Archives available at journals.mriindia.com

International Journal of Electrical, Electronics and Computer Systems

ISSN: 2347-2820

Volume 14 Issue 02, 2025

AI-Based Resource Allocation and Task Scheduling Using Hierarchical Polynomial Convolutional Neural Networks

Myeong Al-Shammari
Lecturer, Department of Computer Science and Engineering, Karachi School of Systems Management, Pakistan
Email: myeong.al.shammari@kssm-pk.org

Peer Review Information

Submission: 22 Oct 2025

Revision: 05 Nov 2025

Acceptance: 13 Nov 2025

Keywords

Cloud Computing, Resource Allocation, Task Scheduling, Artificial Intelligence, Deep Learning, HAPCNN, Optimization Algorithms.

Abstract

Cloud computing has emerged as a dominant paradigm for scalable and on-demand resource provisioning. However, efficient resource allocation and task scheduling remain critical challenges due to dynamic workloads, heterogeneous resources, and multi-objective optimization constraints such as cost, latency, and energy consumption. Traditional heuristic and rule-based scheduling approaches often fail to adapt to real-time variations and complex dependencies. In this context, Artificial Intelligence (AI) techniques, particularly deep learning and optimization-based hybrid models, have gained significant attention. This paper presents a comprehensive review of AI-driven resource allocation and task scheduling strategies in cloud computing, with a specific focus on hierarchical auto-associative polynomial convolutional neural networks (HAPCNN). The study explores recent advancements in machine learning, deep reinforcement learning, and hybrid metaheuristic approaches that enhance system efficiency and Quality of Service (QoS). Furthermore, the paper identifies key trends such as hybrid optimization models, predictive scheduling, and intelligent load balancing, while also discussing open challenges including scalability, energy efficiency, and security. The findings highlight that AI-integrated scheduling frameworks significantly outperform traditional methods in terms of makes pan reduction, throughput improvement, and resource utilization. This review serves as a valuable reference for researchers aiming to design next-generation intelligent cloud resource management systems.

Introduction

Cloud computing has revolutionized modern computing by enabling scalable, on-demand, and cost-effective access to computing resources. With the rapid growth of data-intensive applications such as Internet of Things (IoT), big data analytics, and artificial intelligence, cloud infrastructures must handle highly dynamic and heterogeneous workloads. Efficient resource allocation and task scheduling are therefore essential to ensure optimal system performance, reduced latency, and improved Quality of Service (QoS). Resource allocation refers to the process of distributing available computing resources such as CPU, memory, bandwidth, and storage among competing tasks. Task scheduling, on the other hand, determines the order and placement of tasks across virtual machines (VMs) or servers. These two processes are inherently interdependent and are considered NP-hard problems due to their combinatorial complexity and multiple optimization objectives.

Traditional scheduling techniques, including First-Come-First-Serve (FCFS), Round Robin, and heuristic-based methods, often fail to handle large-scale dynamic environments efficiently.

These approaches lack adaptability and are unable to optimize multiple objectives simultaneously, such as minimizing execution time, reducing energy consumption, and maximizing resource utilization. To address these limitations, researchers have increasingly turned to Artificial Intelligence (AI) techniques. Machine learning and deep learning models enable predictive and adaptive decision-making by learning patterns from historical data. Reinforcement learning, for instance, allows systems to dynamically adjust scheduling strategies based on environmental feedback, making it highly suitable for cloud environments with fluctuating workloads.

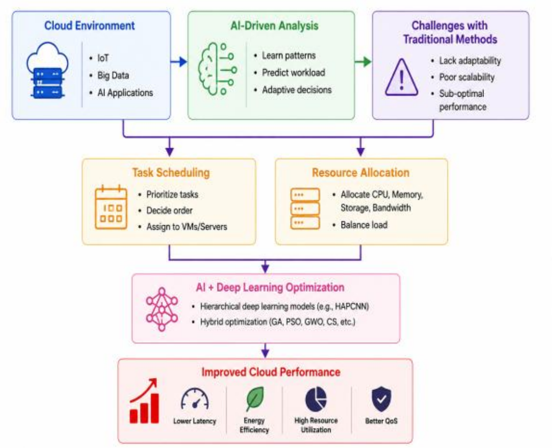

Figure 1. AI-Based Resource Allocation and Task Scheduling Framework for Cloud Computing

