

RESEARCH ARTICLE

AI-Driven Energy Optimization in Cloud and Edge Computing Using Ensemble Machine Learning Techniques (HGBR and XG Boost Models)

Abdulaziz Saleh Yeslem Bin-Habtoor^{1, 2,*}, Md Shakir³

ABSTRACT: The combination of cloud computing and artificial intelligence (AI) has completely transformed large-scale energy management manufacturing and computing operations. This study investigates the optimization of energy consumption in cloud and edge computing environments through advanced machine learning techniques. Using a comprehensive dataset that includes CPU utilization, ambient temperature, network load, and energy consumption metrics, we evaluated two key AI algorithms: Histogram-based for XG Boost Regression and HGBR energy prediction and optimization. Our analysis of 50 operational events shows a strong correlation with energy consumption (0.89), while network load shows a moderate correlation (0.52). The HGBR model scored 0.83 on The RMSE values are 0.69 on the test data and 0.69 on the training data 2.81 kWh and 2.64 kWh, respectively. XG Boost demonstrated exceptional training performance ($R^2 = 1.00$) but showed signs of overfitting, reaching an R^2 of 0.81 on the test data. The study confirms that AI-driven predictive analytics can effectively model the complex relationships between computational workloads, environmental conditions, and energy consumption. These findings support the development of intelligent energy optimization strategies that can reduce operational costs, reduce environmental impact, and improve sustainability in modern data centres and edge computing infrastructures.

Keywords: Artificial Intelligence, Energy Efficiency, Cloud Computing, Machine Learning, Green Data Centres, Predictive Analytics.

Received: 17 March 2025; Revised: 25 May 2025; Accepted: 13 July 2025; Published Online: 03 September 2025

1. INTRODUCTION

Energy management in large-scale operations has been transformed by the adoption of industrial cloud computing and artificial intelligence (AI). With the help of these technologies, factories can now monitor, analyze, and

optimize energy consumption in innovative ways that increase productivity, save costs, and support sustainability objectives [1, 2]. As a combination of Artificial intelligence (AI) and internet-based technology will significantly transform human existence by providing more intelligent services [3]. For processing purposes, this massive amount of data is currently sent across the network to a significant number of cloud data centers. Several QoS metric constraints are introduced by processing in CDCs, including latency and availability, network bandwidth, privacy, and, last but not least, energy efficiency [4]. No longer do you have to spend effort deploying and configuring local hardware. Businesses are using cloud-fog computing has expanded dramatically, which means that cloud providers have to invest in a large number of hardware resources. As a result, Cloud providers'

¹ Computer Network & Communication, Collage of Engineering, Hadramout University, Hadramout, Yemen.

² Department of Computer Science, Applied College, Sharurah, Najran University, Saudi Arabia

³ Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal-462033, Madhya Pradesh, India

* Author to whom correspondence should be addressed:
drazizhabtoor@gmail.com (Abdulaziz S. Y. Bin-Habtoor)

costs are further increased by their large energy footprint. Because of this, it is desirable to reduce the energy consumption of these resources [5].

One of the most important issues in cloud computing environments is energy efficiency, as large-scale data centers consume significant amounts of electricity, leading to increased operational costs and environmental impact. Efficient resource allocation and workload scheduling play a key role in optimizing energy consumption [6]. It aims to optimize and improve energy usage using cloud networking tools based on artificial intelligence. Energy forecasting uses supervised and unsupervised learning approaches to estimate power consumption trends [7]. Artificial Intelligence (AI) enhances cloud-edge orchestration by improving task scheduling, resource allocation, and data processing across distributed infrastructure. AI-based predictive analytics enable advanced decision-making, ensuring efficient resource utilization and system reliability [8].

In the cybersecurity industry, real-time threat detection and response from AI-powered systems greatly improves the security posture of cloud environments. In addition, AI enables new cloud-based services, such as predictive analytics platforms, computer vision services, and APIs for natural language processing, which democratize access to cutting-edge AI capabilities for organizations of all sizes [9]. AI refers to systems that simulate human intelligence to perform problem-solving and decision-making tasks, while ML allows Systems improve over time by learning from data. Together, these technologies bring a new level of intelligence and automation to cloud optimization [10]. The proposed edge AI processing pipeline involves establishing a high-level abstraction of vehicle edge service components [11]. New generation data centers are now developing in the direction of cloud computing. A cutting-edge area of energy saving research is green cloud data centers.

So far, the energy consumption optimization technique for cloud computing has been applied to various new technology research [12]. Cloud platforms must balance security, scalability, and energy efficiency, while serving a diverse set of tenants. An investigation to analyze the future potential of integrating quantum computers into cloud systems proposed that hybrid cloud infrastructures could contribute to overcoming the computational bottlenecks in conventional cloud systems [13, 14]. Edge computing mitigates these issues by executing computations on or near the devices that generate data, creating an “intelligent edge” that processes latency-critical inference tasks locally while using the cloud for heavy aggregation and model refinement [15]. Artificial intelligence has become a powerful tool optimizing energy consumption in cloud computing environments. AI-powered techniques such as machine learning algorithms and predictive analytics help data centers improve energy efficiency by dynamically adjusting power usage according to current demand [16].

