Journal of Global Research in Electronics and Communication
Volume 2, No. 06, June, 2026
Available Online at: www.jgrec.info
REVIEWPAPER

Review of Artificial Intelligence Based Cloud Cost and Security Optimization Techniques Systems

Mr. Mohit Sahu
Assistant Professor
Department of Computer Sciences and Applications
Mandsaur University, Mandsaur
mohit.sahu@meu.edu.in

***Abstract*—Cloud computing has gained immense importance as a critical technological infrastructure for present scalable, flexible, and cost-efficient computing services to businesses across the globe. But achieving efficient operations cost-wise along with providing high-level security continues to pose major challenges owing to the growing complexity of cloud environments. Artificial Intelligence (AI) has proven to be a potent tool that helps with intelligent resource management, prediction, automated scaling, and enhanced security. The current paper proposals an extensive review of AI-powered cloud cost and security optimization strategies. The paper reviews the main AI technologies, such as Machine Learning (ML), Deep Learning (DL), and Reinforcement Learning, and analyzes their application in resource prediction, intelligent workload scheduling, auto-scaling, energy-efficient resource management, threat detection, intrusion detection, anomaly detection, and access control. It has been found that AI-based strategies result in better utilization of resources, cost efficiency, scalability, and enhanced security of cloud systems against emerging cybersecurity threats. The paper highlights the limitations of AI technologies and the existing challenges, including lack of explain ability, data privacy, adversarial attacks, high computational cost, and scalability issues. The overall study has shown the capabilities that can be achieved by using artificial intelligence to develop intelligent, secure, and efficient cloud computing systems.**

Keywords—Artificial Intelligence (AI), Cloud Computing, Cost Optimization, Cloud Security, Machine Learning, Resource Allocation, Intrusion Detection Systems (IDS).

I. INTRODUCTION

The necessity for smart systems that can efficiently optimize costs while offering strong security and superior service delivery has increased due to rapid expansion of cloud-based computing capabilities. One of key technology for providing scalable and affordable computer resources is clouds [1]. As cloud adoption becomes increasingly popular, operational cost management and security issues have become crucial for companies. Existing cloud management strategies tend to focus on resource allocation and monitoring based on predefined rules, leading to resource waste, higher operational costs, and security risks [2]. Artificial Intelligence (AI) technologies are developing fast and new solutions to these problems become possible thanks to intelligent automation and data analysis [3].

The concept of Artificial Intelligence is being more widely applied in cloud computing technologies to increase resource efficiency, automation of workload management and cloud computing cost optimization. Machine learning, deep learning, and other AI-based approaches facilitate the ability of predicting resource needs, load balancing, and classification of workloads. The use of those features makes possible to minimize the risks of under-provisioning and over-provisioning of resources [4] Moreover, cloud management systems that use AI enable autonomous decision-making and real-time optimization, making cloud environments more adaptive and cost-effective.

Apart from being used for optimizing costs, artificial intelligence is another crucial element that has to be taken into account when implementing cloud security. Large volumes of data from networks, system logs, and user behavior may be analyzed by sophisticated artificial intelligence algorithms to find anomalies, possible threats, and weaknesses. Deep learning, machine learning, and several other artificial intelligence techniques for cyber have demonstrated a high degree of accuracy in identifying hostile activity [5].

The use of artificial intelligence in cloud computing has resulted in the emergence of self-healing and self-optimizing systems, which take care of cost, performance, and security needs simultaneously. Nevertheless, there are still many issues relating to data privacy, model interpretability, computational overhead, and security concerns in relation to AI models [3]. It is therefore necessary to conduct a comprehensive review of the cloud cost and security optimization approaches based on AI [6], Identify current problems and give directions for future research and applications.

In this paper, it review Artificial Intelligence-based cloud cost and security optimization methodologies, applications, benefits, and problems. It also identify current research problems, gaps in the literature, and future directions in the development of intelligent and secure cloud computing systems.

A. Structure of The Paper

This paper continues to be structured as follows. In section II, AI technology in cloud computing is explained. The sections III and IV describe cost and security optimization using AI technology, respectively. Section V identifies some of the challenges that currently exist, Section VI summarizes the literature, and Section VII concludes the paper.