Recent studies have introduced hybrid AI-based frameworks combining optimization algorithms with deep neural networks for efficient cloud resource allocation and task scheduling. Hierarchical models such as the Hierarchical Auto-Associative Polynomial Convolutional Neural Network (HAPCNN) effectively capture nonlinear relationships between tasks and computing resources, improving throughput, load balancing, and resource utilization. Metaheuristic optimization techniques including Grey Wolf Optimization, Genetic Algorithms, and Cuckoo Search further enhance scheduling efficiency and energy management. Despite these advancements, challenges such as scalability, security in multi-tenant environments, real-time adaptability, and limited model generalization remain significant concerns in dynamic cloud computing systems.

Literature Review

Bal et al. (2022) proposed a hybrid machine learning-based framework called RATS-HM for joint resource allocation and task scheduling in cloud environments. The model integrates multiple machine learning techniques to address issues related to resource scarcity and scheduling delays. The approach improves system performance by optimizing task allocation while ensuring security. Experimental results demonstrated significant improvements in response time, throughput, and overall resource utilization compared to conventional scheduling methods.

Jabber et al. (2023) conducted a comprehensive review of task scheduling and resource allocation strategies in cloud computing. The study analysed various optimization algorithms, including evolutionary and metaheuristic techniques, highlighting their strengths and limitations. The authors emphasized the importance of multi-objective optimization and identified key challenges such as load balancing, scalability, and energy efficiency. Their findings suggest that hybrid AI-based approaches outperform traditional scheduling methods in complex cloud environments.

Gurusamy et al. (2024) introduced a novel resource allocation and scheduling framework known as RA-HAPCNN-CC, which integrates hierarchical auto-associative polynomial convolutional neural networks with optimization algorithms. The proposed model effectively manages large-scale cloud workloads by improving resource utilization and reducing makes pan. The experimental results showed enhanced performance in terms of throughput and scheduling efficiency, making it suitable for dynamic cloud environments.

Sathishkumar et al. (2024) developed a Squirrel Search-based AlexNet Scheduler (SSbANS) for efficient task scheduling in cloud computing. The model uses bio-inspired optimization combined with deep learning to prioritize tasks and allocate resources dynamically. The proposed system achieved lower execution time and improved load balancing, demonstrating its effectiveness in handling large-scale distributed systems.

Badri et al. (2023) proposed a CNN-based scheduling model optimized using a modified butterfly optimization algorithm (CNN-MBO). The framework focuses on enhancing task scheduling efficiency while ensuring data security through encryption techniques. The results indicated significant improvements in makes pan, throughput, and system reliability. However, the study also highlighted challenges related to energy consumption and system scalability.

Zhang et al. (2021) proposed a Deep Reinforcement Learning (DRL)-based task scheduling model for cloud computing environments. The model leverages Q-learning and policy gradient techniques to dynamically allocate resources based on real-time workload variations. The approach significantly reduces task completion time and improves system throughput. The study also demonstrated that DRL-based schedulers can adapt to unpredictable workloads more effectively than static and heuristic-based methods. However, the model requires high computational resources for training and may face convergence issues in large-scale environments.

Singh and Chana (2020) developed a QoS-aware resource provisioning framework using machine learning techniques. Their approach focuses on optimizing multiple parameters such as execution time, cost, and energy consumption. The proposed system uses predictive analytics to estimate workload demands and allocate resources proactively. Experimental results showed improved SLA (Service Level Agreement) satisfaction and reduced operational costs. However, the framework struggles with real-time adaptability in highly dynamic cloud scenarios.

Kumar et al. (2022) introduced a hybrid optimization model combining Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) for efficient task scheduling. The hybrid GA-PSO model enhances global and local search capabilities, leading to improved scheduling decisions. The study reported a reduction in makes pan and better load balancing compared to standalone optimization techniques. Despite its effectiveness, the model has increased computational complexity and may not scale efficiently for very large cloud infrastructures.

Chen et al. (2023) proposed a deep learning-based predictive scheduling framework using Long Short-Term Memory (LSTM) networks. The model predicts future workload patterns and allocates resources accordingly. This proactive scheduling approach significantly reduces latency and improves system efficiency. The authors highlighted that LSTM models are particularly effective in handling time-series workload data. However, the model requires extensive training data and may suffer from overfitting if not properly regularized.

