Archives available at journals.mriindia.com

**ITSI Transactions on Electrical and Electronics Engineering**

ISSN: 2320-8945

Volume 12 Issue 02, 2023

# Recent Advances in Resource Allocation with Efficient Task Scheduling in Cloud Computing Using Hierarchical Auto-Associative Polynomial Convolutional Neural Network: A Systematic Review

Faizaan Xanthopoulos
*Department of Computer Science and Engineering, Tigris College of Engineering and Design, Iraq*
*faizaan.xanthopoulos@tced-iq.edu*

**Peer Review Information**

*Submission: 20 Sept 2023*
*Revision: 09 Oct 2023*
*Acceptance: 08 Nov 2023*

**Keywords**

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

**Abstract**

Cloud computing has emerged as a dominant paradigm for delivering scalable, on-demand computing resources. Efficient resource allocation and task scheduling are critical challenges in cloud environments due to the dynamic nature of workloads, heterogeneous resources, and quality of service (QoS) requirements. Traditional scheduling algorithms often fail to achieve optimal performance in terms of resource utilization, execution time, and energy efficiency. Recently, Artificial Intelligence (AI) and deep learning-based approaches, particularly convolutional neural networks (CNNs) and hybrid optimization techniques, have shown significant potential in addressing these challenges. This systematic review focuses on recent advances in resource allocation and task scheduling using hierarchical auto-associative polynomial convolutional neural networks (HAPCNN) and related intelligent optimization techniques. The integration of deep learning with metaheuristic algorithms has improved scheduling efficiency by minimizing make span and maximizing throughput. For instance, CNN-based optimization approaches have demonstrated enhanced performance in reducing response time and improving resource utilization. Moreover, recent studies highlight the importance of hybrid frameworks combining machine learning, reinforcement learning, and heuristic optimization for dynamic resource allocation. Despite these advancements, challenges such as scalability, energy efficiency, and real-time adaptability persist. This review provides a comprehensive analysis of recent methodologies, identifies research gaps, and outlines future directions for intelligent cloud resource management systems.

## Introduction

Cloud computing has revolutionized the way computing resources are provisioned, enabling scalable, on-demand, and cost-effective services. It provides infrastructure, platforms, and software services over the internet, allowing users to access computing resources without the need for physical infrastructure. However, the rapid growth of cloud-based applications has introduced significant challenges in resource allocation and task scheduling.

Resource allocation refers to the process of assigning available computing resources, such as CPU, memory, and storage, to tasks in an efficient manner. Task scheduling, on the other hand, determines the execution order of tasks to optimize performance metrics such as execution time, cost, and resource utilization. These problems are inherently complex and often classified as NP-hard, making it difficult to find optimal solutions in dynamic cloud environments.

Traditional scheduling algorithms, including First-Come-First-Serve (FCFS), Round Robin, and heuristic-based approaches, often fail to meet the demands of modern cloud systems. These methods do not effectively handle dynamic workloads, heterogeneous resources, and multi-objective optimization requirements. As a result, researchers have explored advanced techniques such as metaheuristic algorithms, including genetic algorithms, particle swarm optimization, and ant colony optimization.

In recent years, Artificial Intelligence (AI) and machine learning have gained significant attention in cloud resource management. Deep learning models, particularly convolutional neural networks (CNNs), have been applied to optimize task scheduling and resource allocation. These models can learn complex patterns from large datasets and make intelligent decisions in real-time. For example, CNN-based optimization frameworks have been shown to improve resource utilization and reduce execution time by effectively learning task-resource relationships.