II. ARTIFICIAL INTELLIGENCE IN CLOUD COMPUTING

AI is one of important facilitators in modern-day cloud computing technology by virtue of intelligent automation, predictive decision-making, optimization of resources, and improved security [7]. The use of AI with cloud computing makes it possible for companies to effectively manage computing resources, lower costs, increase efficiency, and

enhance security. This is because AI cloud computing is able to analyze cloud computing data in large amounts and come up with patterns, predictions, and even detect any anomaly [8]. As cloud environments become bigger and more complicated, it becomes essential for us to use AI technologies in order to create intelligent, adaptable, and economical cloud infrastructures.

- **Machine Learning (ML):** ML is an ability of the cloud system to learn from past and live data without any explicit coding. Algorithms of ML such as Decision Trees, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Naïve Bayes are extensively applied to predict workloads, allocate resources, estimate costs, detect anomalies, and monitor security. Through identifying the usage pattern of the cloud, ML increases the resource efficiency and minimizes the cost of operation.
- **Deep Learning (DL):** DL is an advanced branch of ML that analyses large amounts of cloud data and finds complicated patterns using numerous layers of neurons. DL models like CNNs, RNNs, LSTM, and GAN are employed for workload prediction, intrusion detection, malware classification, and threat analysis. It ensures high prediction accuracy and enable intelligent cloud management and optimization.
- **Reinforcement Learning (RL):** Through RL, it is possible to learn an optimal behavior in cloud computing through constant interactions with the environment. RL has been extensively used in job planning, automated scale, load balance, flexible resource distribution, and cloud efficiency. Through constant adaptation to varying workloads, RL facilitates optimization in cloud computing.

III. AI-Based Cloud Cost Optimization Techniques

Artificial Intelligence has become a vital technology to optimize the cost of clouds through intelligent resource management and workload processing [9]. AI-based cost optimization techniques allow businesses to cut down costs, use resources efficiently, and maintain quality services. Through analysis of cloud utilization patterns and infrastructure performance, AI makes it possible to operate clouds cost-effectively.

A. Predictive Resource Allocation

Predictive Resource Allocation applies AI to predict resource demands for allocating cloud resources in the future. To evaluate historical cloud data and project future tasks, methods such as ML, DL, and Predictive Analytics are employed [10][11]. Such an approach eliminates over-provisioning and under-provisioning, increases resource usage, and cuts down on costs without compromising on service quality and SLA compliance.

B. Intelligent Workload Scheduling

Optimization of Workload scheduling through Artificial Intelligence (AI) allows the intelligent scheduling and processing of workloads within cloud computing environments by applying various AI strategies. Workload scheduling through Machine Learning, Reinforcement Learning, and Optimization Strategies can be achieved effectively [12]. Such strategies increase system efficiency, improve resource usage, and minimize cloud operation costs.

C. Auto-Scaling and Elastic Resource Provisioning

Cloud resources are dynamically adjusted according to demand needs thanks to auto-scaling and elastic resource delivery. The AI method uses deep learning and machine learning approaches to forecast workload fluctuation and make proactively scale options [13]. This dynamic allocation of resources enhances scalability and efficiency while minimizing waste and cutting down operational costs.

D. Resource Utilization Optimization

Optimization of Resource Utilization makes use of AI techniques for management of computing, storage, and network resources in a cloud environment. It helps in minimizing the waste of resources by making efficient use of idle resources through artificial intelligence.

E. Energy-Efficient Resource Management

Energy-Efficient Resource Management makes use of AI technologies for efficient consumption of energy in the cloud data center environment by effective management of computational and storage resources. Machine learning is used to decrease energy consumption, enhance energy efficiency, and lower costs.

IV. AI-Based Cloud Security Optimization Techniques

The significance of Artificial Intelligence in securing cloud technology is evident in intelligent monitoring, threat analysis, and defense mechanisms that are put into place through the help of artificial intelligence. The techniques used for securing cloud technology are constantly analyzing cloud operations in order to able to detect any threats, respond to security incidents, and protect cloud assets from all types of cyber-attacks.

