor - AA Battery Powered	12	Supply air duct, Return air duct, Exhaust air duct	Monitoring and recording temperature and relative humidity of supply, return, and exhaust air	+/- 2% humidity accuracy (between 0% – 100% RH) +/- 0.5°C temperature accuracy (between 0°C–100°C)"
ALTA Industrial (IP65) Wireless Humidity & Temperature	2	Outside data centre	Monitoring and recording temperature and relative humidity of ambient air	+/- 2% humidity accuracy (between 0% – 100% RH) +/- 0.5°C temperature accuracy (between 0°C–100°C)"
ALTA Wireless Air Velocity / Speed Sensor	12	Supply air duct, Return air duct, Exhaust air duct	Monitoring and recording air velocity of supply, return, and exhaust air	+/- 0.5 m/s

Table 2 – Details of integrated sensors

3.2 XGBoost Forecasting Model

To assess the effectiveness of the trained XGBoost model, performance metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared were applied. MSE measures the average of the squared differences between predicted and actual values, giving higher weight to larger errors. MAE represents the average of the absolute differences between predicted and actual values, providing a straightforward measure of prediction accuracy in the same units as the data. R-squared is a statistical metric that indicates the proportion of variance, with values closer to 1 indicating a better fit. These metrics, under the default settings, are outlined in Table 3, while Figure 4 presents the comparison between predicted and actual values.

Targets	MSE	MAE	R-squared
Cooling Capacity	3.197	1.339	-5.340
COP	79.742	6.502	-4.868

Table 3 – Performance metrics of XGBoost model's target variables in default setting

[insert Figure 4.]

Figure 4 – Prediction and actual values of XGBoost model's Cooling Capacity and COP of default setting

From both metrics, prediction and actual plot, the performance of this prediction is not satisfactory, a negative R-squared indicates that the prediction is no better than a straight line. To address this issue, a Hyperparameter Tuning Method was proposed.

Hyperparameter tuning is an essential step in the development of machine learning models, as it aims to identify the optimal combination of hyperparameters that maximise the model's predictive performance.[30] Hyperparameters are key settings that govern the model's learning process, such as the learning rate, maximum depth of decision trees, and the number of estimators, minimum samples split and leaf in the context of the two algorithms applied in this paper. Common hyperparameter tuning methods include grid search, random search, and Bayesian optimisation. Grid search systematically explores predefined hyperparameter values, while random search randomly samples combinations, often leading to efficient results in less time. Bayesian optimisation, on the other hand, uses probabilistic models to guide the search towards promising regions in the hyperparameter space, improving efficiency and accuracy.

Numerous studies have successfully implemented hyperparameter tuning for enhancing machine learning models in HVAC and cooling systems. For instance, grid search hyperparameter optimisation was applied to an interpretable neural network, demonstrating a significant improvement in predictive accuracy of building HVAC system.[31] Similarly, a random search enhanced multi-layer perceptron network was employed for estimating dynamic preconditioning time in residential buildings, resulting in considerable energy saving and system payback period reduction.[32] Bayesian optimisation hyperparameter tuning was integrated of a data-driven model predictive controls for near-optimal control performance, achieving notable energy efficiency gains in HVAC operations compared with the original building automation control.[33]

Grid search systematically explores all possible combinations of hyperparameters within specified ranges, providing a more comprehensive coverage of the search space compared other methods. As the results with default hyperparameters are demonstrating considerable potential for exhaustive improvement, grid search is selected for optimal tuning efficacy.

After performing hyperparameter tuning with grid search, the performance metrics and prediction with actual values are presented in Table 4 and Figure 5.

Targets	MSE	MAE	R-squared
Cooling Capacity	0.728	0.654	0.543
COP	33.780	3.771	0.485

Table 4 – Performance metrics of XGBoost model's target variables after hyperparameter tuning

[insert Figure 5.]

Figure 5 – Prediction and actual values of XGBoost model's Cooling Capacity and COP after hyperparameter tuning

Substantial enhancements can be observed in both metrics, the prediction, and actual plot. Nevertheless, various strategies can be explored for further refinement, including 1) Incorporating temporal features: investigating the integration of temporal aspects such as day of the week, time of day, or seasonal indicators to capture nuanced patterns not discernible with the existing features, and 2) Feature selection: reassessing the significance of features and contemplating the exclusion or combination of features with low importance. This approach has the potential to streamline the model and potentially elevate its performance.

To gain a deeper comprehension of the interplay between features and target variables, Figure 6 and Figure 7 depict the feature importance and correlation matrix of the XGBoost forecasting model.

[insert Figure 6.]

Figure 6 – Feature importance of XGBoost model's Cooling Capacity and COP

[insert Figure 7.]