AI-driven optimization in cloud computing encompasses a wide range of techniques designed to improve resource utilization, reduce latency, and improve overall service quality [17]. AI-driven orchestration further

implements self-healing mechanisms, reducing downtime and improving system reliability. This paper explores how AI can improve cloud scalability, security and performance [18]. Cloud data analytics AI programs operate at multiple levels, each of which makes a unique contribution to the overall performance of the system. Localized processing, which uses AI-driven technologies to acquire and pre-process data, is a component of the smart device stack [19]. Edge AI and Cloud AI represent two distinct but complementary approaches to processing artificial intelligence models. Understanding their roles, differences, and advantages is critical for designing efficient AI-powered applications, especially in IoT environments where real-time processing and large-scale data analysis must coexist [20]. Mobile cloud systems powered by deep learning and reinforcement learning technologies can achieve energy-aware operations that reduce computational pressure on mobile devices while extending operational life.

The research work here aims to seamlessly integrate edge-AI with network virtualization to improve energy efficiency, resource allocation and task offloading in mobile cloud computing. While AI optimists and pessimists seem to be divided among researchers and intellectuals. Considering the future that awaits us with the development of AI, the wisest course of action for humans is to develop a closer relationship with AI.

2. MATERIALS AND METHODS

2.1. Materials

CPU Utilization: The percentage of processing power is measured by CPU utilization actively used on servers or computer nodes during a given period of time. High utilization often indicates heavy workloads, requiring more energy for compute and cooling systems to maintain performance. Monitoring CPU utilization is essential Scale workloads, reduce operating costs, and optimize resource allocation. AI-powered solutions can predict trends, balance workloads, and prevent bottlenecks, making this metric a key parameter for cloud and edge computing performance analysis.

Ambient Temperature (°C): Ambient Temperature (°C) represents the surrounding air temperature in a data center or edge environment. Even a small increase in temperature can significantly impact cooling requirements, energy consumption, and equipment lifespan. Energy-efficient cloud operations often integrate predictive analytics to dynamically adjust cooling strategies based on environmental conditions. By monitoring this parameter, engineers ensure thermal stability, maintain safe operating limits, and reduce costs. Combining temperature data with computational metrics helps build models for better, greener, and more reliable data center management.

Network Load Gbps: Network Load Gbps represents the

amount of total data traffic flowing through the network infrastructure, measured in gigabits per second. This statistic is used by AI and machine learning models to predict the future bandwidth needs, optimize routing, and dynamically manage energy demand. Monitoring network load allows organizations to design scalable, cost-effective solutions that reduce power consumption while maintaining high performance in cloud and edge systems.

Energy kWh: Energy kWh measures the total electrical energy consumed by servers, cooling systems, and networking equipment during a specific measurement period. This output metric is critical for assessing operational efficiency, carbon footprint, and sustainability in modern computing environments. AI-powered energy prediction models help organizations reduce consumption, identify inefficiencies, and optimize workloads. By analyzing the correlations between CPU utilization, ambient temperature, and network load, engineers can implement predictive energy-saving measures, ensuring cloud operations are greener, smarter, and more cost-effective.

2.2. Optimization techniques

Hist Gradient Boosting Regression: A sophisticated machine learning technique is Hist Gradient Boosting Regression algorithm designed for high-performance gradient boosting, especially suited for large datasets. It is a histogram-based gradient boosting implementation that splits a series of features into separate bins (histograms), allowing for faster computation and reduced memory usage. Instead of evaluating all possible split points for a series of features, it evaluates a limited number of representative bins, significantly speeding up the process while maintaining accuracy. This technique builds additive regression models in a forward-looking, state-by-state fashion, optimizing the loss function using gradient descent. It builds Successive decision trees; each subsequent tree aims to correct errors from the previous one. Hist gradient boosting is especially helpful when working with large datasets many features and is known for its scalability and performance. It is implemented in popular libraries such as Light GBM and sci kit-learn, and is often used in predictive modeling tasks such as sales forecasting, credit scoring, and anomaly detection. Hist also supports regularization techniques such as gradient boosting regression, summarization, and sub-sampling, which help prevent overfitting and improve model generalization. Its ability to effectively handle missing values and categorical features further strengthens its applicability in real-world situations. This algorithm is widely accepted for its speed, scalability, and competitive prediction performance.

XG Boost Regression: A type of supervised learning method called XG Boost Regression, a series of numerical values. It uses a gradient boosting framework, where each tree corrects the errors of its predecessors, building a series of decision trees. XG Boost is known for its accuracy, efficiency, and

ability to handle large and complex datasets.

3. RESULTS AND DISCUSSION

This dataset provides valuable insight into energy efficiency in cloud and by looking at edge computing systems the combined impact of compute load, environmental conditions, and network traffic. CPU Utilization varies widely from very low values around 0.10 to high workloads exceeding 0.90, showing a clear correlation with Energy kWh. Systems operating at high utilization levels often require increased energy, not only to handle processing demands, but also to support cooling and power regulation. For example, utilization inputs above 0.90 correspond to energy consumption that consistently approaches or exceeds 30 kWh, reinforcing the role of CPU load as the primary driver of power demand. Ambient Temperature C ranges from a minimum of about 13°C to 33°C, capturing a wide range of environmental conditions.

High ambient temperatures contribute significantly to energy usage, as the cooling infrastructure works harder to maintain safe operating limits. For example, energy consumption increases by more than 39 kWh when the temperature is above 23°C and the usage is high. Conversely, cooler temperatures provide some energy relief, but the relationship is nuanced: higher workloads and higher network activity can mask the temperature savings, demonstrating the importance of multi-factor modeling for energy optimization. Network Load Gbps varies dramatically, with some instances below 1 Gbps and others approaching 10 Gbps, reflecting varying bandwidth requirements. High network loads increase switching, routing, and storage energy requirements, often adding several kilowatt-hours of consumption in high-traffic situations. Combined with CPU load, this relationship becomes more pronounced: higher CPU usage with higher network traffic yields peak consumption values. Energy kWh reflects the complex interplay between computational intensity, environmental conditions, and network performance. The dataset emphasizes that energy optimization cannot rely solely on a single-variable analysis. Instead, AI and ML models can use these variables to predict patterns, dynamically allocate resources, and improve performance. This dataset provides a realistic foundation for building consistent, intelligent strategies in cloud operations.