A. Threat Detection and Prevention

Threat detection and prevention are vital in protecting the cloud from cyber threats and attacks [14] . The AI-driven security systems constantly analyze network traffic, user activity, and system behavior to detect threats and security breaches. Such smart systems allow real-time threat detection and prevention, thus ensuring effective cloud security management [15].

B. Intrusion Detection Systems (IDS)

IDS are extensively employed in monitoring cloud infrastructure in an effort to identify malicious activities that could threaten the security of systems [16]. AI-enabled IDS solutions help offer adaptive and scalable protection by detecting abnormal activities and new types of attacks. This ensures improved detection results, reduced false alarms, and automation of response actions to ensure the security of cloud computing systems.

C. Anomaly Detection

Anomaly detection is an important security method based on artificial intelligence that helps detect activities different from usual cloud operation [17][18]. The analysis of cloud behavior and utilization of

resources by AI enable early detection of any unauthorized access or activities, thus ensuring improved cloud reliability, increased threat awareness, and better cybersecurity management.

D. Access Control and Authentication

Access and authentication methods based on AI technology can be used to enhance cloud security through analysis of user behavior, login patterns, and access rights for identification and detecting any attempt of accessing the cloud inappropriately. Intelligent authentications create adaptive security systems that improve identity and data management and minimize threats.

E. Malware Detection and Classification

The methods that are used to detect and classify malware and ransomware in cloud computing utilize artificial intelligence algorithms [19]. Using features of files, behavior patterns, and networks' activities, artificial intelligence can correctly identify malware and categorize it, allowing rapid actions and increasing the overall security and stability of cloud computing environments.

V. Challenges In AI- Based Cloud Cost And Security Optimization

Although numerous improvements have been made in optimizing cloud cost and security using Artificial Intelligence (AI), there are various obstacles that hinder the use of AI in the process. There are problems like explain ability of models, data privacy, adversarial attacks, high computational costs, scalability, and accuracy of predictions among others [20]. Dealing with these issues is important for formulating effective cloud optimization frameworks.

- **Limited Explain ability of AI Models:** Because it may be difficult to comprehend how these models make decisions, particularly in DL, they are frequently referred to as "black boxes". Absence of visibility decreases the trust of cloud administrators in the model.
- **Susceptibility to Adversarial and Cyber Attacks:** These types of AI systems are susceptible to adversarial attacks whereby attackers use malicious inputs to control the output generated by the models. Adversarial attacks can weaken threat-detection systems and cause improper resource distribution and even breaches.
- **Data Privacy and Confidentiality Issues:** Cloud data plays a critical role in cloud computing optimization using AI technologies due to the need for huge amounts of cloud data in making decisions. This poses questions about the privacy and security of data.
- **High Computational and Operational Overhead:** The use of advanced AI approaches such as DL and reinforcement learning involves high computational cost of training and implementation. This processing cost affect the overall cost and affect the practicality of using AI in those environments.
- **Scalability in Dynamic and Multi-Cloud Environments:** The cloud infrastructure is highly dynamic and may consist of more than one cloud service provider. The AI models need to adjust themselves in accordance with different loads, different resources and different networking environments.
- **Reliability and Prediction Accuracy:** The performance of cloud optimization using artificial intelligence is largely influenced by the accuracy of predictions and decision-making. Errors in prediction can result in inefficiencies in the usage of resources, cost escalation, alarm generation, or even security breaches, which make the entire system unreliable.

VI. Literature Review

The literature review gives information about the recent developments in field of AI-based online cost and security optimization using approaches like prediction, resource allocation, automation, and threat detection.

T. K. Chatterjee (2026) explores how autonomous burden optimization, predictive analytics, and dynamic allocating are using AI technology to transform cloud resource allocation. The combination of ML algorithms with cloud infrastructure enables previously unheard-of levels of accuracy in workload classification, automated scaling, and resource forecasting. These capabilities allow businesses to significantly lessen the overcapacity and under provisioning scenarios that plague traditional threshold-based management approaches [21].

