International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 11 | Nov 2026 www.irjet.net p-ISSN: 2395-0072

Green AI Evaluation of Data-Centre Facility-Power Prediction: Accuracy, Computational Cost, and Environmental Impact

Soumya Shetty[1], Triveni Katkar[2], Tanuja Patil[3], Dr. Saloni Bhushan[4]

[123]MCA Department, D. Y. Patil Deemed to be University, School of Humanities & Sciences, Navi Mumbai, India
[4]Associate Professor, School of Humanities and Sciences, D. Y. Patil Deemed to be University, School of Humanities & Sciences, Navi Mumbai, India

---

Abstract - *The increasing use of Artificial Intelligence (AI) in data-centre operations has highlighted the importance of balancing predictive accuracy with computational energy efficiency. This study evaluates this trade-off by comparing three machine-learning models Linear Regression, Random Forest, and XGBoost for short-term facility-power forecasting using publicly available operational data from the Frontier supercomputer. The prepared dataset contained 48,816 observations, divided chronologically into 34,171 training, 7,322 validation, and 7,323 test observations. Two feature configurations were examined: full operational telemetry and a cooling/workload-focused configuration. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). Linear Regression achieved the best observed test performance in both configurations. Under the cooling/workload-focused configuration, it recorded an MAE of 0.025657 MW, an RMSE of 0.038213 MW, and an R^2 of 0.730887. Paired statistical tests with Holm correction identified differences in absolute prediction errors; however, the moving-block bootstrap confidence interval for the focused Linear Regression–XGBoost comparison included zero, indicating uncertainty in the estimated difference. Computational energy consumption during training and inference was estimated using CodeCarbon on an Apple MacBook Air M4. These estimates remain provisional because a concurrent-tracker warning requires further verification. The findings indicate that Linear Regression offers a favourable balance between predictive accuracy and estimated computational cost under the tested conditions. However, the study forecasts facility-related power demand rather than directly measuring cooling-system electricity consumption. It therefore does not establish actual reductions in facility energy use, cooling demand, water consumption, or carbon emissions.*

Keywords: Green AI, facility-power prediction, machine-learning efficiency, computational energy, carbon-emissions estimation, Frontier supercomputer.

1. INTRODUCTION

Machine-learning models require computational resources for training and prediction, making their energy consumption relevant when assessing their overall efficiency [1], [2]. This consideration is particularly important in data centres, where computing workloads and supporting infrastructure both contribute to electricity demand. This study uses publicly available operational data from the Frontier supercomputer [11] to predict Facility Accessory Power at the next recorded timestamp, usually around ten minutes after the current observation. Linear Regression, Random Forest, and XGBoost are evaluated to determine how accurately they predict this facility-related power demand.

The experiments use two feature configurations: Full Operational Telemetry and Cooling/Workload-Focused Telemetry. Comparing them helps assess whether a narrower set of cooling- and workload-related variables can provide useful predictions without relying on the complete set of operational measurements. This distinction is important because Frontier Total Power includes Facility Accessory Power as a component, so using it as an input could introduce target leakage. The study evaluates predictive accuracy alongside estimated energy consumption during model training and inference. It does not directly measure cooling-energy savings or test a real-time cooling-control system; its findings are limited to the forecasting experiments conducted on the available dataset.

2. PROBLEM STATEMENT

A model's predictive accuracy does not, by itself, indicate how much energy is required to train and run it.This study addresses this limitation by comparing three regression models Linear Regression, Random Forest, and XGBoost for predicting the next recorded Facility Accessory Power value. Their prediction errors and estimated training and inference energy are assessed under two feature configurations: Full Operational Telemetry and Cooling/Workload-Focused Telemetry. The comparison also considers whether including Frontier Total Power affects the reliability of the results, since total power contains Facility Accessory Power as a component. In this study, Facility Accessory Power is used as an indicator of facility-related demand; it is not treated as a direct measurement of cooling-only energy.

3. LITERATURE REVIEW

3.1 Green AI and Sustainable Computing

The growing adoption of artificial intelligence has increased the computational resources required for model development, training, and deployment. Schwartz et al. [1] introduced the concept of Green AI, emphasizing the importance of considering computational costs alongside model performance. Henderson et al. [2] highlighted the need for systematic reporting of energy consumption and carbon emissions associated with machine learning experiments. Similarly, Lannelongue et al. [3], [4] discussed approaches for quantifying the environmental footprint of computation and promoting sustainable computing practices.