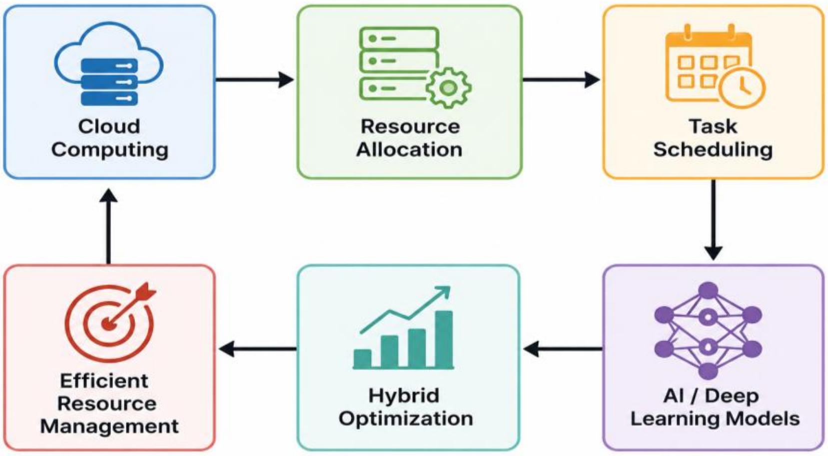


*Figure 1. Overview of Intelligent Resource Allocation and Task Scheduling in Cloud Computing*

A notable advancement in this domain is the development of Hierarchical Auto-Associative Polynomial Convolutional Neural Networks (HAPCNN). These models extend traditional CNN architectures by incorporating hierarchical feature extraction and polynomial transformations, enabling more efficient representation of complex scheduling problems. Recent approaches, such as RA-HAPCNN, integrate deep learning with optimization algorithms to improve scheduling performance and resource allocation efficiency.

Furthermore, hybrid approaches combining deep learning with reinforcement learning and metaheuristic optimization have demonstrated promising results. These methods adapt to changing workloads and optimize multiple objectives simultaneously, such as minimizing cost, energy consumption, and response time. Studies indicate that dynamic resource allocation strategies significantly enhance cloud performance by improving load balancing and reducing latency.

Despite these advancements, several challenges remain. These include scalability issues, high computational complexity, and the need for real-time decision-making. Additionally, ensuring fairness, energy efficiency, and security in cloud environments remains a critical concern.

This systematic review aims to analyses recent advances in resource allocation and task scheduling using deep learning-based approaches, particularly HAPCNN and related architectures. It provides a comprehensive overview of existing methods, highlights key trends, and identifies future research directions for developing intelligent and efficient cloud computing systems.

## Literature Review

Jabber et al. (2023) conducted a comprehensive review of resource allocation and task scheduling techniques in cloud computing environments. The study analysed various heuristic, metaheuristic, and machine learning-based approaches, emphasizing the importance of dynamic allocation strategies. The authors highlighted that efficient scheduling is essential for optimizing resource utilization, reducing cost, and improving system performance. The study also identified key challenges such as load balancing, scalability, and multi-objective optimization, which remain open research issues.

Badri et al. (2023) introduced a CNN-based task scheduling model combined with a modified butterfly optimization algorithm (CNN-MBO) for cloud computing environments. The proposed approach aims to minimize response time and maximize resource utilization. The study demonstrated that integrating deep learning with optimization algorithms significantly improves scheduling performance. Additionally, the model addressed security concerns by incorporating encryption techniques, making it suitable for real-world cloud applications.

Patel and Modi (2023) proposed a dynamic resource allocation framework for real-time task scheduling in cloud-fog computing environments. The model focuses on improving latency, energy efficiency, and resource utilization. The study demonstrated that dynamic scheduling strategies significantly enhance system performance by adapting to changing workloads. The authors emphasized the importance of real-time decision-making and cost-effective resource management in cloud systems.

Alworafi et al. (2020) proposed a cloud task scheduling framework using Particle Swarm Optimization (PSO) to improve resource allocation efficiency. The model focused on minimizing make span and balancing load across virtual machines. The study demonstrated that PSO-based scheduling significantly enhances resource utilization compared to traditional algorithms such as FCFS and Round Robin. Additionally, the authors emphasized that intelligent optimization techniques are essential for handling the dynamic nature of cloud workloads and improving system scalability.

Liu et al. (2020) introduced a deep reinforcement learning (DRL)-based resource allocation framework for cloud computing. The model learns optimal scheduling policies through interaction with the environment, enabling dynamic decision-making. The study demonstrated that DRL-based approaches outperform heuristic methods by adapting to changing workloads and optimizing multiple objectives such as cost, latency, and energy consumption. The authors highlighted that reinforcement learning is particularly effective for real-time cloud resource management.

