IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, April 2026
Impact Factor: 8.2

Deep Learning-Based Architecture for Energy-Efficient Green Cloud Computing with Intelligent Resource Management

Jayshree Pasalkar, Tanvi Sonune, Hrishikesh Wadile, Simoni Raghatate
Assistant Professor, Department of Information Technology
Third Year Students, Department of Information Technology
AISSMS Institute of Information Technology, Pune, India
jayshree.pasalkar@aissmsioit.org, tanvisonune3@gmail.com
hrishikeshwadile2510@gmail.com, simoniraghatate11@gmail.com

Abstract: *Cloud computing services, as of today, being one of the fastest growing sectors, is also associated with great concerns such as high energy consumption and carbon emission, Service Level Agreements (SLA) violations and poor third- party resource utilization in large data centres. Thus, for this, Green Cloud Computing (GCC) has developed as a solution to overcome these challenges in large scale with energy aware resources generating. Smart workload scheduling, and developing sustainable infrastructure In this paper we present an AI based deep learning driven architecture that dynamically optimises resources allocations and workload management in the green cloud environments. The proposed framework employs predictive deep learning models and technologies to analyse historical workload patterns, network traffic, and resource utilization for minimization of energy consumptions and dynamic allocation of computing resources. Since architecture has a predictive nature of forecasting the demand for workloads and status of the system in advance, it becomes quite helpful for cloud providers, as this can lead to automatically creating virtual machines (VM), automated intelligent workload scheduling, managing idleness hence making buffer management efficient. In this paper we present an AI based deep learning driven architecture that dynamically optimises resources allocations and workload man- agement in the green cloud environments. The proposed frame- work employs predictive deep learning models and technologies to analyse historical workload patterns, network traffic, and resource utilization for minimization of energy consumptions and dynamic allocation of computing resources. Since architecture has a predictive nature of forecasting the demand for workloads and status of the system in advance, it becomes quite helpful for cloud providers, as this can lead to automatically creating vir- tual machines (VM), automated intelligent workload scheduling, managing idleness hence making buffer management efficient. The proposed effective architecture is proven for simulation and evaluation of cloud which indicates that it can enhance the resource utilization resulting in lower energy consumptions as well as fewer processing devices buffers against heuristic logic-based optimization methods and techniques used widely by clouds*

Keywords: Green Cloud Computing, Deep Learning, En- ergy Efficiency, Resource Optimization, Virtual Machine Migration, Buffer Optimization, Workload Prediction, Sustainable Cloud Infrastructures

I. INTRODUCTION

In the modern world, the backbone of the infrastructure of the digital world is cloud computing, which provides on-demand access for the computation of resources like storage, processing, networking, and software services.[4] Currently, the infrastructure of the modern digital world is largely dominated by cloud computing, which acts as the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

backbone, providing on-demand access to the resources of the cloud com- puting infrastructure. Worldwide organizations are showing in- creased adoption of cloud platforms depending on supporting applications such as big data analysis, scientific computing, artificial intelligence, e-commerce, etc.[3] As a result of this, the scalability of cloud computing is accelerating the digital revolution in the world; as a result of the rapid development of cloud computing, data centres are developed on a large scale, but as a result of this, electrical energy is consumed in a huge amount, thereby contributing to the emission of carbon gases into the atmosphere. As a result of this, a new research challenge has emerged, i.e., enhancing sustainability in cloud computing.[11] These data centres run 24/7 to ensure the availability of the system and the reliability of the services being provided. Thousands of physical servers and numerous virtual machines are present in these infrastructures.[10] These systems consume a lot of power to perform the computations, manage the network activities, cool the servers, etc. This is where the problem of the rising power requirements is becoming a cost factor for the cloud service providers as well as a problem from the environmental aspect.[8] Here, the idea of Green Cloud Computing (GCC) was introduced to reduce the power consumption and the carbon footprint of the cloud services while providing the required Quality of Service (QoS). One of the main challenges in the efficient management of resources in a green cloud environment is the efficient management of resources. The conventional methods used in the scheduling and allocation of resources in a cloud environment often make use of a heuristic-based method.[14] This method may not work well in a dynamic environment. Inefficient allocation may cause the servers to become idle, may increase the energy consumption rate, may increase the buffer delay, and may cause the virtual machines to migrate. Recent advancements in artificial intelligence and deep learn- ing provide promising solutions to these challenges.[?] Deep learning models can analyse historical workload data, network traffic patterns, and resource utilization metrics to predict future workload and system behaviour. This predictive ability can help provide better resource provisioning, VM allocation, and workload scheduling. Predicting workload changes can help cloud systems reduce idle resource usage, server energy consumption, and excessive buffer delays due to network congestion and inefficient scheduling. Moreover, incorporat- ing deep learning with green cloud computing architectures can provide better buffer management and load balancing techniques. Deep learning models can dynamically manage buffers and allocate resources according to predicted workload intensity. This can reduce latency and provide more efficient system operation while maintaining energy consumption and meeting Service Level Agreement (SLA) compliance. Thus, the focus of the research is to develop a deep learning architecture for Green Cloud Computing, which will help improve the efficiency of the environment and reduce the delays in the processing buffers. This will help develop a more efficient, adaptive, and sustainable cloud environment using the benefits of machine learning decisions and cloud resource allocation. This will help fulfil the needs and requirements of the modern computing environment.