Strubell et al. [5] examined the energy and policy implications of computationally intensive deep learning models, highlighting the environmental costs associated with model training. Anthony et al. [6] introduced Carbontracker, a tool for tracking and estimating the energy consumption and carbon footprint of machine learning training. Desislavov et al. [7] extended this discussion to AI inference, emphasizing the importance of considering energy consumption during model deployment and operation.

3.2 Machine Learning for Data-Centre Energy and Cooling Optimisation

Data centres require substantial electrical power to operate computing infrastructure and maintain suitable thermal conditions. Consequently, cooling systems and thermal management have become important areas of research for improving data-centre energy efficiency. Chen et al. [8] investigated machine-learning-based parameter prediction and optimisation for data-centre heat dissipation systems. Mahbod et al. [9] examined energy-saving evaluation through a model-free reinforcement learning approach, demonstrating the application of intelligent methods to data-centre energy management.

Further research has explored the combination of machine learning with physical principles for predicting the energy consumption of data-centre cooling systems. Li et al. [10] investigated an approach combining physical laws and artificial neural networks for cooling-energy prediction. Ma et al. [12] examined AI-enabled predictive energy-saving planning for liquid-cooling systems. Jadhav and Liu [14] also explored machine-learning-guided cooling-system optimisation for data centres.

3.3 Energy-Aware Workload Management and Sustainable Technology

In addition to cooling-system optimisation, workload management is an important factor in improving data-centre energy efficiency. Smith et al. [13] investigated machine-learning-based energy-efficient workload management, highlighting the role of intelligent workload allocation in sustainable data-centre operations.

The broader development of connected infrastructure and advanced analytics also provides a foundation for intelligent monitoring and sustainable technology applications. Bhushan et al. [15] discussed advanced analytical frameworks in the context of the Industrial Internet of Things. Shanker and Bhushan [16] examined the relationship between technological advancement and sustainable city development. These studies provide a broader technological and sustainability context for data-driven approaches to infrastructure-related challenges.

Table 1. Summary of Related Literature

Study	Main contribution	Relevance to this study
Schwartz et al. [1]	Research perspective that accounts for computational cost.	Supports evaluating predictive performance alongside resource requirements.
Henderson et al. [2]	Reporting energy use and emissions from machine-learning workloads.	Provides context for measuring and reporting computational energy.
Lannelongue et al. [3]	Green Algorithms framework for estimating computational environmental impact.	Provides background for environmental-footprint estimation.
Chen et al. [8]	Machine-learning-based prediction and optimisation of data-centre heat dissipation.	Provides context for machine learning applications in data-centre thermal management.
Li et al. [10]	Research on predicting data-centre cooling energy.	Relates to forecasting energy demand in facility systems.
Sun et al. [11]	Frontier operational dataset.	Provides the data used in this research.

Ma et al. [12]	Predictive planning related to liquid-cooling energy savings.	Connects forecasting with data-centre energy planning.
Smith et al. [13]	Machine-learning-based energy-efficient workload management.	Relates to the use of machine learning for improving data-centre energy efficiency.
Jadhav and Liu [14]	Research on machine-learning-based cooling optimization.	Provides context for AI applications in cooling-system management.

4. RESEARCH GAP, OBJECTIVES, AND QUESTIONS

4.1 Research Gap

Existing research examines Green AI, data-centre power prediction, and cooling-system optimisation. However, the present study focuses on comparing predictive performance and estimated computational energy under a consistent experimental setup for the specific task of forecasting Frontier Facility Accessory Power. It evaluates three regression models across two feature configurations and considers how including Frontier Total Power may affect the interpretation of predictive results because total power contains the target as a component. The contribution is a task-specific empirical comparison of accuracy and estimated workload cost, rather than a new prediction algorithm or a demonstration of actual cooling-energy savings.

4.2. Research Objectives

The study aims to achieve the following objectives:

1. To forecast the next recorded Facility Accessory Power value using operational data from the Frontier supercomputer.
2. To evaluate Linear Regression, Random Forest, and XGBoost using chronological training, validation, and test splits.
3. To compare model predictions using mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2).
4. To estimate the computational energy associated with model training and inference using CodeCarbon.
5. To compare predictive performance and estimated computational energy under Full Operational Telemetry and Cooling/Workload-FocusedTelemetry configurations.
6. To investigate the potential influence of target leakage and assess the limitations of the computational-energy estimates.