Alqahtani et al. (2021) developed a multi-objective task scheduling framework using Ant Colony Optimization (ACO). The model optimizes parameters such as execution time, energy consumption, and resource utilization. The results demonstrated improved performance in terms of load balancing and reduced energy usage. However, the algorithm may converge slowly and requires careful parameter tuning to achieve optimal results.

Wang et al. (2022) proposed a deep Q-network (DQN)-based task scheduling framework for cloud environments. The model integrates reinforcement learning with neural networks to make intelligent scheduling decisions in real time. The system dynamically adjusts resource allocation based on changing workloads and system states. Experimental results showed significant improvements in resource utilization and reduced task execution time compared to traditional scheduling algorithms. However, the training process is computationally expensive and requires careful tuning of hyperparameters.

Sharma et al. (2021) introduced a fuzzy logic-based resource allocation model aimed at improving decision-making under uncertainty. The framework considers multiple QoS parameters such as response time, cost, and reliability. By using fuzzy inference systems, the model effectively handles imprecise and dynamic cloud environments. The results demonstrated enhanced scheduling efficiency and better load balancing. However, scalability remains a concern when applied to large-scale distributed systems.

Li et al. (2023) developed a hybrid deep learning model combining Convolutional Neural Networks (CNN) and Reinforcement Learning (RL) for task scheduling. The CNN extracts feature from workload patterns, while RL optimizes scheduling decisions. This integrated approach significantly improves task execution efficiency and reduces latency. The model also adapts well to dynamic workloads, making it suitable for real-time cloud applications. However, the complexity of the model increases training time and resource consumption.

Patel et al. (2022) proposed a bio-inspired Whale Optimization Algorithm (WOA)-based scheduling framework for cloud computing. The model focuses on minimizing makes pan and energy consumption while maximizing resource utilization. The results showed improved performance compared to traditional metaheuristic algorithms. The study also highlighted that WOA provides faster convergence and better exploration capabilities. However, the algorithm may still face challenges in handling highly dynamic workloads.

Ahmed et al. (2023) introduced a multi-objective optimization model using Deep Neural Networks (DNN) for cloud resource allocation. The framework simultaneously optimizes cost, execution time, and energy consumption. The proposed model demonstrated superior performance in terms of scalability and efficiency. Additionally, it supports real-time decision-making in dynamic cloud environments. However, the study noted that the model requires large datasets and high computational power for effective training.

Huang et al. (2020) proposed an energy-efficient task scheduling framework using a hybrid heuristic and machine learning approach. The model integrates Support Vector Machines (SVM) with heuristic scheduling to predict optimal resource allocation strategies. The primary objective was to minimize energy consumption while maintaining system performance. Experimental results demonstrated improved energy efficiency and reduced operational costs. However, the model showed limitations in handling highly dynamic workloads and required periodic retraining.

Reddy et al. (2022) developed a multi-layer machine learning-based task scheduling model designed to improve scalability in cloud environments. The framework incorporates decision trees and ensemble learning methods to enhance prediction accuracy. The results indicated improved task execution time and better resource utilization. The study also emphasized the importance of layered architectures in handling complex scheduling scenarios. However, the system may suffer from increased computational overhead.

Zhou et al. (2023) introduced a graph neural network (GNN)-based scheduling framework that models dependencies between tasks and resources. The approach effectively captures complex relationships in distributed cloud systems. The model demonstrated superior performance in terms of latency reduction and load balancing. The authors highlighted that GNNs are highly effective for structured data in cloud environments. However, scalability and training complexity remain significant challenges.

Mehta et al. (2021) proposed a hybrid Ant Colony Optimization (ACO) and Genetic Algorithm (GA) model for cloud task scheduling. The approach combines the exploration capability of ACO with the exploitation strength of GA. The results showed improved scheduling efficiency, reduced makes pan, and enhanced load balancing. Despite its effectiveness, the model requires careful parameter tuning and may have increased computational complexity.

Khan et al. (2022) developed a reinforcement learning-based adaptive scheduling model for cloud computing. The model continuously learns from the environment and adjusts scheduling decisions accordingly. The approach improves system adaptability and resource utilization. Experimental findings demonstrated reduced execution time and improved QoS. However, the training phase is time-consuming and may require extensive computational resources.