L. Emma (2025) Machine learning models are used by AI-powered cloud resource management to forecast typical workloads, flexibly distribute resources, and optimize expenses in real time. This study investigates how scalability and cost control techniques in traditional cloud services like AWS, Azure, and Google Cloud are improved by AI-driven methods including reinforcement learning, DL, and time-series forecasting. Key considerations include predictive scaling, anomaly detection, and optimization of compute, storage, and networking resources [22].

T. J. Akinbolaji (2024) investigates cutting-edge methods for improving cloud-based system efficiency and expense optimization. By analyzing emerging technologies and methodologies, such as serverless computing, AI-driven optimization, and edge computing, propose comprehensive approaches that organizations can implement to maximize the value of their cloud investments [23].

S. Ratnayake (2024) provides a thorough analysis of AI-driven cloud computing methods with an emphasis on fault tolerance, security, optimization, resource allocation, and performance improvement. AI methods, including ML, DL, neural networks, and reinforcement learning, improve scalability, efficiency, and resilience. Furthermore, AI-based resource management and security techniques enable cost reduction, predictive scaling, threat detection, and intrusion prevention in cloud environments [24].

S. Deochake (2023) examines a number of methods for optimising cloud costs, including as resource allocation plans, cloud pricing, and analysis. A discussion of these methods' efficacy and important lessons is provided, along with real-world case examples. This paper's study shows that using online cost optimization strategies can help businesses save a substantial amount of money [25].

R. Vadisetty et al. (2022) discusses AI-based cybersecurity and cloud optimization, highlighting recent developments, challenges, and future research directions. It compares traditional methods with AI-driven approaches and evaluates their impact on response time, fault tolerance, energy efficiency, and security resilience. The study concludes that integrating AI into cloud computing improves operational efficiency and enables autonomous, self-healing cloud infrastructures [26].

Table I summarizes the key studies by comparing their contributions, advantages, limitations, recommendations, and

research gaps, thereby highlighting the need for integrated AI-driven frameworks for secure and cost-efficient cloud environment.

TABLE I. SUMMARY OF THE STUDY ON AI-BASED CLOUD COST AND SECURITY OPTIMIZATION TECHNIQUES

Authors	Study on	Contributions	Advantages	Limitations	Recommendations
Tarun Kumar Chatterjee (2026)	AI-driven cloud resource management	Dynamic resource allocation, predictive analytics, automated workload optimization	Improves forecasting accuracy and scaling efficiency	Limited focus on integrated security and cost-security trade-offs	Develop unified frameworks combining cost, performance, and security
Lawrence Emma (2025)	AI-powered cloud resource and cost management	Predictive scaling, anomaly detection, and real-time cost optimization	Enhances autoscaling and reduces operational costs	Primarily emphasizes cost optimization with limited security integration	Incorporate security-aware AI models for cloud management
Taiwo Joseph Akinbolaji (2024)	Cost and performance optimization in cloud systems	Serverless computing, AI-driven optimization, edge computing	Improves cost efficiency and system performance	Security implications of optimization techniques are not extensively explored	Investigate secure and adaptive optimization strategies
Sanjeewa Ratnayake (2024)	AI techniques in cloud computing	Performance improvement, fault tolerance, security, and resource allocation	Improves scalability, resilience, and threat detection	Mostly survey-based with limited practical validation	Develop real-world AI-enabled cloud optimization frameworks
Saurabh Deochake (2023)	Cloud cost optimization techniques	Pricing analysis, resource allocation, and case studies	Achieves significant cost savings	Focuses mainly on cost aspects, ignoring security concerns	Explore AI-based approaches integrating security with cost optimization
Rahul Vadisetty et al. (2022)	AI-based cybersecurity and cloud optimization	AI-driven security, fault tolerance, energy efficiency, self-healing clouds	Enhances security resilience and operational efficiency	Limited discussion on real-time cost optimization mechanisms	Investigate adaptive and autonomous cloud management models

VII. CONCLUSION AND FUTURE WORK

Artificial Intelligence (AI) is transforming cloud computing by enabling intelligent, adaptive, and efficient management of cloud resources and security services. A thorough analysis of AI-based cloud cost and security optimization strategies was provided in this study. This research looked at the application of AI methods such as ML, DL, and reinforcement learning, among others, to optimize resource allocation, workload scheduling, auto-scaling, power consumption, threat detection, intrusion detection, and access control. It was found from the literature survey that using AI in cloud computing offers great advantages in terms of reducing operational costs, optimizing the usage of resources, and securing against new and emerging cyber-attacks. However, issues such as explain ability, data privacy, adversarial attacks, overheads in computations, and scalability still pose challenges in research. Generally, it can be said that AI is an effective technology in cloud computing.

