International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0138
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 899 - 904

AI-Driven Energy Prediction Models for Sustainable Cloud Infrastructure: A Comprehensive Survey and Unified Framework

Kavyanjali S[1], Arundhati Prasanthan[2] ,Dr. Gurunath R[3]
[1,2]PG-Master of Computer Application, Dayananda Sagar College of Arts, Science and Commerce, Bangalore, Karnataka
[3]Associate Professor, Department of Computer Application, Dayananda Sagar College of Arts, Science and Commerce, Bangalore, Karnataka
Emails: *skavyanjali645@gmail.com[1], arundhatinair125@gmail.com[2], gurunath@dayanandasagar.edu[3]*

Abstract
The Rapid rise of AI and cloud services are increasing in the modern data centers, which consume a large amount of energy consumption. There are certain key concerns like Sustainability, operational cost and environmental impact. However, many AI-based approach had been implemented to maintain workload prediction, Resource optimization and renewable energy integration in cloud infrastructure. In existing studies, there is a lack of unified sustainability-oriented framework. This research does a Comprehensive survey which includes AI-based energy prediction model for unified sustainability, that includes techniques like machine learning, Deep learning and reinforcement learning that does predictions such as workload forecasting, intelligent scheduling, predictive maintenance and energy-aware resource management. while doing this study we found some problems in current cloud systems, which includes limited cross-layer coordination, lack of carbon-aware scheduling, data heterogeneity, explainability issues and minimal real-world validation. To overcome these issues we can use a unified multi-layer sustainable cloud framework, this is nothing but building one complete system that connects all layers of the cloud together that is data acquisition, AI-driven prediction, optimization and sustainability monitoring. A Modified and simpler mathematical model is introduced to reduce the maximum energy consumption along with satisfying Quality of Service (QoS), carbon intensity, and renewable usage constraints. This proposed models approach is to provide a better point of view for building intelligent, energy-efficient and environmentally applicable cloud infrastructures that ensures the sustainability as well as supporting the future research in green computing and sustainable digital ecosystems.

Keywords: *Carbon awareness, Cloud optimization, Energy prediction, Green computing, Sustainable cloud*

.

1. Introduction
The Exponential growth of cloud services and Artificial Intelligence applications has presented noticeable innovation but it also needs lot of computing power to work in cloud and edge environments. Examples like deep learning training, Data centers, energy consumption, AI, rising demand, energy efficiency, global electricity use, sustainability. Although efficiency has improved and infrastructure management. Overall electricity consumption increasing as people more dependence on digital services expands [1], [2]. As cloud platforms grows to facilitate AI-driven solutions, sustainability, operational cost and environmental impact are becoming equally important. Parameters like Power Usage Effectiveness (PUE) help evaluate how efficiently data centers use of energy and support efforts toward green computing operations [3]. Inorder to solve energy and efficiency problems, AI based research management is emerging as a more intelligent alternative to traditional cloud operations. Traditional systems react only after demand increases, which can lead to allocating extra resources and wasting energy. Instead of reacting, think if cloud systems could predict the demand in advance. Predictive models go through the historical workload patterns and predict the future demands, so resources can be allocated in advance and more

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0138
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 899 - 904

efficiently [4]. Machine learning and reinforcement learning techniques helps in combining virtual machines, scheduling tasks properly [4], [5].These smart methods can reduce energy waste without affecting system performance. Regarding these progress, research in this area is still scattered. Many studies focus at only one part such as virtualization efficiency or AI-based scheduling without integrating them into a broader plan for sustainability. This lack of connecting between predicting workloads, saving energy and checking sustainability moves it slow to use ideas in the real adoption. Identifying these difficulties, takes a holistic perspective on sustainable cloud computing. It checks out recent updates in AI-based energy prediction, sorting resource efficiency models, identifies key missing, and suggests a single way to fix it links future trends with better plaining, sustainability monitoring. The main aim is to help the building of intelligent, energy-efficient, and environmentally responsible cloud infrastructures.