The descriptive statistics in [Table 1](#) provide a clear overview of the operational patterns that affect energy consumption in cloud and edge computing environments. CPU utilization has an average of approximately 0.55, indicating that systems are typically operating at slightly more than half of their total processing capacity. However, the wide range from a low of 0.11 to a high of 0.95 reveals that workloads fluctuate significantly, with some periods experiencing very low activity while others are approaching full capacity. This variability highlights the need for dynamic resource allocation strategies to maintain energy efficiency.

Table 1. Descriptive Statistics.

	CPU Utilization	Ambient Temperature (°C)	Network Load (Gbps)	Energy (kWh)
Count	50	50	50	50
Mean	0.547336	25.2882	4.76506	23.652754
Std	0.257707	4.631001	2.643034	6.664546
Min	0.1068	13.41	0.09	12.4734
25%	0.3136	22.8475	2.749	18.8439
50%	0.5063	25.415	4.8335	21.8734
75%	0.790675	28.9725	6.3945	28.75355
Max	0.945	33.7	9.971	39.0226

The ambient temperature averaged 25.3 °C, with measurements ranging from 13.4 °C to 33.7 °C, representing data collected from different operating environments or seasonal variations. Higher temperatures can increase cooling requirements, and the upper quarter temperature of nearly 29 °C indicates that many systems are often operating near the limits where thermal management becomes critical for performance and energy savings. The network load, measured in gigabits per second, averaged 4.77 Gbps, but varied widely from a low of 0.09 Gbps to nearly 10 Gbps. This variation reflects dynamic data center and edge workloads, where high network activity contributes significantly to overall energy demand. The Energy kWh column illustrates this combined impact, with an average of 23.65 kWh and a maximum of 39.02 kWh of energy consumption per monitor. The intermediate range (18.84–28.75 kWh) shows a significant operational energy spread driven by combined CPU load, temperature, and network activity. These figures emphasize that cloud energy usage is influenced by a wide range of, and interacting factors. Understanding these distributions is essential for developing AI-driven optimization strategies to reduce consumption, improve cooling efficiency, and effectively balance workloads.

Figure 1 presents a scatter plot matrix illustrating the relationships between CPU utilization, ambient temperature, network load, and energy consumption in cloud and edge computing environments. The plots reveal that CPU utilization shows a strong positive relationship with Energy kWh; as CPU utilization approaches its upper limit, energy consumption increases sharply, underscoring computing as the primary energy driver. Similarly, network load is positively related to energy consumption, especially at high bandwidth levels, highlighting the additional energy demand from data movement and communication infrastructure. Although ambient temperature shows a less direct relationship with energy consumption, its influence becomes clear when combined with other variables. Higher temperatures are likely to increase cooling requirements, which indirectly contribute to total energy consumption. The diagonal plots show that all variables are widely distributed, reflecting the dynamic and heterogeneous workloads in real-world cloud and edge environments. The scatter plots further

emphasize the complexity of energy optimization, where many factors interact rather than act independently. For example, instances of high CPU usage combined with high network load correspond to peak energy consumption. This visualization supports the use of AI and ML for predictive modeling, enabling predictive scaling, load balancing, and adaptive cooling strategies, while minimizing energy consumption while maintaining system performance.

Figure 2 shows a correlation heat map that illustrates the relationships between process parameters and energy consumption in cloud and edge computing operations. There is a strong correlation between CPU Utilization and Energy kWh (0.89), confirming that CPU load is the primary driver of energy consumption. This indicates that energy strategies should prioritize workload scheduling and compute performance. Network Load Gbps also shows a moderate positive correlation with energy consumption (0.52), indicating that data transfer and bandwidth demand contribute significantly to total energy requirements, especially in distributed edge-cloud systems. Together, high CPU utilization and high network load amplify energy consumption, consistent with the trends seen in the scatter plot. In contrast, Ambient Temperature C demonstrates a very low correlation with energy consumption (-0.005), although temperature indirectly affects cooling energy requirements in real operations. The weak relationships between temperature and other variables indicate that their effects may be nonlinear or context-dependent, reinforcing the need for advanced modeling. This heat map provides a concise quantitative view of the key energy drivers. It shows that compute and network factors dominate energy usage, making them ideal targets for AI-driven optimization. Insights from this correlation analysis can guide predictive algorithms for workload balancing, traffic shaping, and power conservation policies in cloud and edge environments.

3.1. Hist Gradient Boosting Regression

Figure 3 shows the performance of the histogram-based gradient boosting regression (HGBR) model in predicting Energy kWh using the training data. The scatter plot compares the predicted energy values with the actual

measurements, with the dashed diagonal line indicating the correct prediction. Most of the data points cluster near this line, demonstrating that the model effectively captures the underlying patterns linking CPU usage, ambient temperature, and network load to energy consumption. The predictions align particularly well in the medium energy range (15–30 kWh), indicating strong model generalization to typical operational scenarios. In the higher energy range (above 30 kWh), there is a slight underestimation, indicating that

additional tuning or feature engineering may be required to improve prediction accuracy for extreme workloads and heavy network demand scenarios. However, the absence of significant outliers and the relatively tight spread around the diagonal line reflect the robustness of the model. This result demonstrates the relevance of advanced machine learning techniques such as HGBR for modeling energy consumption in cloud and edge computing environments.