II. RELATED WORK

The recent research work carried out under the domain of Green Cloud Computing mainly emphasizes the improvement of energy efficiency in cloud data centres, along with Ser- vice Level Agreement (SLA) compliance and performance. Researchers have conducted a number of studies to evaluate the application of optimization-based resource management approaches to improve energy efficiency and resource uti- lization. Researchers have extensively investigated the appli- cation of energy-aware load balancing and virtual machine consolidation approaches to minimize the number of active servers, thereby reducing idle power consumption. [?], [?] In addition, a number of research works have been conducted to evaluate the application of intelligent resource allocation and scheduling approaches to improve the sustainability of cloud computing infrastructure. Researchers have employed predictive models and machine learning approaches to evaluate workload characteristics and allocate resources accordingly, thereby improving resource utilization and eliminating un- necessary virtual machine migration.[?] Moreover, efficient virtual machine migration approaches and priority-based ser- vice management approaches have been proposed to maintain

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

workload balance, thereby reducing migration overhead and unnecessary energy consumption. Workflow scheduling has also been extensively addressed for large-scale scientific and data-intensive computing applications.[?], [?] Techniques like dynamic task clustering and DAG-based workflow scheduling aim to minimize the overhead of workflow scheduling and maximize parallel execution in distributed cloud environments. Moreover, recent studies also show that there is a potential to integrate IoE with cloud computing to provide real-time mon- itoring and intelligent decision-making and adaptive energy management for interconnected clouds.[?] However, there are many challenges to be addressed to realize sustainable clouds. For example, most of the existing techniques and solutions are based on heuristic and simulation-based optimization tech- niques that are not real-time adaptive in nature. In addition, problems like network congestion, buffer-related delays, poor prediction of workload, and lack of intelligent learning models are also reported to impact energy efficiency and performance of clouds. This indicates that more adaptive and predictive models like deep learning-based models can be more effective in real-time optimization of resource allocation and buffers while maintaining energy efficiency in clouds.[?], [?]

- Energy-aware load balancing
- Virtual machine consolidation
- Predictive resource allocation
- Workflow scheduling optimization

However, many approaches rely on heuristic or simulation- based methods, lacking real-time adaptability.

III. LITERATURE REVIEW

Several research efforts have been made to improve energy efficiency and sustainability in cloud computing environments. Geetanjali and S. J. Quraishi (2022) addressed the issue of high energy consumption and environmental impact of traditional cloud data centres through a conceptual analysis and survey of green cloud strategies. Their work introduced the concept of green cloud computing and emphasized energy saving and sustainability using technologies such as cloud computing, load balancing, blockchain, and IoT. However, the study lacked experimental validation or quantitative evaluation.

C. Sailesh et al. (2023) reviewed the adoption of green cloud computing, focusing on energy usage and server load issues in cloud infrastructure. They analyzed adoption factors and cloud service models, highlighting VM load management and job allocation. Despite providing insights into SaaS, IaaS, and PaaS models, the study lacked an implementation framework and real-world case studies.