Further research on cloud computing should focus on developing explainable and privacy-preserving AI models for cloud computing. Moreover, it is imperative to incorporate cost optimization and security features in the unified AI frameworks, improving scalability in multi-cloud systems, and enhancing resilience against adversarial attacks will be essential for building autonomous, secure, and efficient cloud infrastructures.

REFERENCES

[1] R. K. Gadiraju, "Artificial Intelligence for Resource Optimization in Cloud Computing Environments," *J. Electr. Syst.*, vol. 20, no. 6s, pp. 3164–3174, Apr. 2024, doi: 10.52783/jes.9485.

[2] G. Karamchand, "View of AI-Optimized Network Function Virtualization Security in Cloud Infrastructure," *Int. J. Humanit. Inf. Technol.*, vol. 7, no. 3, pp. 1–12, 2025, doi: 10.21590/ijhit.07.03.01.

[3] G. Olaoye, "The Impact of AI on Cloud Cost Optimization and Resource Management," 2025.

[4] C. Surianarayanan, J. J. Lawrence, P. R. Chelliah, E. Prakash, and C. Hewage, "A Survey on Optimization Techniques for Edge Artificial Intelligence (AI)," *Sensors*, vol. 23, no. 3, p. 1279, Jan. 2023, doi: 10.3390/s23031279.

[5] S. R. Mamidi, "Enhancing Cloud Computing Security Through Artificial Intelligence-Based Architecture," *J. Artif. Intell. Gen. Sci. ISSN3006-4023*, vol. 5, no. 1, pp. 63–72, Jun. 2024, doi: 10.60087/jaigs.v5i1.166.

[6] J. B. Mehta, "Autonomous Workload Right-Sizing for Multi-Cloud Cost Optimization," in *2026 International Conference on Artificial Intelligence, Systems, and Emerging Technologies (ICAISET)*, Cairo, Egypt: IEEE, 2026, pp. 1–11, June. doi: 10.1109/ICAISET66439.2026.11541899.

[7] S. Duan *et al.*, "Distributed Artificial Intelligence Empowered by End-Edge-Cloud Computing: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 25, no. 1, pp. 591–624, 2023, doi: 10.1109/COMST.2022.3218527.

[8] Srikanth Chakravarthy Vankayala, "Intelligent Quality Assurance in Cloud-Native Systems: A Deep Learning and Reinforcement Learning Approach to Adaptive Test Coverage Optimization," *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 11, no. 6, pp. 546–553, Nov. 2024, doi: 10.32628/IJSRSET2512555.

[9] S. B. Sunil Kumar Parisa, "AI-Enabled Cloud Security Solutions : A Comparative Review of Traditional vs . Next-Generation Approaches International Journal of Statistical Computation and," *Int. J. Stat. Comput. Simul.*, p. 15, 2024.

[10] C. Lekkala, "AI-Driven Dynamic Resource Allocation in Cloud Computing: Predictive Models and Real-Time Optimization," *J. Artif. Intell. , Mach. Learn. Data Sci.*, vol. 2, no. 2, pp. 1–7, June, 2024, doi: 10.2139/ssrn.4908420.

[11] H. N. Dholariya, "GVIF: A Governed Vector Intelligence Framework for AI-Driven Cloud Data Modernization in Regulated Financial Systems," *Int. J. Comput. Exp. Sci. Eng.*, vol. 12, no. 1, pp. 371–386, January, 2026, doi: 10.22399/ijcesen.4797.