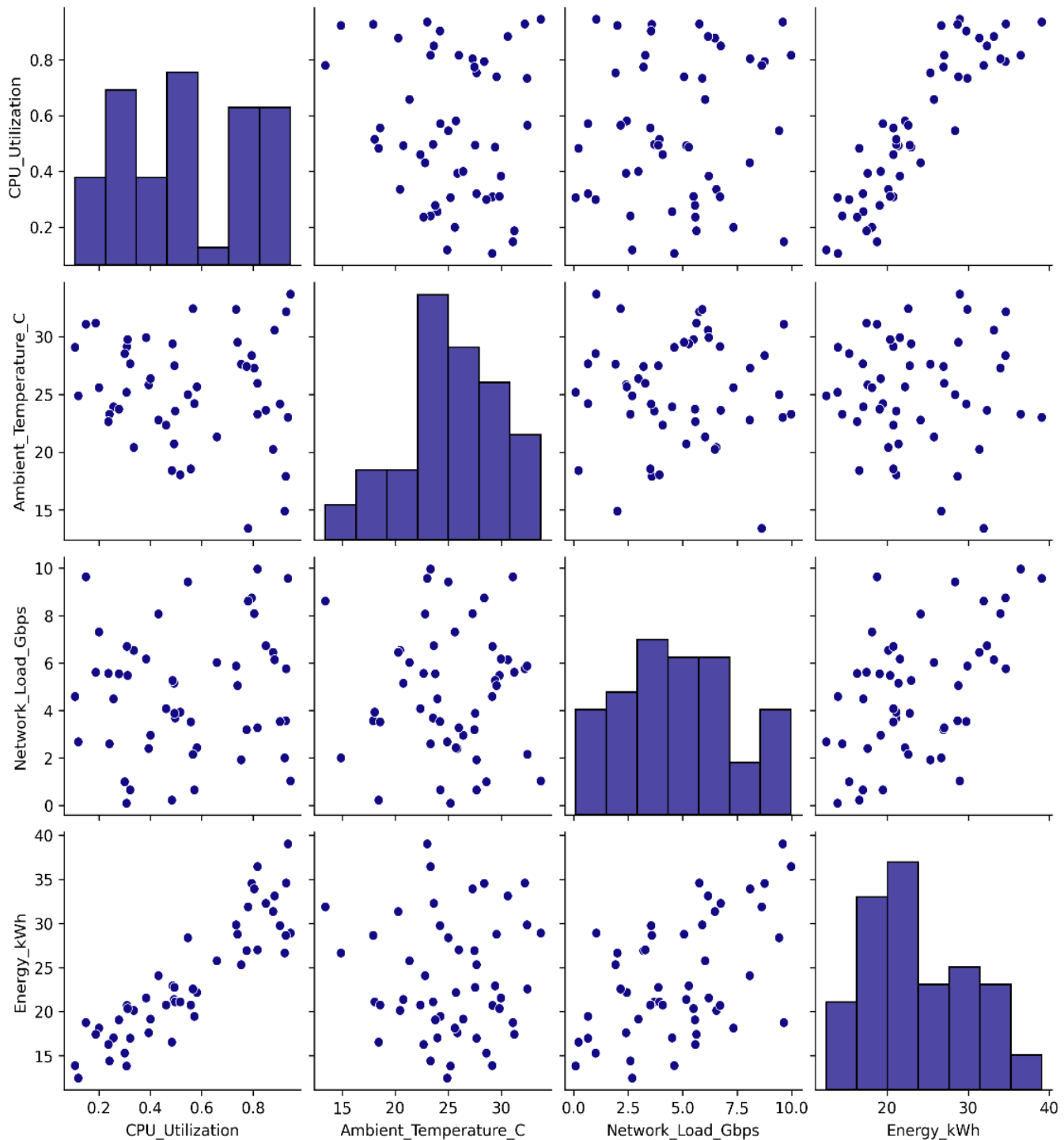


Fig. 1. Scatter Plot of Various AI for Energy Efficient Cloud Operations AI and ML at the Edge and in the Cloud.