2. Literature review

As stated in [10] AI-driven models are widely used by edge-cloud data centers to Enhance the resource utilization through machine learning, Deep learning, reinforcement learning , predictive modelling, workload forecasting and energy forecasting and fault prediction using Abnormality detection. This study in [11] shows that Deep Reinforcement Learning techniques like DNQ ,PPO for renewable-integrated microgrid energy prediction and showing improved version of renewable resources used to reduce operational cost the challenges are no real-world validation and scalability untested. According to study in [12] focuses on Predictive Analytics for Sustainable Cloud and Edge Infrastructure by integrating AI techniques and Optimization techniques and lack of standardized sustainability metrics. The research in [13] shows that Predictive maintenance of renewable energy using AI-driven method the objective of this study is IoT Integration, Analytical Techniques , cost and cybersecurity. The study in [14] paper shows that by using AI-driven predictive analytics and automation that can reduce carbon emissions and achieving sustainable building management. Demonstration in [15] shows that Predictive Scaling of techniques like ARIMA , LSTM and Reinforcement Learning used to reduce cost of infrastructure in cloud applications but it required High convergence time for RL. According to [16] this paper shows that AI-Driven Cloud Storage Optimization to reducing Sustainability ,better Efficiency and improve Security to run the data centers more effectively but the challenges are optimization and sustainability. The work demonstrated in [17] shows that deploying of AI in Cloud Architectures with supports green computing and workload management. The study in [18] disclose that using real-time AI and IoT Sensors can reduce energy consumption but long-term performance unknown and validation is required. Research in [19] tell that green cloud computing strategy and environment friendly cloud methodologies to reduce workload planning and energy-efficient strategies in cloud computing for efficient energy. The paper [20] aims that AI-driven predictive analytics for multi-cloud management Predictive Analytics Applications and Security. The paper [21] focuses on Semantic AI Infrastructure and sets up a conceptual and technical basis for aligning informatics infrastructures with the sustainable development agenda. As stated in [22] demonstrates that Artificial Intelligence in Renewable Energy and it uses Artificial Intelligence, Energy Optimization, Predictive Maintenance, Renewable Energy, Smart Grid, Machine Learning but the challenge is that it takes require larger datasets, stronger integration and practical large-scale testing. Research in [23] emphasizes on AI-powered predictive models in enhancing energy efficiency in cloud computing resource allocation. This chapter in [24] stresses on innovative model with the objectives are to develop an AI-Driven Framework for Sustainable Energy Management in Edge Cloud Computing.

3. Research gaps identified

Even though AI has significantly contributed to improving energy efficiency in cloud computing still many practical challenges exist . Issues such as the absence of unified frameworks, poor coordination between system layers, limited data availability, and scalability issues continue to affect the development of fully sustainable cloud infrastructures. Figure 1

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0138
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 899 - 904

illustrates the key research gaps identified in this area.

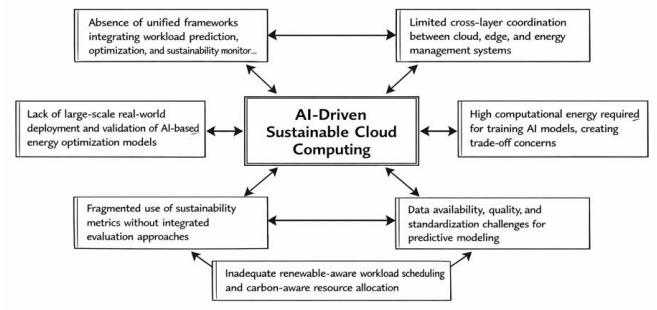

Figure 1. Gaps identified on AI-driven Sustainable Cloud Computing

4. Proposed research framework

To address the limitations identified in previous studies, this research introduces a conceptual multi-layer framework for sustainable cloud computing. The framework integrates data collection, intelligent analysis, optimization techniques, and sustainability monitoring within a single structure. Instead of treating workload management, energy efficiency, and environmental impact as separate problems, the proposed approach combines them into a unified system. The goal is to develop cloud infrastructures that are not only scalable and high-performing but also energy-efficient and environmentally sustainable.

4.1.Proposed Multi-Layer Sustainable Cloud Framework

The proposed model consist of five layers (as shown in the figure 2), each layer plays an important role in enhancing the energy efficiency and sustainability in cloud operations.