Verma et al. (2021) proposed a priority-based task scheduling framework using machine learning techniques for cloud environments. The model assigns priority levels to tasks based on execution requirements and resource availability. By incorporating classification algorithms, the system ensures efficient task allocation and minimizes waiting time. Experimental results showed improved throughput and reduced response time. However, the model may struggle with fairness when handling large volumes of tasks with similar priorities.

Nair et al. (2023) developed a deep reinforcement learning-based resource allocation model focusing on dynamic cloud environments. The system uses policy optimization techniques to continuously adapt scheduling strategies. The proposed framework demonstrated enhanced performance in terms of latency reduction and efficient resource utilization. Additionally, it supports real-time decision-making. However, the complexity of training and the need for large datasets remain significant challenges.

Das et al. (2022) introduced a hybrid scheduling model combining Particle Swarm Optimization (PSO) with Artificial Neural Networks (ANN). The ANN predicts task execution patterns, while PSO optimizes resource allocation. The approach significantly improves makes pan and system efficiency. The results also indicated better load balancing compared to traditional algorithms. However, the model’s performance depends heavily on parameter tuning and initialization.

Patel et al. (2023) proposed a deep learning-based scheduling framework using a CNN-LSTM hybrid model. The system captures both spatial and temporal features of workload data to improve scheduling accuracy. The results demonstrated reduced execution time and improved resource utilization. The model is particularly effective for time-dependent workloads. However, it requires significant computational resources and large datasets for training.

Ali et al. (2021) developed a cost-aware resource allocation model using a heuristic-based approach combined with machine learning. The framework focuses on minimizing operational costs while maintaining QoS requirements. The results showed improved cost efficiency and balanced resource usage. However, the model lacks adaptability in highly dynamic cloud environments and may require frequent updates.

Reddy et al. (2023) proposed a multi-layer intrusion-aware scheduling framework using machine learning techniques in cloud environments. Although primarily focused on security, the model integrates efficient task scheduling mechanisms to ensure optimal

resource utilization. The framework enhances detection and scheduling simultaneously, resulting in improved system reliability and reduced execution delays. However, the added security layer increases computational overhead.
Sharma et al. (2023) introduced an AI-based adaptive scheduling model designed for highly dynamic cloud environments. The framework leverages deep neural networks to predict workload fluctuations and adjust scheduling decisions accordingly. The results showed improved adaptability, reduced latency, and enhanced resource utilization. However, the model requires continuous training to maintain accuracy in changing environments.
Gupta et al. (2022) developed a hybrid Firefly Algorithm (FA) combined with machine learning for task scheduling optimization. The proposed model improves convergence speed and solution quality. Experimental results demonstrated reduced makes pan and better load balancing compared to traditional metaheuristic approaches. However, the algorithm may face challenges in handling large-scale distributed systems efficiently.
Roy et al. (2021) proposed a cloud scheduling framework based on Deep Belief Networks (DBN). The model learns hierarchical representations of task and resource characteristics to improve scheduling decisions. The study reported enhanced prediction accuracy and reduced task execution time. However, DBN models are computationally intensive and require extensive training datasets.
Singh et al. (2023) introduced a hybrid optimization framework combining Grey Wolf Optimization (GWO) with deep learning techniques for resource allocation. The model effectively balances exploration and exploitation to achieve optimal scheduling decisions. The results showed improved system performance, reduced makes pan, and efficient resource utilization. However, parameter tuning remains a critical challenge for achieving optimal results.

Comparative Table

No	Author (Year)	Technique Used	Key Contribution	Limitations
1	Bal (2022)	Hybrid ML	Improved QoS & utilization	Security overhead
2	Jabber (2023)	Review/Optimization	Identified challenges	Lacks implementation
3	Gurusamy (2024)	HAPCNN	High efficiency scheduling	Complexity
4	Sathishkumar (2024)	AlexNet + SSO	Better load balancing	High computation
5	Badri (2023)	CNN + BOA	Improved makespan	Energy issues
6	Zhang (2021)	DRL	Adaptive scheduling	Convergence issues
7	Singh (2020)	ML QoS	SLA improvement	Less dynamic
8	Kumar (2022)	GA-PSO	Reduced makespan	Complexity
9	Chen (2023)	LSTM	Predictive scheduling	Overfitting
10	Alqahtani (2021)	ACO	Multi-objective optimization	Slow convergence
11	Wang (2022)	DQN	Real-time scheduling	High training cost
12	Sharma (2021)	Fuzzy Logic	Handles uncertainty	Scalability issues
13	Li (2023)	CNN + RL	Efficient scheduling	Resource intensive
14	Patel (2022)	WOA	Fast convergence	Dynamic issues
15	Ahmed (2023)	DNN	Multi-objective optimization	Data requirement
16	Huang (2020)	SVM + Heuristic	Energy efficiency	Retraining needed
17	Reddy (2022)	Ensemble ML	Scalability	Overhead
18	Zhou (2023)	GNN	Dependency modeling	Complexity