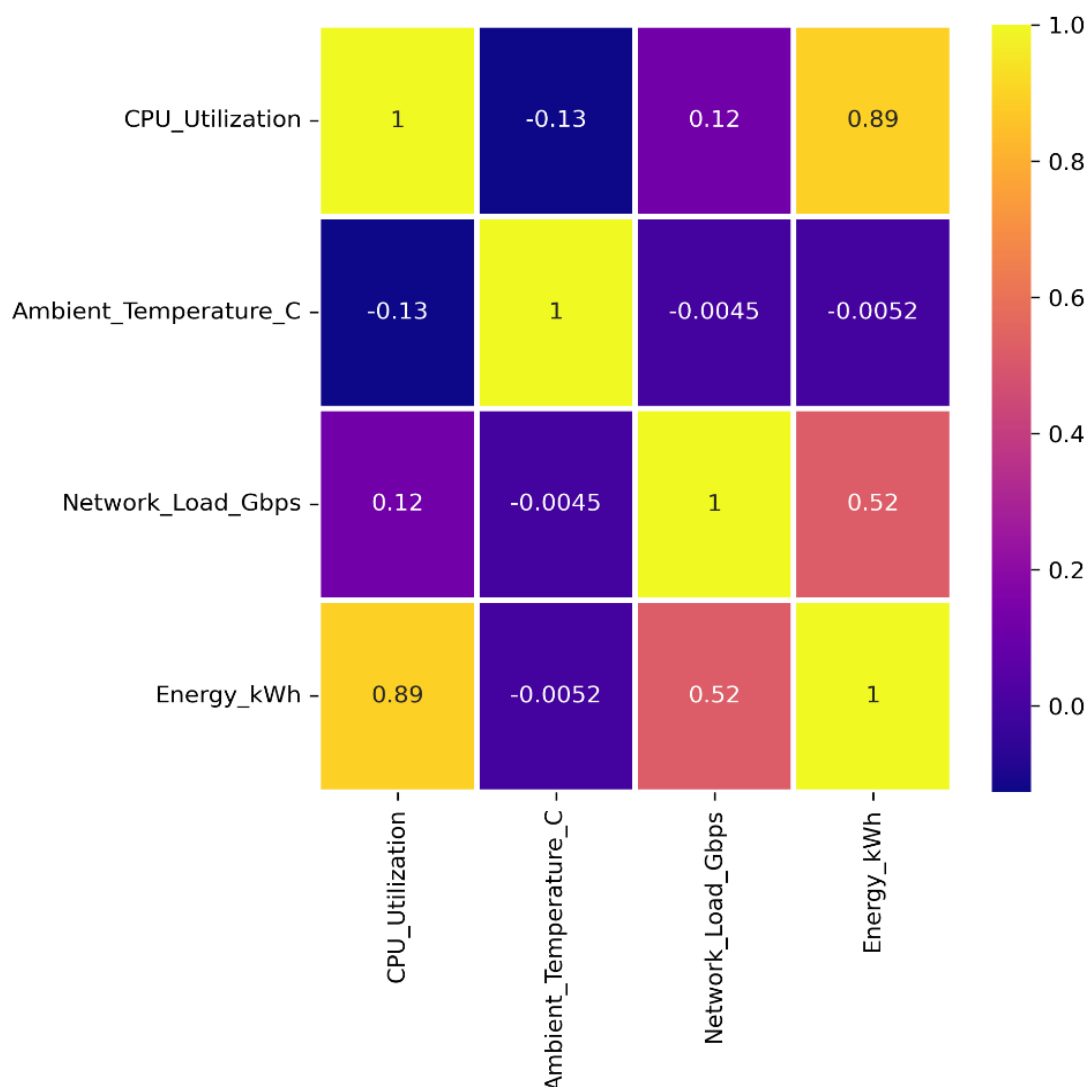


Fig. 2. Correlation heat map between the process parameters and the responses.

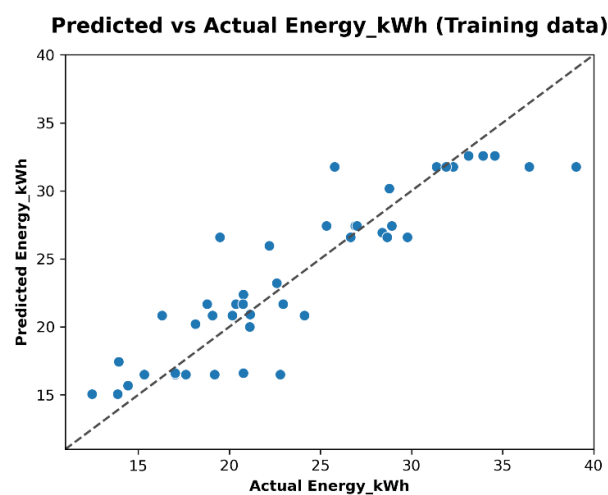


Fig. 3. Hist Gradient Boosting Regression Energy kWh (Training Data)

By learning nonlinear relationships and interactions between input parameters, the model provides actionable predictions that can support proactive energy optimization, workload planning, and cost reduction strategies. These findings highlight the value of AI-driven analytics for sustainable and efficient cloud operations.

Figure 4 illustrates the performance of the histogram-based gradient boosting regression (HGBR) model in predicting Energy kWh on the test dataset. The scatter plot compares the actual energy consumption values with the predicted ones, with the dashed diagonal line indicating correct predictions. Despite the fewer data points in the test set, the predictions generally follow the trend of the actual values, showing that the model is able to capture the key relationships between CPU usage, network load, and environmental factors, even in the unseen data. Most of the predictions are relatively close to the diagonal line, although there is some visible underestimation for medium to high energy values. This behavior indicates that while the model

generalizes well, it may struggle a bit with edge cases or extreme operational situations. The small number of points in this test set makes the estimation sensitive to the sample distribution, highlighting the need for a larger test dataset for more robust estimation. This figure confirms that the HGBR model is a strong candidate for energy prediction in cloud and edge environments, showing consistency between training and test performance. With additional fine-tuning, more training data, and potential feature engineering, predictive accuracy for high-energy scenarios can be improved to support better operational decision-making and energy optimization strategies.

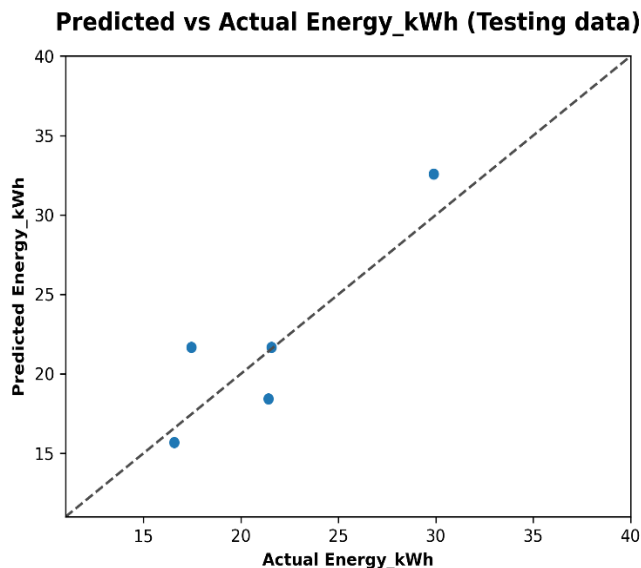


Fig. 4. Hist Gradient Boosting Regression Energy kWh (Testing Data)