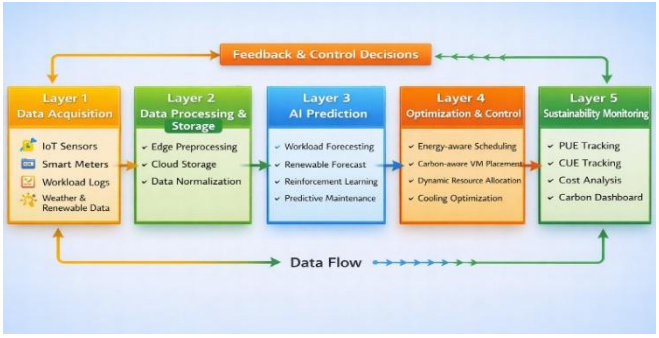

Figure 2. Proposed Multi-Layer Sustainable Cloud

4.2. Data Acquisition

the first stage of the framework mostly focuses on collecting information about how the cloud environment operates. Both real-time and past data are collected so that system can understand the useful patterns and environmental aspects.
Information is gathered from :

- IoT sensors placed within data centers
- Smart energy meters
- Workload and server utilization logs
- Weather conditions and renewable energy sources
- Collecting this data provides a clear view of system performance, energy consumption , and workload behaviour.

4.3. Data Processing and Storage Layer

The raw data will be useful only if it is processed properly then only it can be used for analysis. This layer handles tasks such as cleaning, organizing and storing the data efficiently . Some preprocessing is performed at the edge level to reduce delays and limit unnecessary data transfer. The processed data is then stored in cloud storage systems that can maintain a large amount of information. Normalizing and structuring the data ensures that it can be analyzed accurately and used effectively for further processing.

4.4. AI-Based Prediction Layer

The prediction layer forms the intelligent component of the framework. Artificial intelligence techniques are applied to identify patterns in the collected data and estimate future system conditions such as:

- Workload forecasting using models like LSTM or Prophet
- Renewable energy prediction using regression or ANN
- Reinforcement learning for smart scheduling
- Predictive maintenance models to prevent system failures

These predictions allow the cloud infrastructure to plan resources proactively.

4.5.Optimization and Control Layer

The motive of this layer is to convert predictions into operational decisions. According to the insights gained from the AI models ,the system performs the actions which improves efficiency and reduce energy

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0138
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 899 - 904

consumption. Examples include scheduling tasks based on energy availability, placing virtual machines in locations with lower carbon impact, dynamically adjusting computing resources, and optimizing cooling mechanisms in data centers . Through these actions, the system can significantly enhance both performance and Environment stability.

4.6. Sustainability Monitoring Layer

This is the final layer which ensures transparency and continuous improvement. Various metrics are monitored to understand the environmental and operational impact of the infrastructure these include the following:

- Power Usage Effectiveness (PUE) for measuring energy efficiency
- Carbon Usage Effectiveness (CUE) for estimating carbon emissions
- Cost analysis of operations
- Carbon footprint dashboards for decision support
- Continuous monitoring helps organizations track progress and identify areas where improvements can be made.

Overall, the proposed framework provides a structured approach for improving the sustainability of cloud computing environments. By integrating data collection, intelligent prediction, optimization, and monitoring, the system enables proactive energy management.

5. Methodology

Table 1 Summary Table of AI-Based Energy Optimization Studies

Study Type	AI Technique	Application	Energy Saving	Limitations
VM Consolidation [25]	ML / Heuristics	Resource allocation	Reduces idle power	May affect performance
Workload Prediction [26]	LSTM / Regression	Demand forecasting	Avoids over-provisioning	Needs historical data
Reinforcement Learning [27]	DRL	Scheduling	Better energy–performance balance	High training cost
Green Data Centers [28]	AI Optimization	Renewable integration	Lowers carbon footprint	Renewable variability
Edge–Cloud Optimization [29]	Hybrid ML models	Distributed workloads	Saves transmission energy	Integration complexity

6. Mathematical model for energy-aware sustainable cloud optimization

To achieve energy-efficient and environmentally responsible cloud operations, the overall energy usage of the cloud infrastructure can be formulated as an optimization problem.

6.1.Objective Function

The primary objective of the model is to minimize the total energy consumption of the cloud system.
Total Energy Consumption:

$$Energy_{total} = E_{Compute} + E_{Cooling} + E_{network}$$

where:

$E_{compute}$: Energy consumed by servers or VMs for computation

$E_{cooling}$: Energy used by cooling systems in data centers

$E_{network}$: Energy consumed in data transmission and communication

6.2. Further Research

- Distributed learning for green cloud: It allows federated AI training an model on many devices to reduce the energy consumption in the central server.
- Digital replicas of data centers: Using this data centers to monitor, cooling the system, energy prediction.
- Explainable AI in energy prediction: Using this model we can understand the AI based systems to they are make decisions and predicting the energy usage, scheduling and optimization.
- Green AI model training: Developing an AI model that utilize less carbon emissions, energy and computing power.
- Integration with SDGS: We use sustainable development goals that support environment sustainabilty which reduce organizations resources and environmental impact.