Figure 7 – Correlation matrix of XGBoost model's Cooling Capacity and COP

The feature importance diagrams reveal that cooling capacity is predominantly influenced by return air relative humidity (RA RH), whereas the most influential feature for COP is outdoor air relative humidity (OA RH). The correlation matrix indicates that all input features exhibit relatively weak correlations with the targets. However, RHs generally show stronger correlations compared to temperatures

The dominance of RH over temperature can be attributed to the cooling mechanism in evaporative cooling systems, which relies on water evaporation. During this process, heat is absorbed as water transitions from liquid to vapour. The efficiency of this phase transition is primarily dictated by the air's RH: lower RH enables greater evaporation and better cooling potential, while higher RH reduces the system's cooling effectiveness. In contrast, temperature primarily reflects the heat load in the system by influencing air density and flow rate. However, its impact is less direct because the system's effectiveness hinges on air's ability to absorb moisture, which is a factor governed by RH.

In terms of feature importance, a key innovation of this highly efficient evaporative cooling system is the introduction of a small portion of outdoor air to utilise the natural cooling source. The majority of the air, however, is return air recirculated from the data centre, making RA RH the primary determinant of cooling capacity. For COP, the deciding factor is the balance between cooling output and energy input. OA RH determines the system's natural cooling potential, which indicates its capacity to achieve free cooling through evaporation. Lower OA RH enhances this potential, allowing the system to deliver the same cooling output with reduced energy input, thereby improving COP.

The correlation matrix shows generally weak correlations between input features and targets, indicating the presence of non-linear relationships. This supports the use of XGBoost, a model well-suited to capturing complex, non-linear patterns in the dataset. For further fine-tuning or training of more sophisticated deep learning models for similar evaporative cooling systems, future work should focus on:

1. Exploring non-linear relationships between weakly correlated features and targets.
2. Integrating or eliminating excessively correlated features to improve model interpretability, performance, and efficiency.

3.3 Random Forest Forecasting Model

Same features and target variables were selected as the XGBoost forecasting model in Table 1, the performance metrics of default setting are presented in Table 5.

Targets	MSE	MAE	R-squared
Cooling Capacity	0.567	0.388	0.849
COP	132.271	2.806	0.491

Table 5 – Performance metrics of Random Forest model's target variables of default settings

While after hyperparameter tuning using grid search, the refined performance metrics of default setting are presented in Table 6.

Targets	MSE	MAE	R-squared
Cooling Capacity	0.569	0.388	0.848
COP	31.98	2.215	0.541

Table 6 – Performance metrics of Random Forest model's target variables after hyperparameter tuning

In the context of the Random Forest model, fine-tuning hyperparameters consistently enhances forecasting accuracy for Cooling Capacity and COP. The results also demonstrates that the enhancement in forecasting precision is more significant for COP than for Cooling Capacity.

4.0 Application Scenario Analysis

The existing training and testing dataset collected during the cold and humid winter in the UK exhibit distinct seasonal and regional patterns. With data acquired from the Met Office,[34] two box charts illustrating annual trends of temperature and relative humidity in the UK are shown in Figure 8 and Figure 9. As shown in the figures, the climatic conditions from September through December are closely aligned with climate conditions from January to May. During these months, temperatures and relative humidity levels tend to be moderate, conditions under which evaporative coolers operate effectively. This seasonal overlap suggests that the dataset collected from September to December can accurately represent conditions in January to May. By using this dataset to train our forecasting model, we established a representative foundation for predicting performance across both periods, allowing the model to provide reliable forecasts from September to May. Research on forecasting methods for time series data highlights that models capturing seasonal patterns can effectively predict future values during comparable seasonal periods as well.[35]

Due to the late August installation of the evaporative cooling units, data is unavailable for the summer months (June to August), during which the data centre was cooled with vapour compression systems. This limits the model's current applicability to the moderate seasons. Future work will focus on gathering summer data to enable comprehensive, year-round performance forecasting.

[insert Figure 8.]

Figure 8 – Annual temperature trend in the UK

[insert Figure 9.]

Figure 9 – Annual relative humidity trend in the UK

With an ample set of input features, the previously developed machine-learning models now possess the capability to forecast operational (cooling capacity) and performance (COP) parameters (targets) for regions sharing similar climate conditions with the 100 kW evaporative data centre cooling system, which confirms the feasibility of the models' application.

Under optimal conditions, with an extensive annual dataset spanning at least a year and encompassing diverse seasonal trends, including moderate spring and autumn, hot and dry summer, and cold and humid winter, these machine learning models could achieve enhanced accuracy in capturing the dynamic operational conditions of various data centres equipped with different types of cooling systems, extending beyond evaporative cooling. Moreover, in such scenarios, the input features could be confined to outdoor weather conditions only, specifically ambient temperature and humidity. This streamlined approach empowers the model to forecast hourly operational and performance parameters solely based on historical and forecasted weather conditions provided by meteorological agencies, such as the UK Met Office.