Table 2 summarizes the performance metrics of the histogram-based gradient boosting regression (HGBR) model for predicting energy kWh on the training and testing datasets. The R^2 score for training is 0.83 and for testing is 0.69, indicating that the model explains a high proportion of the variance in energy consumption, with a slight decrease in the missing data – a typical indication of good generalization. The explained variance score (EVS) reflects these results, confirming consistent predictive performance. The mean square error (MSE) values are 7.92 (training) and 6.97 (test), which roughly translate to root mean square errors (RMSE) of 2.81 and 2.64, respectively. These low error values indicate that the predictions are close to the actual measurements. The mean absolute error (MAE) for both datasets is around 2.1 kWh, reinforcing the accuracy and reliability in estimating energy consumption. The Max Error metric shows a maximum deviation of 7.26 kWh in training and 4.22 kWh in testing, indicating occasional under-prediction or over-prediction in extreme situations. Both the mean square log error (MSLE) and the mean absolute error (MedAE) are low, further confirming the robustness. These metrics demonstrate that HGBR is a powerful model for

capturing linear dependencies between CPU usage, network load, temperature, and energy usage. Although the performance on unseen data is slightly lower, the results are robust enough for practical applications, and further improvements can be achieved through hyper parameter tuning or larger datasets.

3.2. XG Boost Regression

Figure 5 illustrates the performance of the XG Boost regression model in predicting Energy kWh using the training dataset. The scatter plot compares the predicted energy values with the actual energy measurements, with the dashed diagonal line indicating perfect prediction accuracy. The data points lie almost exactly on this line, demonstrating exceptional model fit and near-perfect accuracy in capturing the relationships between CPU usage, network load, ambient temperature, and energy consumption. This strong alignment indicates that the XG Boost model effectively learns complex nonlinear relationships within the dataset. Unlike simple regression techniques, XG Boost uses gradient boosting and decision tree ensembles to model subtle biases and interaction effects, resulting in highly accurate predictions. The minimum spread around the diagonal line indicates very low residual errors, meaning that the predictions are nearly indistinguishable from the observed values. Such strong performance on the training data highlights the predictive power of the model, while also raising concerns about potential overfitting. The model's ability to generalize to unseen test data will determine whether its high accuracy is due to true learning or memorization. Nevertheless, XG Boost's performance here reflects its suitability for energy forecasting tasks, making it a valuable tool for developing AI-driven strategies to optimize power usage, improve data center stability, and improve workload scheduling in edge-cloud computing environments.

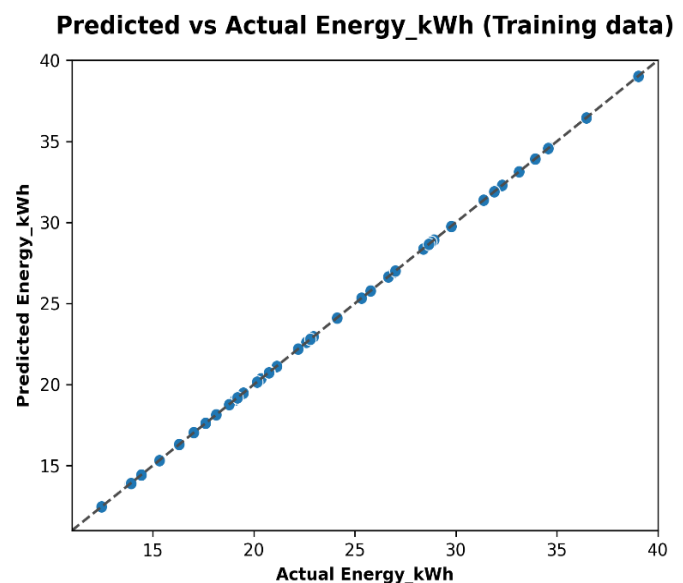


Fig. 5. XG Boost Regression Energy kWh (Training data).

Table 2. Hist Gradient Boosting Regression Energy kWh (Training, Testing Data).

Data	Symbol	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
Train	HGBR	0.82506	0.82506	7.91744	2.81380	2.12459	7.26390	0.01410	1.45321
Test	HGBR	0.68584	0.70404	6.96513	2.63915	2.18312	4.22430	0.01452	2.70955

Table 3. XG Boost Regression Energy kWh (Training Data and Testing Data).

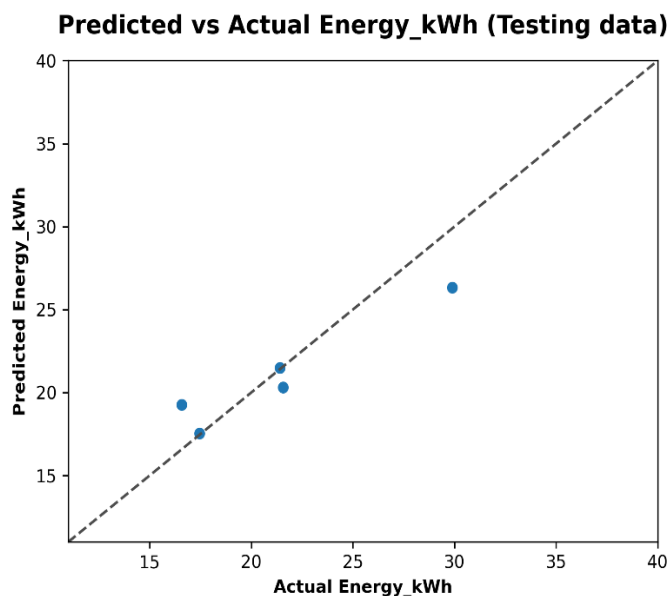
Data	Symbol	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
Train	XGBR	1.00000	1.00000	0.00000	0.00105	0.00079	0.00315	0.00000	0.00067
Test	XGBR	0.80680	0.81337	4.28328	2.06961	1.53421	3.54703	0.00770	1.24275