M. S. Raza et al. (2021) explored intelligent computa- tional techniques to address inefficient resource utilization and energy waste. Their comparative review analyzed AI-based techniques for improving energy efficiency, but the study had limited performance benchmarking.

S. Kaur and N. Chaurasia (2021) investigated fault toler- ance in green cloud computing, focusing on reducing energy consumption in fault-prone cloud networks. Their study linked reliability with green cloud principles but lacked simulation or quantitative energy metrics.

R. Sadhu et al. (2024) presented a conceptual review ad- dressing environmental damage caused by large-scale cloud usage. The study summarized key approaches, challenges, and opportunities using renewable energy and sustainable data centres, but did not propose a technical model.

N. Singla (2023) conducted an analytical study on green cloud infrastructure, addressing the increasing carbon footprint of cloud servers. The study detailed advantages and challenges of green cloud computing but lacked experimental validation.

R. Doss et al. (2022) proposed a micro smart grid-based approach to reduce CO2 emissions from ICT infrastructure through case study and empirical analysis. Although the study introduced smart grid-powered data centres, it had limited market demand analysis.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

T. Shree et al. (2020) discussed green computing approaches to address heat generation and inefficient energy usage in cloud computing. Their descriptive analysis explained eco-friendly frameworks but relied on outdated technologies with minimal evaluation.

A. Kaushik et al. (2022) conducted a survey on energy- efficient load balancing algorithms to reduce power wastage due to idle resources. While the study compared various heuristic algorithms, it lacked experimental comparison.

D. Chen (2022) applied bibliometric analysis to identify green building research trends using cloud and big data technologies. Although insightful, the study was not directly focused on cloud energy optimization.

E. H. Alharbi et al. (2020) proposed a framework for green cloud adoption in Saudi Arabia, addressing the lack of structured implementation. However, the framework was region-specific and not generalized.

R. Sharma et al. (2022) reviewed virtual machine migration techniques to address resource underutilization in data centres. Their work highlighted migration strategies for energy efficiency but lacked performance experiments.

A. Badhoutiya (2022) surveyed energy modelling tech- niques for eco-friendly cloud workstations, identifying future research challenges. However, the study did not provide a working prototype.

A. Singha et al. (2022) proposed architectural features for sustainable green cloud computing, addressing energy inefficiency in growing networks. The study remained conceptual without implementation.

R. Kumar and M. Ali Khan (2022) conducted empirical case studies on smart grid-based green cloud solutions for private cloud hosting. While advocating green SLAs and sustainability metrics, the dataset used was limited.

N. Suratia et al. (2023) performed a systematic literature review on CO2 emissions in cloud data centres, classifying challenges and research directions. However, the study lacked practical validation.

C. Lai et al. (2024) applied optimization modelling to reduce energy waste due to low CPU utilization. Their approach achieved significant improvements, including 20–23% host reduction and 42% CPU utilization increase, though implementation complexity was high.

R. Wakankar et al. (2024) used iFogSim simulation to address network congestion and energy usage in fog and cloud computing. Their approach achieved 98.7% latency reduction but requires real-world deployment.

C. L. Stergiou et al. (2020) proposed an energy-efficient allocation algorithm for big data processing using CloudSim-based simulation. While effective in simulation, the study lacked real-world validation.

IV. GREEN CLOUD COMPUTING

Green Cloud Computing emphasizes the development and delivery of cloud infrastructures that consume less energy and have a reduced impact on the environment. In the recent past, the growth of cloud data centres has been tremendous in supporting the increasing number of digital services. Therefore, the efficient management of computer resources is crucial in reducing the cost of operations and the environ- mental impact.[?] Optimization methods, including energy- efficient resource allocation, virtual machine consolidation, and workload management, play a crucial role in enhancing the efficiency of cloud infrastructures.[?], [?] Sustainability in cloud computing involves the integration of energy-efficient technologies, renewable energy sources, and intelligent man- agement of infrastructures to ensure the environment is not adversely impacted. In the modern era, the main focus is the achievement of a balance between different requirements, including energy efficiency, service level agreement, scalabil- ity, and stability. By using intelligent resource management techniques, it is possible to dynamically adjust the resources used in the cloud environment according to the needs of the workload. This helps in reducing the amount of idle time and enhancing the usage of the hardware. Recent developments in the field of artificial intelligence and machine learning have further helped to augment the concept of green cloud optimization. Intelligent mechanisms such as predictive anal- ysis of the workload and the provision of intelligent resource allocation can be integrated into the cloud platform to ensure the optimization of the virtual machines deployed on the cloud platform. By integrating optimization algorithms with green computing practices, the concept of Green Cloud Computing can be considered a potential way to develop scalable cloud infrastructures with greater emphasis on green computing.[?]