The goal is to assist in determining the cost investment and payback period from the designing stage of any data centres with various cooling systems and locations.

5.0 Conclusion

This study investigated the application of time-series machine learning models to enhance the energy efficiency of data centres by accurately forecasting the performance of a highly efficient 100 kW evaporative cooling system. Using a dataset collected over four months, the study developed and optimised two predictive models, XGBoost and Random Forest, evaluating their performance using key metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared. Through a comprehensive process of model training, evaluation, and hyperparameter tuning, the study aimed to optimise the accuracy of these predictions and contribute to more sustainable data centre operations.

The initial models with default settings showed suboptimal performance, with the XGBoost model yielding an MAE of 1.339 for cooling capacity and 6.502 for COP, along with a negative R-squared, indicating poor predictive accuracy. After hyperparameter tuning, the Random Forest model significantly improved, achieving an MAE of 0.388 and an R-squared of 0.849 for cooling capacity, and an MAE of 2.215 and an R-squared of 0.541 for COP, demonstrating its superior ability to accurately forecast cooling performance.

This optimisation presents opportunities for additional energy-saving measures. Further endeavours could focus on in-depth data exploration, advanced feature engineering, meticulous cross-validation, and a more resilient model selection process. Pursuing these strategies has the potential to achieve even more precise forecasting outcomes. Currently, in the given circumstances, the Random Forest model outperforms XGBoost in managing this time-series dataset, exhibiting lower MSE, MAE, and higher R-squared for Cooling Capacity. It is noteworthy, however, that both models encounter challenges in forecasting COP.

The current models capture seasonal trends effectively, allowing for accurate predictions across all evaporative cooling months (September to May). Further efforts will involve expanding the dataset to encompass a broader temporal and spatial scope, including summer months (June to August), to enable comprehensive, year-round performance forecasting. To further extend the applicability of the developed forecasting models, future work will aim to explore non-linear relationships of weakly correlated features, integrate input features with excessive correlations, and refine the models to support the design of data centres and associated cooling systems through precise hourly predictions.

Additionally, preliminary findings indicate that Water Usage Efficiency (WUE) is a critical factor in evaporative cooling systems. Future research will involve experimental data collection on WUE to enable the model to accurately forecast water consumption.

CRediT Author Statement

Zhichu Wang: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Cheng Zeng: Conceptualization, Formal analysis, Methodology, Supervision, Investigation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Zishang Zhu: Methodology, Investigation

Yunhai Li: Investigation, Formal Analysis

Xiaoli Ma: Investigation, Formal Analysis

Xudong Zhao: Conceptualization, Supervision, Funding acquisition

Reference

1. National Grid ESO. [Internet]. Available from: https://www.nationalgrideso.com/document/246446/download (accessed 1 December 2023).

2. International Energy Agency (IEA). Data Centres & Networks [Internet]. Available from: https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks (accessed 1 December 2023).

3. Wan J, Gui X, Kasahara S, Zhang Y, Zhang R. Air flow measurement and management for improving cooling and energy efficiency in raised-floor data centers: A survey. *IEEE Access*. 2018;6:48867-901. doi:10.1109/access.2018.2866840

4. Xu P, Ma X, Zhao X, Fancey K. Experimental investigation of a super performance dew point air cooler. *Appl Energy*. 2017;203:761-77. doi:10.1016/j.apenergy.2017.06.095

5. Zhang Q, Meng Z, Hong X, Zhan Y, Liu J, Dong J, et al. A survey on Data Center Cooling Systems: Technology, Power Consumption Modeling and control strategy optimization. *J Syst Archit*. 2021;119:102253. doi:10.1016/j.sysarc.2021.102253

6. Ran Y, Hu H, Zhou X, Wen Y. DeepEE: Joint optimization of Job Scheduling and cooling control for data center energy efficiency using Deep Reinforcement Learning. *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. 2019. doi:10.1109/icdcs.2019.00070

7. Wang R, Zhang X, Zhou X, Wen Y, Tan R. Toward physics-guided safe deep reinforcement learning for Green Data Center Cooling Control. *2022 ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS)*. 2022 May. doi:10.1109/iccps54341.2022.00021

8. Li Y, Wen Y, Tao D, Guan K. Transforming cooling optimization for Green Data Center via deep reinforcement learning. *IEEE Trans Cybern*. 2020 May;50(5):2002-13. doi:10.1109/tcyb.2019.2927410

9. Zhang W, Wen Y, Wah Wong Y, Chuan Toh K, Chen C-H. Towards joint optimization over ICT and cooling systems in Data Centre: A Survey. *IEEE Commun Surv Tutor*. 2016;18(3):1596-616. doi:10.1109/comst.2016.2545109

10. Afram A, Janabi-Sharifi F. Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods. *Energy Build*. 2015 May 1;94:121-49.