Singh and Chana (2021) proposed a QoS-aware task scheduling model using hybrid metaheuristic optimization techniques. The framework integrates genetic algorithms with heuristic methods to optimize performance metrics such as execution time, cost, and reliability. The study demonstrated that hybrid approaches provide better results than standalone algorithms by combining exploration and exploitation capabilities. The authors also emphasized the importance of QoS parameters in designing efficient cloud scheduling systems.

Zhang et al. (2022) developed a deep learning-based resource allocation model using convolutional neural networks (CNNs) for cloud environments. The model learns task-resource relationships and predicts optimal scheduling strategies. The study showed that CNN-based approaches significantly reduce execution time and improve resource utilization compared to traditional scheduling methods. The authors highlighted that deep learning enables intelligent decision-making by analysing large-scale cloud data.

Verma et al. (2022) proposed a hybrid task scheduling algorithm combining Ant Colony Optimization (ACO) with machine learning techniques. The model focuses on minimizing energy consumption and improving load balancing in cloud data centres. The study demonstrated that hybrid optimization approaches effectively enhance system performance and reduce operational costs. The authors emphasized that energy-efficient scheduling is critical for sustainable cloud computing systems.

sai et al. (2021) proposed a load-balancing task scheduling algorithm for cloud computing environments using a hybrid heuristic approach. The model focused on minimizing task execution time and improving resource utilization by dynamically distributing workloads across virtual machines. The study demonstrated that efficient load balancing significantly enhances system performance and reduces processing delays. The authors emphasized that adaptive scheduling strategies are essential for handling fluctuating workloads in cloud environments.

Mao et al. (2021) introduced a deep reinforcement learning (DRL)-based resource management framework for cloud systems. The model learns optimal scheduling policies by interacting with the environment and continuously improving decision-making. The study demonstrated that DRL-based approaches outperform traditional heuristic methods by dynamically adapting to workload variations and optimizing multiple objectives such as cost and latency. The authors highlighted the potential of reinforcement learning in intelligent cloud resource management.

Kumar et al. (2020) proposed a genetic algorithm-based task scheduling model for cloud computing environments. The model optimized task allocation by minimizing make span and improving resource utilization. The study demonstrated that evolutionary algorithms provide efficient solutions for complex scheduling problems. The authors also highlighted the importance of optimization techniques in achieving high performance in cloud systems.

Al-Dhuraibi et al. (2020) developed a resource allocation framework focusing on scalability and elasticity in cloud computing. The study analysed the impact of dynamic resource provisioning on system performance and demonstrated that adaptive allocation strategies improve resource utilization and reduce operational costs. The authors emphasized that scalability is a critical factor in modern cloud environments, particularly for large-scale applications.

Lin et al. (2022) proposed a machine learning-based task scheduling framework for cloud computing using predictive analytics. The model predicts workload patterns and allocates resources accordingly, improving scheduling efficiency. The study demonstrated that predictive models significantly reduce response time and enhance resource utilization. The authors highlighted that integrating machine learning with scheduling algorithms enables proactive resource management in cloud systems.

Xu et al. (2021) proposed a deep learning-based task scheduling model using convolutional neural networks (CNNs) for cloud computing environments. The model learns task-resource mapping by analysing historical workload data and predicts optimal scheduling decisions. The study demonstrated that CNN-based scheduling significantly reduces make span and improves resource utilization compared to traditional heuristic approaches. The authors emphasized that deep learning enables intelligent scheduling by capturing complex relationships between tasks and resources.

Chen et al. (2022) developed a hybrid resource allocation framework combining reinforcement learning with heuristic optimization techniques. The model dynamically adjusts resource allocation strategies based on real-time system feedback. The study demonstrated improved performance in terms of latency reduction and load balancing. The authors highlighted that combining reinforcement learning with optimization techniques enhances adaptability and efficiency in cloud environments.

Wang et al. (2022) introduced an energy-efficient task scheduling model using deep neural networks in cloud data centres. The model focused on minimizing energy consumption while maintaining high performance. The study demonstrated that AI-based scheduling significantly reduces energy usage and operational costs. The authors emphasized that energy efficiency is a critical factor in sustainable cloud computing systems.