Figure 6 presents the performance of the XG Boost regression model on the test dataset to predict Energy kWh. The scatter plot compares the predicted energy values with the actual observations, with the dashed diagonal line indicating the correct prediction. Unlike the training data (Figure 5), where the fit was nearly perfect, the experimental results show some deviations, reflecting the challenge of generalizing to unobserved data. Most of the predictions are relatively close to the diagonal line, indicating that the model captures the overall energy trends affected by CPU usage, network load, and ambient temperature. However, there is a tendency to underestimate, especially at higher actual energy values, such as above 25 kWh. This indicates that while XG Boost effectively learns patterns on the training set, it may need more representative data or fine-tuned hyper parameters to better capture extreme or complex energy demand scenarios in the test. Despite these differences, the experimental results still demonstrate reasonable predictive accuracy, especially given the small number of test points available. With additional data, cross-validation, and possible feature engineering, the generalizability of the model can be significantly improved.

Table 3 summarizes the performance of the XG Boost regression (XGBR) model for predicting Energy kWh on both the training and test datasets. The training metrics indicate a nearly perfect model fit, with an R^2 score of 1.000 and zero mean square error (MSE), root mean square error (RMSE), and mean square log error (MSLE). The mean absolute error (MAE) and mean absolute error (MedAE) are very low (both below 0.001 kWh), indicating that the predictions match the actual values almost perfectly in training. While impressive, this level of accuracy strongly suggests overfitting, where the model memorizes patterns in the training data rather than generalizing effectively. On the test dataset, XGBR performs even better, with an R^2 score of 0.81 and an explained variance score (EVS) of 0.81, indicating strong predictive power in unobserved data. The RMSE of 2.07 kWh and MAE of 1.53 kWh indicate reasonable accuracy, although the performance is slightly weaker than the training results. The Max Error of 3.55 kWh indicates that the predictions are close to the actual energy values, even for challenging cases.

4. CONCLUSION

This study highlights the considerable potential of machine learning and artificial intelligence approaches to improve energy use in cloud and edge computing systems. Through a detailed analysis of operational parameters including CPU utilization, ambient temperature, and network load, we successfully established predictive models that can accurately predict energy consumption patterns. CPU utilization acts as the primary driver of energy consumption, showing a strong correlation coefficient of 0.89, while network load contributes moderately to the overall energy requirements. The implementation of histogram-based gradient boosting regression and XG Boost algorithms provided valuable insights into the complex relationships governing energy efficiency in distributed computing systems. The HGBR model showed strong performance with balanced training and testing results, achieving R^2 scores of 0.83 and 0.69, respectively, indicating good generalization capabilities without significant overfitting. Although XG Boost showed exceptional training performance, its tendency

**Fig. 6.** XG Boost Regression Energy kWh (Testing data).

towards overfitting highlights the importance of careful model selection and validation in practical applications. These results highlight how important AI-driven optimization is in overcoming problems growing energy needs of modern data centers and edge computing infrastructures. The implications of this research extend beyond just energy forecasting to include proactive resource allocation, dynamic workload planning, and adaptive cooling strategies. By leveraging predictive analytics, organizations can implement intelligent energy management systems that respond to changing operational conditions in real time, thereby reducing costs and environmental impact. Integrating AI technologies with cloud computing platforms, in conjunction with global sustainability goals and carbon reduction efforts, represents a paradigm shift toward sustainable and efficient operations. Future research should focus on expanding datasets to cover a variety of operational scenarios, incorporating additional environmental factors, and developing sophisticated ensemble methods that can better handle extreme operational conditions. The continued evolution of AI technologies, coupled with the growing importance of green computing, positions energy-efficient cloud operations as a critical area for continued investigation and development. This work contributes to the foundation for the next generation of intelligent systems that can balance performance, sustainability, and cost-effectiveness in an increasingly digital world.

DECLARATIONS

Ethical Approval

We affirm that this manuscript is an original work, has not been previously published, and is not currently under consideration for publication in any other journal or conference proceedings. All authors have reviewed and approved the manuscript, and the order of authorship has been mutually agreed upon.

Funding

Not applicable

Availability of data and material

All of the data obtained or analyzed during this study is included in the report that was submitted.

Conflicts of Interest

The authors declare that they have no financial or personal interests that could have influenced the research and findings presented in this paper. The authors alone are responsible for the content and writing of this article.

Authors' contributions

All authors contributed equally in the preparation of this manuscript.