4.3. Research Questions

The study addresses the following research questions:

RQ1: How accurately can Linear Regression, Random Forest, and XGBoost forecast the next recorded Facility Accessory Power value using Frontier operational data?
RQ2: How do the three models compare in predictive performance and estimated computational energy under the experimental conditions?
RQ3: How does the choice of feature configuration, particularly the inclusion of Frontier Total Power, affect prediction accuracy and the interpretation of model results?
RQ4: Which model offers the most appropriate balance between predictive performance and estimated computational energy for this forecasting task?

5. HYPOTHESIS

For each pairwise comparison of models within each feature configuration, the null hypothesis (H_0) states that there is no systematic difference in their absolute prediction errors, while the alternative hypothesis (H_1) states that a difference exists. The paired Wilcoxon signed-rank tests use a significance level of 0.05, with Holm correction applied to account for multiple comparisons. The study also examines whether Linear Regression can provide competitive predictive accuracy with lower estimated computational energy than Random Forest and XGBoost. This expectation is evaluated using the experimental results rather than assumed in advance. Statistical conclusions are interpreted alongside effect sizes and confidence intervals, rather than relying on p-values alone.

6. METHODOLOGY

6.1. Dataset and Preparation

The experiments use the Frontier HPC & Facility Data dataset described by Sun et al. [11]. The dataset contains operational records collected during 2023, with most observations approximately ten minutes apart. Data preparation involved removing the metadata row and two empty columns and converting the timestamp field into datetime format. The resulting modelling dataset contained 48,816 observations across 17 columns. Overall Coolant Flow, Overall Waste Heat, and Power Usage Effectiveness (PUE) were omitted because the exported file did not provide valid measurement series for these variables.

The forecasting target was Frontier Facility Accessory Power, measured in megawatts (MW). For each observation, the models estimated the target at the following recorded timestamp. Although this interval was generally around ten minutes, missing records meant that some prediction intervals were longer.

Facility Accessory Power was used to represent demand associated with facility operations, including cooling and auxiliary systems. It was not treated as a direct measurement of cooling-system electricity consumption.

To preserve the chronological order of the data, 70% of the observations were assigned to training, 15% to validation, and the remaining 15% to testing. These sets contained 34,171, 7,322, and 7,323 observations, respectively. The final performance comparison used predictions from the test set, which was held out from model training and validation.

6.2. Feature Scenarios

The experiments used two input configurations to examine how the selection of operational measurements affected forecasting performance.

The Cooling/Workload-Focused Telemetry configuration contained 14 features. These included compute power, average coolant return temperature, coolant supply temperature, coolant-flow readings from the three subloops, three lagged accessory-power values (lags 1, 3, and 6), two lagged compute-power values (lags 1 and 3), hour, day of the week, and month.

The **Full Operational Telemetry** configuration contained the same 14 features and added Frontier Total Power, resulting in 15 input features. Because Total Power is the sum of compute power and Facility Accessory Power, including it may expose information directly related to the target. Performance under this configuration must therefore be interpreted with the possibility of target leakage in mind.

The focused configuration removes Total Power but retains previous accessory-power readings. It assesses whether historical observations and other operational variables can support short-term forecasting. Neither configuration, by itself, establishes a causal relationship between cooling-system behaviour and facility power demand.

6.3. Models

The study compared three regression algorithms: Linear Regression, Random Forest, and XGBoost.

Linear Regression was implemented using the standard scikit-learn estimator. The Random Forest model used 50 trees, a maximum tree depth of 12, a minimum of two observations per leaf, random_state=42, and n_jobs=2.

The XGBoost model used 300 boosting iterations, a maximum depth of 6, and a learning rate of 0.05. Both row subsampling and feature subsampling were set to 0.8. The model used the squared-error objective, random_state=42, and n_jobs=2.

The specified parameters were kept consistent between the predictive-performance experiments and the energy-estimation runs. This helped maintain comparable experimental settings across the three algorithms.