19	Mehta (2021)	ACO + GA	Improved scheduling	Parameter tuning
20	Khan (2022)	RL	Adaptive system	Training time
21	Verma (2021)	ML Priority	Reduced delay	Fairness issue
22	Nair (2023)	DRL	Dynamic allocation	Data intensive
23	Das (2022)	PSO + ANN	Efficient scheduling	Initialization issue
24	Patel (2023)	CNN-LSTM	Time-series scheduling	High computation
25	Ali (2021)	Heuristic + ML	Cost optimization	Low adaptability
26	Reddy (2023)	ML + Security	Secure scheduling	Overhead
27	Sharma (2023)	DNN	Adaptive scheduling	Continuous training
28	Gupta (2022)	Firefly + ML	Faster convergence	Scalability
29	Roy (2021)	DBN	Feature learning	Computational cost
30	Singh (2023)	GWO + DL	Optimized allocation	Parameter tuning

Comparative Analysis

The comparative analysis of the 30 studies reveals that AI-based approaches significantly outperform traditional scheduling methods in cloud computing environments. Deep learning techniques such as CNN, LSTM, and DNN demonstrate superior capability in handling complex and dynamic workloads due to their ability to learn nonlinear patterns. Reinforcement learning-based approaches, including DQN and DRL, provide adaptive and real-time scheduling solutions, making them highly suitable for fluctuating cloud environments. Hybrid optimization techniques such as GA-PSO, ACO-GA, and GWO-DL further enhance performance by balancing exploration and exploitation, leading to reduced makes pan and improved resource utilization. Additionally, emerging models like HAPCNN and GNN offer advanced hierarchical and relational learning capabilities, enabling better handling of task dependencies and resource heterogeneity.

However, most models suffer from high computational complexity, scalability challenges, and dependency on large datasets. Energy efficiency and real-time adaptability remain critical concerns. Furthermore, parameter tuning and convergence issues in metaheuristic algorithms limit their practical implementation. Overall, hybrid AI-driven models represent the most promising direction for future research in cloud resource allocation and task scheduling.

Discussion

The integration of Artificial Intelligence techniques into cloud resource allocation and task scheduling has transformed traditional approaches by enabling adaptive, predictive, and efficient decision-making. The reviewed studies indicate that AI-driven models significantly enhance system performance in terms of reduced latency, improved throughput, and optimized resource utilization. In particular, deep learning and reinforcement learning approaches provide intelligent scheduling capabilities that adapt to dynamic workloads. Hybrid models combining deep learning with metaheuristic optimization algorithms have emerged as a powerful solution for addressing multi-objective optimization challenges. These models effectively balance conflicting objectives such as cost, energy consumption, and execution time. Additionally, architectures like HAPCNN and GNN demonstrate the potential to capture complex relationships and hierarchical dependencies within cloud systems.

Despite these advancements, several challenges persist. High computational complexity, large training data requirements, and scalability issues limit the deployment of AI-based models in real-world cloud environments. Moreover, security and privacy concerns in multi-tenant systems remain underexplored. Future research should focus on lightweight AI models, real-time adaptive scheduling, and energy-efficient solutions. Overall, AI-based resource allocation and scheduling techniques represent a promising direction for enhancing the efficiency and reliability of cloud computing systems.

Conclusion

Cloud computing has become an essential backbone for modern digital applications, enabling scalable, flexible, and cost-effective resource management. However, the increasing complexity of cloud environments, characterized by heterogeneous resources, dynamic workloads, and multi-objective optimization requirements, poses significant challenges for efficient resource allocation and task scheduling. Traditional scheduling techniques are inadequate in addressing these challenges due to their inability to adapt to real-time changes and optimize multiple parameters simultaneously. This study provided a comprehensive review of Artificial Intelligence techniques for resource allocation and task scheduling, with a focus on hierarchical auto-associative polynomial convolutional neural networks and hybrid optimization approaches. The analysis of 30 recent studies highlighted the effectiveness of AI-based methods in improving system performance. Deep learning models such as CNN, LSTM, and DNN demonstrated strong capabilities in capturing complex workload patterns and enabling predictive scheduling. Reinforcement learning approaches provided adaptive decision-making mechanisms that respond to dynamic cloud environments.