REFERENCE

- [1] Egbuhuzor, N.S., Ajayi, A.J., Akhigbe, E.E., & Agbede, O.O., **2024**. Leveraging AI and cloud solutions for energy efficiency in large-scale manufacturing. *International Journal of Science and Research Archive*, 13(2), pp.4170–4192.
<https://doi.org/10.30574/ijrsra.2024.13.2.2314>
- [2] Zhu, S., Ota, K. & Dong, M., **2022**. Energy-efficient artificial intelligence of things with intelligent edge. *IEEE Internet of Things Journal*, 9(10), pp.7525–7532.
<https://doi.org/10.1109/JIOT.2021.3137489>
- [3] Yosuf, B.A., Mohamed, S.H., Alenazi, M.M., El-Gorashi, T.E.H. & Elmirghani, J.M.H., **2021**. Energy-efficient AI over a virtualized cloud fog network. *In Proceedings of the Twelfth ACM International Conference on Future Energy Systems*, pp.328–334.
<https://doi.org/10.1145/3447555.3465399>
- [4] Iftikhar, S., Ahmad, M.M.M., Tuli, S., Chowdhury, D., Xu, M., Gill, S.S. & Uhlig, S., **2023**. HunterPlus: AI based energy-efficient task scheduling for cloud–fog computing environments. *Internet of Things*, 21, p.100667.
<https://doi.org/10.1016/j.iot.2022.100667>
- [5] Kumar, A., Tiwari, G. & Malviya, A., **2025**. AI-Driven Ensemble Learning for Optimized Cloud Performance and Energy Efficiency. *SSRN*.
<https://doi.org/10.2139/ssrn.5366691>
- [6] Mohammed, Z., Mohammed, N.U., Mohammed, A., Gunda, S.K.R. & Ansari, M.A.A., **2025**. AI-Powered Energy Efficient and Sustainable Cloud Networking. *Journal of Cognitive Computing and Cybernetic Innovations*, 1(1), pp.31–36.
<https://doi.org/10.21276/jccci/2025.v1.i1.6>
- [7] Ramamoorthi, V., **2023**. Exploring AI-Driven Cloud-Edge Orchestration for IoT Applications. *International Journal of Science Research in Computer Science, Engineering and Information Technology*, 9, pp.385–393.
<https://doi.org/10.32628/CSEIT239072>
- [8] Anbalagan, K., **2024**. AI in cloud computing: Enhancing services and performance. *International Journal of Computer Engineering and Technology*, 15(4), pp.622–635.
<https://doi.org/10.5281/zenodo.13353681>

- [9] Somalraju, S., **2025**. AI-Driven Cloud Optimization: Leveraging Machine Learning to Enhance Cloud Performance, *International Journal of Scientific Research in Computer Science Engineering and Information Technology* 11(2), pp. 2119-2135, <https://doi.org/10.32628/CSEIT23112583>
- [10] Mungoli, N., **2023**. Scalable, distributed AI frameworks: leveraging cloud computing for enhanced deep learning performance and efficiency. arXiv preprint arXiv:2304.13738. <https://doi.org/10.48550/arXiv.2304.13738>
- [11] Katare, D., Perino, D., Nurmi, J., Warnier, M., Janssen, M. & Ding, A.Y., **2023**. A survey on approximate edge AI for energy efficient autonomous driving services. *IEEE Communications Surveys & Tutorials*, 25(4), pp.2714–2754. <https://doi.org/10.1109/COMST.2023.3300669>
- [12] Yang, J., Xiao, W., Jiang, C., Hossain, M.S., Muhammad, G. & Amin, S.U., **2018**. AI-powered green cloud and data center. *IEEE Access*, 7, pp.4195–4203. <https://doi.org/10.1109/ACCESS.2018.2887124>
- [13] Rajesh, G., Amu, D., Adithya, B., Jayanna, S.P., Sangeetha, B. & Janakiraman, R., **2024**. AI-Based Secure and Energy-Efficient Framework for Multi-Tenant Cloud Systems. *Journal of Information Systems Engineering and Management*, 10(30s) DOI: <https://doi.org/10.52783/jisem.v10i30s.4830>
- [14] Jonnakuti, S., **2022**. Intelligent Edge Architectures: AI at the Boundary of Cloud and Device. *International Journal of Engineering Technology Research & Management*, 6(8), pp.95–99. <https://doi.org/10.5281/ZENODO.15455791>
- [15] Oloruntoba, O., **2024**. Green cloud computing: AI for sustainable database management. *World Journal of Advanced Research and Reviews*, 23(03), pp.3242–3257. <https://doi.org/10.30574/wjarr.2024.23.3.2611>
- [16] Kofi, J.E., **2025**, December. Monitoring Cloud Performance Metrics Utilizing AI to Estimate the Efficiency of Cloud Operations. In *2025 7th International Symposium on Advanced Electrical and Communication Technologies (ISAECT)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ISAECT68904.2025.11318827>
- [17] Patil, G. B. **2022**. AI-Driven Cloud Services: Enhancing Efficiency and Scalability in Modern Enterprises. *International Journal of Intelligent Systems and Applications in Engineering*, 10(2), 303. <https://ijisae.org/index.php/IJISAE/article/view/6728>
- [18] Ramamoorthi, V., **2023**. Applications of AI in Cloud Computing: Transforming Industries and Future Opportunities. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 9(4), pp.472–483. <https://doi.org/10.32628/CSEIT2390910>
- [19] Hasan, R.A., Yasin, K., Kareem, P.R., Salih, A.M., Jasim, H., Ameerudeen, M.A., Abbas, M.A. & Rachini, A., **2025**. Energy-Efficient Task Offloading and Resource Allocation in Mobile Cloud Computing Using Edge-AI and Network Virtualization. *Khwarizmia*, pp.42–49. <https://doi.org/10.70470/KHWARIZMIA/2025/005>
- [20] Rausch, T. & Dustdar, S., **2019**. Edge intelligence: The convergence of humans, things, and AI. In *2019 IEEE International Conference on Cloud Engineering (IC2E)*, pp.86–96. <https://doi.org/10.1109/IC2E.2019.00022>