6.4. Evaluation and Statistical Testing

Forecasting accuracy was evaluated using mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2). MAE summarises the average magnitude of prediction errors in MW. RMSE gives greater weight to large errors, making it useful for identifying models that occasionally produce substantial deviations. R^2 measures predictive performance relative to a baseline that predicts the mean target value.

To compare models at corresponding timestamps, paired Wilcoxon signed-rank tests were applied to their absolute prediction errors. Holm correction was used to adjust the significance tests for multiple comparisons.

Moving-block bootstrap confidence intervals were also calculated using blocks of 144 consecutive observations. At the usual ten-minute sampling interval, a block represents approximately 24 hours of data. This approach retains some of the temporal structure that would be lost if individual observations were resampled independently.

All reported predictive results are based on a single chronological test period. The findings therefore describe performance on that period and should not be taken as proof that the same ranking will hold under different workloads, operating conditions, or time periods.

6.5. Experimental Environment and Energy Estimation

All experiments were performed on an Apple MacBook Air M4 with an ARM64 processor architecture, 16 GB of RAM, and 10 logical CPU cores. CodeCarbon was used to estimate the energy associated with model fitting and prediction, together with the corresponding carbon-dioxide-equivalent (CO_2e) emissions.

To obtain estimates for repeated operations, measurements were collected in sequential runs lasting at least 30 seconds. The experimental procedure included a minimum of 20 model-training fits and 100 complete prediction runs on the test dataset. For each model, the accumulated energy estimate was divided by the number of completed runs to calculate the average energy per training fit and per full test-set prediction run. Since the test set contained 7,323 observations, each prediction run covered the entire set rather than a single observation.

CodeCarbon sampled energy-related information at one-second intervals. The resulting figures are estimates produced by software, not readings obtained from a dedicated power meter. During the experiments,

CodeCarbon reported that more than one tracker could operate at the same time. This raises uncertainty about whether the energy attributed to each operation was isolated correctly. The estimates should therefore be regarded as preliminary until the tracking process has been checked and the experiments repeated with tracker isolation confirmed.

These measurements describe the computational cost of running the experiments on the specified MacBook Air. They do not directly measure energy consumption on the Frontier supercomputer, and they should not be used as estimates of Frontier's operational energy demand.

7. RESULTS AND DATA ANALYSIS

7.1. Predictive Performance

Table 2 reports the test-set performance of Linear Regression, Random Forest, and XGBoost under the Full Operational Telemetry and Cooling/Workload-Focused Telemetry configurations. The models were evaluated using MAE, RMSE, and R^2.

Table 2. Predictive Performance on the Test Set

Scenario	Model	MAE (MW)	RMSE (MW)	R^2
Full Operational Telemetry	Linear Regression	0.020017	0.031665	0.815214
Full Operational Telemetry	Random Forest	0.031040	0.043083	0.657924
Full Operational Telemetry	XGBoost	0.028202	0.039543	0.711824
Cooling/Workload-Focused Telemetry	Linear Regression	0.025657	0.038213	0.730887
Cooling/Workload-Focused Telemetry	Random Forest	0.031019	0.043097	0.657699
Cooling/Workload-Focused Telemetry	XGBoost	0.027888	0.039125	0.717894

Linear Regression recorded the lowest MAE and RMSE and the highest R^2 in both configurations. Its performance was stronger with Full Operational Telemetry than with the focused features. In contrast, XGBoost performed slightly better with the focused configuration, while Random Forest produced the largest prediction errors in both scenarios.

The difference between the two Linear Regression results requires careful interpretation. Frontier Total Power includes Facility Accessory Power as a component, so using it as an input may expose information directly related to the target. The stronger results obtained with Full Operational Telemetry therefore cannot be attributed solely to better forecasting capability. The focused configuration avoids this particular source of leakage and provides a more defensible basis for comparing performance using the selected features.

7.2. Estimated Computational Energy

Table 3 summarises the estimated energy required for one model-training fit and one complete test-set inference pass. It also reports the estimated CO_2-equivalent emissions associated with training.