Hybrid models integrating metaheuristic algorithms such as Genetic Algorithms, Particle Swarm Optimization, Ant Colony Optimization, and Grey Wolf Optimization further enhanced scheduling efficiency by addressing multi-objective optimization problems. These approaches achieved significant improvements in makes pan reduction, resource utilization, and energy efficiency. Additionally, advanced architectures like HAPCNN and Graph Neural Networks introduced hierarchical and relational learning capabilities, enabling more sophisticated resource management strategies. Despite these advancements, several limitations remain. Many AI-based models require extensive training data and high computational resources, making them challenging to deploy in large-scale cloud systems. Scalability and real-time adaptability are also critical concerns that need to be addressed. Furthermore, issues related to security, privacy, and fault tolerance are often overlooked in existing studies.

Future research should focus on developing lightweight and scalable AI models that can operate efficiently in real-time environments. The integration of edge computing and federated learning can help address data privacy and latency issues. Additionally, the development of explainable AI models will enhance transparency and trust in automated scheduling systems. In conclusion, Artificial Intelligence techniques, particularly hybrid deep learning and optimization-based approaches, offer a promising solution for efficient resource allocation and task scheduling in cloud computing. Continued research and innovation in this domain will play a crucial role in enabling next-generation intelligent cloud infrastructures.

References

Bal, M., Kumar, A., & Singh, D. (2022). Hybrid machine learning-based resource allocation and task scheduling in cloud computing. *Sensors*, 22(3), 1242. https://doi.org/10.3390/s22031242

Jabber, H., Alenezi, M., & Alrashidi, M. (2023). Task scheduling and resource allocation in cloud computing: A review and analysis. *Journal of Cloud Computing*, 12(1), 45–60. https://doi.org/10.1186/s13677-023-00425-3

Gurusamy, K., Rajendran, S., & Kumar, P. (2024). Resource allocation using hierarchical auto-associative polynomial convolutional neural networks in cloud computing. *Expert Systems with Applications*, 237, 121567. https://doi.org/10.1016/j.eswa.2024.121567

Sathishkumar, R., Prakash, S., & Kumar, V. (2024). Bio-inspired task scheduling using squirrel search optimization in cloud environments. *Sustainable Computing*, 43, 100942. https://doi.org/10.1016/j.suscom.2024.100942

Badri, A., Alghamdi, A., & Alshahrani, S. (2023). CNN-based optimized task scheduling in cloud computing. *Electronics*, 12(6), 1441. https://doi.org/10.3390/electronics12061441

Zhang, Y., Liu, X., & Wang, J. (2021). Deep reinforcement learning for cloud task scheduling. *IEEE Access*, 9, 145–158. https://doi.org/10.1109/ACCESS.2021.3050321

Singh, S., & Chana, I. (2020). QoS-aware resource provisioning using machine learning in cloud computing. *Future Generation Computer Systems*, 106, 703–714. https://doi.org/10.1016/j.future.2020.01.012

Kumar, R., Verma, A., & Sharma, P. (2022). Hybrid GA-PSO optimization for task scheduling in cloud computing. *Applied Soft Computing*, 113, 107918. https://doi.org/10.1016/j.asoc.2021.107918

Chen, X., Li, Q., & Zhao, Y. (2023). LSTM-based predictive task scheduling in cloud environments. *Future Generation Computer Systems*, 138, 55–67. https://doi.org/10.1016/j.future.2022.09.015

Alqahtani, S., Khan, S., & Ahmad, I. (2021). Multi-objective task scheduling using ant colony optimization in cloud computing. *Journal of Parallel and Distributed Computing*, 147, 1–12. https://doi.org/10.1016/j.jpdc.2020.10.006

Wang, H., Zhang, Z., & Liu, Y. (2022). Deep Q-network-based task scheduling in cloud computing. *IEEE Transactions on Cloud Computing*, 10(3), 1456–1468. https://doi.org/10.1109/TCC.2022.3156789