Conclusion

The research paper demonstrates that how AI- based and predictive techniques are used in cloud computing for sustainable infrastructure. this also

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0138
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 899 - 904

highlights the machine learning, Deep learning and reinforcement learning that does predictions such as workload forecasting, intelligent scheduling, predictive maintenance and energy-aware resource management. where existing methodologies or paper do not have Fully sustainable solutions for carbon emissions and real time monitoring. To overcome the address the problem this paper provides a five layered architecture which includes Data Acquisition, Data Processing and Storage, AI-Based Prediction Layer, Optimization and Control and Sustainability Monitoring. It also established ha basic mathematical model which improves energy-efficient and maintaining performance. This work conclude by future enhancement we have Sustainable cloud, Energy prediction, Green computing, Cloud optimization, Carbon awareness and green AI training models to reduce energy usage.

References

[1]. E. Masanet *et al.*, "Recalibrating global data center energy-use estimates," *Science*, 2020.
[2]. A. Shehabi *et al.*, "United States data center energy usage report," Lawrence Berkeley National Laboratory, 2016.
[3]. The Green Grid, "PUE: A comprehensive examination of the metric," 2012.
[4]. A. Beloglazov and R. Buyya, "Energy efficient dynamic consolidation of virtual machines," *Concurrency and Computation*, 2012.
[5]. H. Mao *et al.*, "Resource management with deep reinforcement learning," *ACM HotNets*, 2016.
[6]. A. Beloglazov and R. Buyya, "Energy efficient resource management in cloud computing," *Future Generation Computer Systems*, vol. 28, no. 5, pp. 755–768, 2012.
[7]. A. Shehabi *et al.*, "United States data center energy usage report," Lawrence Berkeley National Laboratory, 2016.
[8]. Google Inc., "Data center efficiency measurements (PUE, CUE, WUE)," White Paper, 2020.
[9]. E. Masanet *et al.*, "Recalibrating global data center energy use estimates," *Science*, vol. 367, no. 6481, pp. 984–986, 2020.
[10]. AI-Driven Predictive Analytics for Optimizing Resource Utilization in Edge-Cloud Data Centers .
[11]. AI-Driven Energy Optimization in Renewable-Integrated Microgrid Infrastructure for Sustainable Smart Communities.
[12]. AI-Driven Predictive Analytics for Sustainable Cloud and Edge InfrastructurePredictive maintenance of renewable energy infrastructure using AI: A comprehensive review
[13]. AI-driven decarbonization of buildings: Leveraging predictive analytics and automation for sustainable energy management AI-Driven Predictive Scaling for Performance Optimization in Cloud-Native Architectures
[14]. AI-Driven Cloud Storage Optimization: Enhancing Efficiency, Security, and Sustainability
[15]. AI for Sustainable Cloud Architectures Smart Energy Optimization in Government Infrastructure using AI-Driven Predictive Models and IoT Sensors
[16]. Green Computing: Energy-Efficient AI for Sustainable Cloud Infrastructure
[17]. AI-driven predictive analytics for multi-cloud management Semantic AI Infrastructure for Sustainable Decision Intelligence
[18]. Artificial intelligence in renewable energy: A review of predictive maintenance and energy optimization Energy-Efficient Cloud Resource Allocation Through AI Powered Predictive Models.
[19]. AI-based sustainable green energy optimization for edge cloud computing with renewable energy resources
[20]. A. Beloglazov and R. Buyya, "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines," Concurrency and Computation: Practice and Experience, vol. 24, no. 13, pp. 1397–1420, 2012.
[21]. Q. Xu, Z. Wang, and M. Chen, "Online

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0138
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 899 - 904

learning for workload prediction in cloud computing," IEEE Transactions on Cloud Computing, 2018.

[22]. H. Mao, M. Alizadeh, I. Menache, and S. Kandula, "Resource management with deep reinforcement learning," in Proc. ACM HotNets, 2016.

[23]. M. Ghamkhari and H. Mohsenian-Rad, "Energy and performance management of green data centers: A profit maximization approach," IEEE Transactions on Smart Grid, vol. 4, no. 4, pp. 1898–1910, 2013.

[24]. W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637–646, 2016.