Table 3. Estimated Computational Energy and Training Emissions

Scenario	Model	Training Energy per Fit (kWh)	Inference Energy per Test Pass (kWh)	Training CO_2e (kg/run)
Full Operational Telemetry	Linear Regression	2.089601×10^{-7}	1.198962×10^{-8}	1.490807×10^{-7}
Full Operational Telemetry	Random Forest	1.362094×10^{-4}	5.487080×10^{-7}	9.717738×10^{-5}
Full Operational Telemetry	XGBoost	2.093144×10^{-5}	3.896958×10^{-7}	1.493335×10^{-5}
Cooling/Workload-Focused Telemetry	Linear Regression	2.075299×10^{-7}	1.211588×10^{-8}	1.480603×10^{-7}
Cooling/Workload-Focused Telemetry	Random Forest	1.507841×10^{-4}	5.615851×10^{-7}	1.075755×10^{-4}

Cooling/ Workload -Focused Telemetry	XGBoost	2.14377×10^{-5}	4.04533×10^{-7}	1.52945×10^{-5}

The reported estimates place Linear Regression lowest in both training and inference energy, followed by XGBoost and Random Forest. Random Forest has the highest estimated training energy in both configurations. The same ordering appears in the inference estimates.

These values require qualification. CodeCarbon reported that multiple tracker instances were permitted to run concurrently, creating uncertainty about whether energy consumption was attributed correctly to each operation. The figures should therefore be treated as preliminary until the tracking setup has been checked and the measurements repeated. They describe the experiments conducted on the MacBook Air and should not be interpreted as measurements of energy consumed by the Frontier supercomputer.

7.3. Overall Interpretation

Across the tested configurations, Linear Regression achieved the strongest predictive performance and the lowest reported computational-energy estimates. XGBoost performed better than Random Forest on the reported prediction metrics, while Random Forest had the highest estimated training and inference energy.

The focused configuration offers a more cautious basis for interpreting the forecasting results because it excludes Frontier Total Power, which contains the target as a component. However, it still uses historical Facility Accessory Power values, so the experiment evaluates short-horizon forecasting rather than causal cooling-system behaviour.

The findings are limited to the available chronological test period and the experimental setup used. The energy estimates also remain provisional because the concurrent-tracker warning has not yet been resolved. Consequently, the results support a comparison of predictive accuracy and estimated model-computation costs, but they do not demonstrate reductions in actual facility electricity use, cooling energy, or water consumption.

8. STATISTICAL HYPOTHESIS TESTING

Differences between model predictions were examined using paired Wilcoxon signed-rank tests on absolute errors calculated at matching timestamps. Holm correction was applied to the resulting p-values to account for multiple comparisons. All adjusted p-values were reported as below 0.001. The analysis also used a moving-block bootstrap with 2,000 resamples and blocks of 144 observations to estimate 95% confidence intervals for differences in mean absolute error (MAE). In the comparisons below, a negative MAE difference means that the first model had a lower MAE than the second.

For Full Operational Telemetry, Linear Regression produced a lower MAE than Random Forest, with a difference of −0.011023 MW (95% CI: −0.014827 to −0.007553). Compared with XGBoost, its MAE was lower by 0.008185 MW (95% CI: −0.010480 to −0.006091). Random Forest also had a higher MAE than XGBoost, with a difference of 0.002838 MW (95% CI: 0.001243 to 0.004515). None of these confidence intervals crossed zero, indicating that the bootstrap analysis supported differences in MAE for all three model pairs under this feature configuration.

For Cooling/Workload-Focused Telemetry, Linear Regression outperformed Random Forest by an MAE difference of −0.005362 MW (95% CI: −0.009327 to −0.001780). Its MAE was also lower than XGBoost's by 0.002231 MW, but the confidence interval ranged from −0.004630 to 0.000062 and included zero. The bootstrap analysis therefore did not establish a clear difference between these two models at the 95% confidence level. Although the Holm-adjusted Wilcoxon result was reported as below 0.001, the two procedures assess the data differently and should not be treated as interchangeable. Random Forest had a higher MAE than XGBoost by 0.003131 MW (95% CI: 0.001595 to 0.004809), and this interval excluded zero.

Overall, the bootstrap results supported MAE differences for most model pairs, while the focused Linear Regression–XGBoost comparison remained uncertain under that analysis. The confidence intervals help show the size and uncertainty of the observed differences rather than relying on p-values alone. These statistical results concern prediction errors only; they do not demonstrate environmental benefits, verified energy savings, or reductions in cooling-energy or water consumption.