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, April 2026
Impact Factor: 8.2

V. PROPOSED ARCHITECTURE

The proposed architecture integrates deep learning models with cloud resource management to optimize energy efficiency, minimize buffer delays, and enhance resource utilization. The architecture is based on a number of interconnected layers that collect data, predict workload, and manage resources intelligently.

The proposed architecture utilizes deep learning together with cloud resource management in order to enhance energy efficiency, buffer delay reduction, as well as resource opti- mization in green cloud computing environments. The topmost

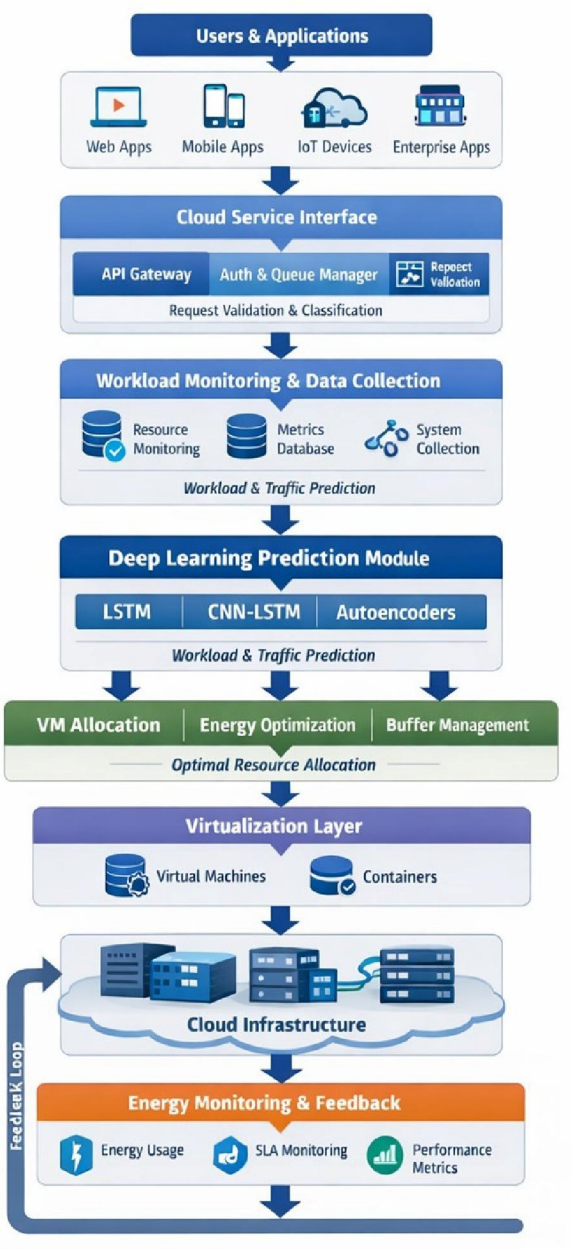

Fig. 1. Proposed Architecture of Green Cloud Computing System

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

layer involves users or applications who submit computing requests via the Cloud Service Interface. The Workload Moni- toring Layer involves continuous monitoring of various param- eters, including CPU, memory, network, virtual machine load, as well as energy consumption, from the cloud computing environment. The parameters are then recorded and sent to the Deep Learning Prediction Module, where predictions are made regarding the workload demands, resource bottlenecks, as well as network congestion. Based on the predictions made by the module, the Intelligent Resource Management Layer utilizes predictions regarding workload demands, resource bottlenecks, as well as network congestion, in order to perform dynamic re- source allocation, virtual machine allocation, as well as buffer optimization in order to enhance efficiency while minimizing energy consumption as well as SLA violations. The Virtualization Layer involves the execution of the decisions made by the Intelligent Resource Management Layer via virtual machine allocation, while the Cloud Infrastructure Layer involves the processing of workload requests via physical servers, storage devices, as well as network devices.