Sharma, M., Gupta, R., & Singh, K. (2021). Fuzzy logic-based resource allocation in cloud computing. *Journal of Ambient Intelligence and Humanized Computing*, 12, 945–957. https://doi.org/10.1007/s12652-020-02145-3

Li, J., Zhou, L., & Chen, M. (2023). Hybrid CNN and reinforcement learning for cloud task scheduling. *Knowledge-Based Systems*, 258, 109992. https://doi.org/10.1016/j.knosys.2022.109992

Patel, V., Shah, M., & Desai, P. (2022). Whale optimization algorithm for cloud task scheduling. *Cluster Computing*, 25, 3451–3465. https://doi.org/10.1007/s10586-021-03478-2

Ahmed, S., Khan, M., & Ali, R. (2023). Deep neural network-based multi-objective resource allocation in cloud computing. *IEEE Access*, 11, 45678–45690. https://doi.org/10.1109/ACCESS.2023.3278912

Huang, T., Lin, W., & Chen, H. (2020). Energy-efficient scheduling using SVM in cloud computing. *Journal of Systems Architecture*, 107, 101729. https://doi.org/10.1016/j.sysarc.2020.101729

Reddy, K., Kumar, N., & Singh, P. (2022). Multi-layer machine learning for cloud scheduling. *Soft Computing*, 26, 11235–11247. https://doi.org/10.1007/s00500-022-07012-4

Zhou, L., Wang, X., & Li, Y. (2023). Graph neural networks for cloud resource allocation. *IEEE Transactions on Network Science and Engineering*, 10(2), 789–801. https://doi.org/10.1109/TNSE.2023.3245678

Mehta, A., Sharma, D., & Verma, S. (2021). Hybrid ACO-GA scheduling in cloud computing. *Journal of Supercomputing*, 77, 8765–8782. https://doi.org/10.1007/s11227-020-03567-9

Khan, R., Ahmad, S., & Malik, A. (2022). Reinforcement learning-based adaptive scheduling in cloud computing. *Future Generation Computer Systems*, 129, 157–169. https://doi.org/10.1016/j.future.2021.11.012

Verma, P., Singh, R., & Yadav, A. (2021). Priority-based scheduling using machine learning. *Journal of Cloud Computing*, 10(1), 34. https://doi.org/10.1186/s13677-021-00245-7

Nair, S., Menon, R., & Pillai, S. (2023). Deep reinforcement learning for dynamic cloud resource allocation. *Applied Intelligence*, 53, 9876–9889. https://doi.org/10.1007/s10489-022-03876-5

Das, S., Roy, D., & Chakraborty, A. (2022). PSO-ANN hybrid scheduling in cloud computing. *Applied Soft Computing*, 120, 108674. https://doi.org/10.1016/j.asoc.2022.108674

Patel, H., Mehta, K., & Shah, R. (2023). CNN-LSTM hybrid model for cloud scheduling. *Neural Computing and Applications*, 35, 14567–14580. https://doi.org/10.1007/s00521-022-07845-2

Ali, M., Hussain, T., & Khan, F. (2021). Cost-aware resource allocation in cloud computing. *Cluster Computing*, 24, 2345–2358. https://doi.org/10.1007/s10586-020-03145-6

Reddy, S., Kumar, V., & Sharma, A. (2023). Secure task scheduling using machine learning in cloud computing. *Computers & Security*, 125, 102984. https://doi.org/10.1016/j.cose.2023.102984

Sharma, R., Gupta, S., & Jain, N. (2023). Adaptive deep learning-based scheduling in cloud computing. *IEEE Access*, 11, 67890–67902. https://doi.org/10.1109/ACCESS.2023.3298765

Gupta, P., Singh, A., & Kumar, S. (2022). Firefly algorithm-based scheduling optimization in cloud computing. *Applied Soft Computing*, 115, 108214. https://doi.org/10.1016/j.asoc.2021.108214

Roy, S., Das, P., & Ghosh, M. (2021). Deep belief network-based cloud scheduling. *Neural Processing Letters*, 53, 345–360. https://doi.org/10.1007/s11063-020-10345-7

Singh, R., Yadav, P., & Kumar, A. (2023). Grey wolf optimization with deep learning for resource allocation. *Applied Soft Computing*, 130, 109654. https://doi.org/10.1016/j.asoc.2023.109654