9. DISCUSSION

Linear Regression recorded the lowest test error and estimated computational energy among the three models under the experimental settings. In the Cooling/Workload-Focused configuration, its MAE was 0.025657 MW, compared with 0.027888 MW for XGBoost and 0.031019 MW for Random Forest. These results suggest that the more complex models did not provide better predictive accuracy for this dataset and forecasting task. However, the bootstrap confidence interval for the Linear Regression–XGBoost comparison included zero, so the difference in their MAE remains uncertain under that analysis. The estimated energy results are also provisional because of the unresolved CodeCarbon tracker warning.

The Cooling/Workload-Focused configuration offers a more defensible basis for comparison because it excludes Frontier Total Power, which contains Facility Accessory Power as a component and may introduce target leakage. However, the focused configuration still uses historical accessory-power measurements, so its results demonstrate short-horizon forecasting rather than causal insight into cooling-system behaviour. Further testing across different operating periods, together with verified energy measurements, is needed to determine whether these findings generalise beyond the current experimental setup.

10. LIMITATIONS

This study evaluates facility-power forecasting rather than the direct effects of AI on data-centre sustainability. It does not measure cooling-system electricity consumption, water use, changes in Power Usage Effectiveness (PUE), or actual reductions in facility energy demand. None of the models was deployed in a live cooling-control system.

The computational-energy estimates were obtained on an Apple MacBook Air M4, not on the Frontier supercomputer. Their accuracy depends on CodeCarbon's estimation method, the experimental hardware, background processes, and the emissions assumptions used. The warning about concurrent tracker instances introduces additional uncertainty, so the reported energy and CO_2e values require verification through repeated measurements with isolated trackers.

The evaluation also relies on a single chronological test period. Performance may differ under other workloads or operating conditions, and temporal dependence in the observations may affect the interpretation of the statistical results. Finally, the exported dataset did not contain valid measurement series for Overall Coolant Flow, Overall Waste Heat, and PUE. These variables could not therefore be included in the analysis, limiting the range of facility-level relationships that could be investigated.

11. RECOMMENDATIONS

Future studies of data-centre power prediction should consider computational cost alongside predictive accuracy. In this study, Linear Regression achieved the lowest MAE in both feature configurations and the lowest reported energy estimates, making it a suitable candidate for further evaluation. The cooling/workload-focused feature set should be prioritised in future experiments because it excludes Frontier Total Power, a variable that includes the target and may introduce target leakage.

The energy measurements should be repeated after ensuring that CodeCarbon trackers run independently. An external power meter could provide an additional basis for validating the estimates. Model performance should also be tested across multiple chronological periods to determine whether the observed results hold under different operating conditions.

Further research should distinguish forecasting accuracy from actual sustainability benefits. Claims about reductions in cooling energy, water use, carbon emissions, or total facility energy should be made only after controlled operational testing demonstrates measurable improvements.

12. CONCLUSION

This study evaluated Linear Regression, Random Forest, and XGBoost for forecasting the next recorded Frontier Facility Accessory Power value and compared their estimated training and inference energy. Linear

Regression achieved the best observed MAE, RMSE, and R^2 in both feature configurations. In the cooling/workload-focused scenario, it recorded an MAE of 0.025657 MW, an RMSE of 0.038213 MW, and an R^2 of 0.730887.

However, the bootstrap confidence interval for the focused Linear Regression–XGBoost MAE difference included zero, so the magnitude of that difference remains uncertain. The reported energy estimates are also provisional because CodeCarbon warned that multiple trackers might have been running concurrently and require further verification.

These findings show the value of evaluating predictive performance alongside computational cost when selecting a model. However, the study examines facility-power forecasting rather than direct cooling-system optimisation. Further validation is needed before drawing conclusions about potential reductions in cooling energy, water use, carbon emissions, or total facility energy.

13. FUTURE SCOPE

Future research should repeat the energy measurements after verifying that CodeCarbon trackers operate independently. Model performance should also be evaluated across multiple chronological test periods and under consistent experimental settings. Additional models could be included to examine whether other approaches offer a better balance between predictive accuracy and computational cost.

Further evaluation should measure deployment latency and hardware-specific energy consumption under different seasonal and operational conditions. A validated simulator or controlled operational trial could help determine whether facility-power predictions are useful for optimising data-centre operations.