VI. METHODOLOGY

The proposed methodology is based on the integration of deep learning-based predictive mechanisms and intelligent resource management techniques to enhance energy efficiency while minimizing processing buffer delays in cloud data centres. The proposed framework involves various stages of operations, including workload monitoring, predictive analy- sis, resource optimization, and feedback adaptation. In the first step of the proposed framework, real-time system metrics are monitored by deploying various monitoring agents across the cloud environment. The agents are responsible for collecting various types of information, including CPU utilization, mem- ory utilization, network traffic, virtual machine load, server energy consumption, and buffer queue length. The monitored information is stored in a centralized dataset, which includes various patterns of workload history.[?] In the next step of the proposed framework, the monitored dataset is processed by using a deep learning-based predictive analysis model. The model analyses the workload history trends and real- time metrics to predict the resource demand in the near future while avoiding network congestion. Predictive analysis helps the system to allocate computing resources in advance while avoiding unnecessary virtual machine migrations or server overload situations.[?] Based on the workload predictions, the intelligent resource management module optimizes the resource allocation in the cloud environment. This module optimizes the allocation of virtual machines in the cloud environment, balances the workload of virtual machines, and consolidates virtual machines during server underutilization. Server consolidation helps in reducing the power consumption of the data centre by placing the idle servers at low power state.[?] Moreover, the system includes an adaptive buffer management mechanism, which optimizes the allocation of buffers based on the predicted network load. This mechanism optimizes the allocation of buffers, thereby reducing queue delays in the network. Finally, a monitoring and feedback mechanism continuously monitors the performance indicators of the system, such as energy consumption, SLA, and latency. The performance indicators are then used to improve the predictive model and optimize the allocation of resources. This process allows the system to adapt and be energy-efficient in the cloud.[?]

VII. EXPECTED OUTCOMES

The proposed architecture, which is based on deep learning technology, is also expected to have a significant impact on the sustainability of the cloud computing environment. By using the proposed architecture, it would be possible to dynamically allocate the required computing resources in real time. This would help reduce the overall energy consumption of the sys- tem, as the servers would be utilized optimally. The proposed framework is also expected to have a significant impact on the reduction of the number of unnecessary virtual machine migrations. By using the proposed framework, it would be possible to avoid the need for frequent virtual machine migra- tions. This would help reduce the network traffic, resulting in a stable system that also supports the required Service Level Agreement (SLA). The adaptive buffer management component of the architecture is also expected to alleviate delays in processing as well as network congestion within the cloud computing environment. This is because the architecture is able to predict the workload intensity accordingly, thereby minimizing delays in the queue as well as response time of the applications.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

In the future, the proposed architecture will be implemented in a cloud computing simulation environment, such as CloudSim or iFogSim, in order to test its effectiveness in real-world situations. Performance metrics of the proposed architecture, including energy consumption, number of VM migrations, SLA violation ratio, as well as latency, will be studied in order to determine the effectiveness of the proposed framework. Moreover, further research could be conducted in the development of more efficient deep learning techniques in order to enhance the effectiveness of the proposed framework in managing cloud computing resources.

VIII. CONCLUSION

The growing trend of cloud computing services has resulted in a significant increase in the energy requirements of large- scale data centres, which have raised concerns regarding their operational costs as well as environmental sustainability. Thus, Green Cloud Computing has emerged as a significant approach to achieve energy savings without compromising the performance and reliability of cloud computing services. In this paper, a deep learning-based approach was proposed to enhance the energy efficiency and resource utilization in cloud computing environments. The proposed approach was based on a framework that combines workload moni- toring, predictive deep learning, intelligent resource alloca- tion, and buffer adaptation to optimize the performance of cloud computing services. The proposed approach aims to minimize unnecessary virtual machine migration, idle servers, and optimize performance. Additionally, the architecture in- cludes a feedback-based optimization technique that constantly observes the performance of the system and improves the allocation of resources. Therefore, the proposed architecture would be able to offer a scalable and energy-efficient solution for the dynamic workloads within the contemporary cloud computing environment. In conclusion, the integration of the proposed architecture, which combines the concepts of deep learning and green cloud computing infrastructure, would be able to offer a viable approach towards the development of a green, intelligent, and high-performance cloud computing environment.