11. Krause J, Perer A, Ng K. Interacting with predictions: Visual inspection of black-box machine learning models. *2016 CHI Conference on Human Factors in Computing Systems*. 2016 May 7. p. 5686-97.

12. Brown LD, Call AC, Clement MB, Sharp NY. Inside the “black box” of sell-side financial analysts. *J Account Res*. 2015 Mar;53(1):1-47.

13. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. *ACM Comput Surv*. 2018 Aug 22;51(5):1-42.

14. Liang Y, Li S, Yan C, Li M, Jiang C. Explaining the black-box model: A survey of local interpretation methods for deep neural networks. *Neurocomputing*. 2021 Jan 2;419:168-82.

15. Zhou Z, Huang G, Chen Y. Machine Learning Models for Predictive Cooling of HVAC Systems in Buildings: A Review. Energy Reports. 2021;7:2025-2041.

16. HAL. Prediction of HVAC System Parameters Using Deep Learning. HAL Science. 2023.

17. Kim JH, Seong NC, Choi W. Cooling Load Forecasting via Predictive Optimization of a Nonlinear Autoregressive Exogenous (NARX) Neural Network Model. Sustainability. 2019;11(23):6535.

18. Ho WT, Yu FW. Predicting chiller system performance using ARIMA-regression models. J Build Eng. 2021 Jan 1;33:101871.

19. Cai W, Wen X, Li C, Shao J, Xu J. Predicting the energy consumption in buildings using the optimized support vector regression model. Energy. 2023 Jun 15;273:127188.

20. 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. p. 785-794.

21. Wang H, Chen Y, Kang J, Ding Z, Zhu H. An XGBoost-Based predictive control strategy for HVAC systems in providing day-ahead demand response. Building and Environment. 2023 Jun 15;238:110350.

22. Chakraborty D, Elzarka H. Early detection of faults in HVAC systems using an XGBoost model with a dynamic threshold. Energy and Buildings. 2019 Feb 15;185:326-44.

23. Le LT, Nguyen H, Zhou J, Dou J, Moayedi H. Estimating the heating load of buildings for smart city planning using a novel artificial intelligence technique PSO-XGBoost. Applied Sciences. 2019 Jul 4;9(13):2714.

24. Kamel E, Sheikh S, Huang X. Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days. Energy. 2020 Sep 1;206:118045.

25. Breiman L. Random Forests. Machine Learning. 2001;45:5-32.

26. Xiao Z, Gang W, Yuan J, Zhang Y, Fan C. Cooling load disaggregation using a NILM method based on random forest for smart buildings. Sustainable Cities and Society. 2021 Nov 1;74:103202.

27. Tun W, Wong JK, Ling SH. Hybrid random forest and support vector machine modeling for HVAC fault detection and diagnosis. Sensors. 2021 Dec 7;21(24):8163.

28. Rizi BS, Pavlak G, Cushing V, Heidarinejad M. Predicting uncertainty of a chiller plant power consumption using quantile random forest: A commercial building case study. Energy. 2023 Nov 15;283:129112.

29. Wang J, Tian Y, Qi Z, Zeng L, Wang P, Yoon S. Sensor fault diagnosis and correction for data center cooling system using hybrid multi-label random Forest and Bayesian Inference. Building and Environment. 2024 Feb 1;249:111124.

30. Bardenet R, Brendel M, Kégl B, Sebag M. Collaborative hyperparameter tuning: Proceedings of the 30th International Conference on Machine Learning - volume 28 [Internet]. *The ACM Digital Library*; 2013. Available from: https://dl.acm.org/doi/10.5555/3042817.3042916 (accessed 5 February 2024).

31. Wang M, Wang Z, Geng Y, Lin B. Interpreting the neural network model for HVAC system energy data mining. Building and Environment. 2022 Feb 1;209:108449.

32. Esrafilian-Najafabadi M, Haghighat F. Occupancy-based HVAC control using deep learning algorithms for estimating online preconditioning time in residential buildings. Energy and Buildings. 2021 Dec 1;252:111377.

33. Lin X, Guo Q, Yuan D, Gao M. Bayesian Optimization Framework for HVAC System Control. Buildings. 2023 Jan 20;13(2):314.

34. Met Office. UK climate averages. Met Office. Available from: https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-climate-averages [Accessed 4 Nov 2024].

35. Liu Z, Zhu Z, Gao J, Xu C. Forecast methods for time series data: a survey. Ieee Access. 2021 Jun 21;9:91896-912.

List of Figures

List of Tables