Future studies should also investigate the relationship between prediction quality and measurable environmental outcomes. Claims about reductions in cooling energy, water consumption, or carbon emissions should be supported by reliable measurements or appropriately sourced data, including relevant carbon-intensity information.

ACKNOWLEDGEMENT

The authors thank the faculty and the department for guidance and feedback during the development and evaluation of this work.

REFERENCES

[1] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," Communications of the ACM, vol. 63, no. 12, pp. 54–63, 2020.

[2] P. Henderson, J. Hu, J. Romoff, E. Brunskill, D. Jurafsky, and J. Pineau, "Towards the systematic reporting of the energy and carbon footprints of machine learning," Journal of Machine Learning Research, vol. 21, no. 248, pp. 1–43, 2020.

[3] L. Lannelongue, J. Grealey, and M. Inouye, "Green Algorithms: Quantifying the carbon footprint of computation," Advanced Science, vol. 8, no. 12, Art. no. 2100707, 2021.

[4] L. Lannelongue, J. Grealey, A. Bateman, and M. Inouye, "Ten simple rules to make your computing more environmentally sustainable," PLoS Computational Biology, vol. 17, no. 9, Art. no. e1009324, 2021, doi: 10.1371/journal.pcbi.1009324.

[5] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for deep learning in NLP," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 3645–3650.

[6] L. F. W. Anthony, B. Kanding, and R. Selvan, "Carbontracker: Tracking and predicting the carbon footprint of training deep learning models," in Proceedings of the ICML Workshop on Challenges in Deploying and Monitoring Machine Learning Systems, 2020.

[7] R. Desislavov, F. Martínez-Plumed, and J. Hernández-Orallo, "Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning," Sustainable Computing: Informatics and Systems, vol. 38, Art. no. 100857, 2023.

[8] X. Chen, R. Tu, and X. Yang, "Parameter prediction optimization of data center's heat dissipation system using machine learning algorithms," Applied Thermal Engineering, vol. 232, Art. no. 121047, 2023, doi: 10.1016/j.applthermaleng.2023.121047.

[9] M. H. B. Mahbod, C. B. Chng, P. S. Lee, and C.-K. Chui, "Energy saving evaluation of an energy efficient data

center using a model-free reinforcement learning approach," Applied Energy, vol. 322, Art. no. 119392, 2022, doi: 10.1016/j.apenergy.2022.119392.

[10] X. Li, X. Wang, Z. He, X. Chen, and Z. Li, "Combining physical laws and ANN for predicting energy consumption of data center cooling systems," Energy and Buildings, vol. 311, Art. no. 114170, 2024, doi: 10.1016/j.enbuild.2024.114170.

[11] J. Sun, Z. Gao, D. Grant, K. Nawaz, P. Wang, C.-M. Yang, P. Boudreaux, S. Kowalski, and S. Huff, "Energy dataset of Frontier supercomputer for waste heat recovery," Scientific Data, vol. 11, Art. no. 1077, 2024, doi: 10.1038/s41597-024-03913-w.

[12] S. Ma, Y. Liu, Y. Liu, J. Wang, Q. Fang, and Y. Huang, "Artificial intelligence-enabled predictive energy saving planning of liquid cooling system for data centers," Advanced Engineering Informatics, vol. 65, Art. no. 103283, 2025, doi: 10.1016/j.aei.2025.103283.

[13] M. Smith, L. Zhao, J. Cordova, X.-F. Jiang, and M. Ebrahimi, "Machine learning-based energy-efficient workload management for data centers," in Proceedings of the IEEE 21st Consumer Communications & Networking Conference (CCNC), 2024, pp. 799–802, doi: 10.1109/CCNC51664.2024.10454842.

[14] S. Jadhav and Z. Liu, "Machine learning guided cooling system optimization for data centers," arXiv preprint arXiv:2601.02275, 2026. [Online]. Available: https://arxiv.org/abs/2601.02275

[15] S. Bhushan, S. Shanker, and K. Cengiz, "The Industrial Internet of Things: Advanced Analytic Framework," in Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities, IGI Global, 2023, ch. 21, doi: 10.4018/978-1-6684-6408-3.ch021.

[16] S. Shanker and S. Bhushan, "The Effect of Technological Advancement on Building a Sustainable City," Studies in Indian Place Names, vol. 40, no. 35, pp. 935–940, Feb. 2020.