[12] Y. Sanjalawe, S. Al-E'mari, S. Fraihat, and S. Makhadmeh, "AI-driven job scheduling in cloud computing: a comprehensive review," *Artif. Intell. Rev.*, vol. 58, no. 7, p. 197, Apr. 2025, doi: 10.1007/s10462-025-11208-8.

[13] S. Chouliaras and S. Sotiriadis, "An adaptive auto-scaling framework for cloud resource provisioning," *Futur. Gener. Comput. Syst.*, vol. 148, pp. 173–183, Nov. 2023, doi: 10.1016/j.future.2023.05.017.

[14] I. Eleweke, M. F. Umakor, C. W. Ndubuisi, C. G. Amomo, S. Adeniji, and M. Temidayo, "Ai-Driven Threat Detection and Prevention in Cloud Computing Environments," *Am. J. Innov. Sci. Eng.*, vol. 4, no. 3, pp. 49–56, Oct. 2025, doi: 10.54536/ajise.v4i3.5041.

[15] S. Kumara, "A Lightweight Deep Learning Based Classification Models for Non-Human Identity Threat Detection," in *2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC)*, IEEE, Feb. 2026, pp. 1–6. doi: 10.1109/ICAIC67076.2026.11395886.

[16] N. Upadhyay, "AI and Machine Learning for Cloud Security: A Comprehensive Survey of IDS and Threat Detection Methods," *J. Glob. Res. Math. Arch.*, vol. 12, no. 6, pp. 34–41, 2025.

[17] C. Nwachukwu, K. Durodola-Tunde, and C. Akwiwu-Uzoma, "AI-driven anomaly detection in cloud computing environments," *Int. J. Sci. Res. Arch.*, vol. 13, no. 2, pp. 692–710, Nov. 2024, doi: 10.30574/ijsra.2024.13.2.2184.

[18] S. Shivam, T. P. Patel, A. K. Padhy, C. Kulkarni, C. Medicherla, and V. Soni, "Anomaly Detection in Financial Payment Transactions Using Efficient Data-Driven Machine Learning Techniques," *2026 IEEE 5th Int. Conf. AI Cybersecurity*, pp. 1–6, Feb. 2026, doi: 10.1109/ICAIC67076.2026.11395758.

[19] R. Keshava, S. K. Pandurangan, M. Sakthivanitha, S. Parmsivan, G. Sunkara, and R. Maruthi, "AI-Powered Algorithms for the Prevention and Detection of Computer Malware Infections," in *2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, IEEE, Sep. 2025, pp. 1673–1680. doi: 10.1109/ICESC65114.2025.11212519.

[20] A. R. Sarabu, "AI-Driven Strategies for Cloud Cost Optimization," *Int. J. Sci. Technol.*, vol. 162, 2025, doi: 10.71097/IJSAT.v16.i2.4714.

[21] T. K. Chatterjee, "AI-Driven Cloud Optimization for Cost Efficiency," in *ICT for Global Innovations and Solutions*, S. Bhattacharya, Ed., Cham: Springer Nature Switzerland, 2026, pp. 537–555. doi: 10.1007/978-3-032-02853-2_38.

[22] L. Emma, "Ai-Powered Cloud Resource Management : Machine Learning For Dynamic Autoscaling And Cost Optimization," 2025.

[23] T. J. Akinbolaji, "Novel Strategies for Cost Optimization and Performance E nhancement in Cloud- B Based Systems," *Int. J. Mod. Sci. Res. Technol.*, vol. 2, pp. 66–79, 2024, doi: 10.5281/zenodo.13982502.

[24] S. Ratnayake, "A Comprehensive Review Of Ai-Driven Optimization, Resource Management, And Security In Cloud Computing Environments," *Int. J. Sustain. Infrastruct. Cities Soc.*, vol. 9, no. 5, 2024.

[25] S. Deochake, "Cloud Cost Optimization: A Comprehensive Review of Strategies and Case Studies," 2026.

[26] R. Vadisetty, "AI-Driven Cybersecurity: Enhancing Cloud Security with Machine Learning and AI Agents," *J. Informatics Educ. Res.*, vol. 2, no. 2, pp. 19–23, Dec. 2022, doi: 10.52783/jier.v2i2.2819.