REFERENCES

[1] Y. A. Alsaaidah, A. Muhammed, M. A. Ala'anzy et al., "Empowering cloud providers: Optimised locust-inspired algorithm for SLA violation mitigation in green cloud computing," Computing, vol. 107, p. 174, 2025, doi: 10.1007/s00607-025-01527-7.
[2] G. Wang, B. Wen, J. He et al., "A new approach to reduce energy consumption in priority live migration of services based on green cloud computing," Cluster Computing, vol. 28, p. 207, 2025, doi: 10.1007/s10586-024-04695-x.
[3] Z. Ma, J. Chen, Y. Yuan, and T. Xu, "The Impact of the Internet of Everything on Green Cloud Computing," in Internet of Everything, Springer, 2025, doi: 10.1007/978-3-031-84426-3 1.
[4] A. Ahmad, R. A. Khan, S. U. Khan et al., "Green cloud computing adoption challenges and practices: A client's perspective-based empirical investigation," Cognition, Technology & Work, vol. 25, pp. 427–446, 2023.
[5] M. Yadav and A. Mishra, "Energy-efficient workflow scheduling using dynamic task clustering for sustainable cloud computing," Discover Computing, vol. 28, p. 201, 2025.
[6] C. Lai, L. Li, Y. Liang, and H. Zhang, "Application of operational research in green computing of cloud host," in Proc. MICCIS, 2024, pp. 104–110.
[7] R. Wakankar, S. Simhachalam, R. Ganesan, and T. Velmurugan, "Im- proved efficiency of fog and cloud computing paradigms," in Proc. AKGEC, 2024, pp. 1–6.
[8] R. Sadhu, H. Kaur, and V. Pattanaik, "Green cloud: Navigating to a sustainable future through green cloud computing," in Proc. ICEMPS, 2024, pp. 1–5.
[9] C. Sailesh et al., "A review on adoption of green cloud computing," in Proc. ICCMC, 2023, pp. 1–6.
[10] N. Singla, "Green cloud infrastructure: Mitigating the environmental impact of cloud computing," in Proc. AECE, 2023, pp. 144–147.
[11] N. Suratia et al., "An extensive analysis of green cloud computing: Overview, associated challenges and research directions," in Proc. ICIMIA, 2023, pp. 1147–1152.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

[12] Geetanjali and S. J. Quraishi, "Energy savings using green cloud computing," in Proc. ICICICT, 2022, pp. 1496–1500.
[13] R. Doss et al., "Efficient green cloud computing through micro smart grid," in Proc. ICACITE, 2022, pp. 336–340.
[14] A. Kaushik, G. Khan, and P. Singhal, "Cloud energy-efficient load balancing: A green cloud survey," in Proc. SMART, 2022, pp. 581–585.
[15] R. Sharma, A. Bala, and A. Singh, "Virtual machine migration for green cloud computing," in Proc. ICDCECE, 2022, pp. 1–7.
[16] A. Badhoutiya, "Green cloud computing—Next step towards eco- friendly workstations," in Proc. ICECA, 2022, pp. 809–813.
[17] A. Singha et al., "Green cloud computing—To build a sustainable tomorrow," in Proc. ICONAT, 2022, pp. 1–6.
[18] R. Kumar and M. Ali Khan, "Green cloud solutions using smart grid," in Proc. SMART, 2022, pp. 1165–1170.
[19] M. S. Raza, J. Wei, and M. M. Ali Muslam, "A succinct review of intelligent computational techniques in green cloud computing," in Proc. ICCSPA, 2021, pp. 1–4.
[20] S. Kaur and N. Chaurasia, "Improved green cloud computing with reduced fault in the network: A study," in Proc. ICSCCC, 2021, pp. 427–431