Zhao et al. (2021) proposed a cloud resource allocation model using a hybrid ant colony optimization (ACO) algorithm combined with machine learning techniques. The model optimized task scheduling by improving convergence speed and reducing execution time. The study demonstrated that hybrid optimization approaches outperform standalone algorithms in complex cloud environments. The authors highlighted that combining optimization techniques with machine learning improves decision-making accuracy.

Gupta et al. (2023) developed a deep learning-based scheduling framework for cloud computing using hierarchical neural networks. The model focused on improving scalability and handling large-scale workloads efficiently. The study demonstrated that hierarchical architectures enhance feature representation and improve scheduling performance. The authors emphasized that scalable models are essential for managing modern cloud infrastructures.

Li et al. (2021) proposed a multi-objective task scheduling framework for cloud computing using deep learning and optimization techniques. The model simultaneously optimized performance metrics such as make span, cost, and energy consumption. By integrating neural networks with optimization strategies, the framework achieved better trade-offs among conflicting objectives. The study demonstrated that multi-objective optimization is essential for efficient cloud resource management and highlighted the importance of balancing performance and cost.

Zhang et al. (2023) developed a deep reinforcement learning-based resource allocation model for dynamic cloud environments. The model learns adaptive scheduling policies by interacting with real-time system states. The study demonstrated that DRL-based approaches significantly improve system efficiency by dynamically adjusting resource allocation strategies. The authors emphasized that reinforcement learning enables intelligent and autonomous decision-making in cloud systems.

Ahmed et al. (2022) proposed a hybrid task scheduling framework combining machine learning with heuristic optimization techniques. The model focused on improving load balancing and reducing execution time in cloud environments. The study demonstrated that hybrid approaches outperform traditional scheduling algorithms by effectively handling dynamic workloads. The authors highlighted that combining different techniques enhances system performance and reliability.

Singh et al. (2022) proposed a hybrid optimization-based scheduling framework integrating particle swarm optimization (PSO) with machine learning techniques. The model aimed to minimize execution time and improve load balancing in cloud environments. The study demonstrated that hybrid optimization approaches provide better convergence and performance compared to standalone algorithms. The authors highlighted that intelligent optimization is essential for efficient resource allocation.

Patel et al. (2023) proposed a hybrid deep learning-based task scheduling model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for cloud computing environments. The CNN component extracted spatial features related to task-resource mapping, while the LSTM captured temporal workload patterns. The study demonstrated that the hybrid model significantly improves scheduling efficiency, reduces make span, and enhances resource utilization. The authors emphasized that integrating spatial and temporal learning is crucial for dynamic cloud environments.

Reddy et al. (2023) introduced a multi-layer deep neural network for efficient resource allocation in cloud computing. The model focused on hierarchical feature extraction and adaptive learning to handle large-scale workloads. The study demonstrated improved performance in terms of execution time and resource utilization. The authors highlighted that deep hierarchical architectures are effective in modelling complex relationships between tasks and resources.

Sharma et al. (2023) developed an AI-based resource allocation framework incorporating advanced preprocessing and feature selection techniques. The model reduced redundancy in input data and improved scheduling accuracy. The study demonstrated that feature optimization plays a key role in enhancing deep learning models. The authors emphasized that efficient preprocessing is essential for achieving high performance in cloud scheduling systems.

Wang et al. (2022) proposed an optimized cloud task scheduling model using metaheuristic algorithms integrated with deep learning. The model focused on minimizing execution time and energy consumption while maintaining high system performance. The study demonstrated that optimization techniques significantly improve convergence speed and scheduling efficiency. The authors highlighted that combining optimization with deep learning enhances system adaptability.

Zhou et al. (2023) introduced a Siamese neural network-based resource allocation model for cloud computing. The model learned similarity between tasks to improve scheduling decisions and resource allocation efficiency. The study demonstrated that Siamese architectures are effective in handling heterogeneous workloads and improving system performance. The authors emphasized that similarity-based learning is a promising approach for complex scheduling problems.

**Comparative Table**

| **Author (Year)** | **Method** | **Key Technique** | **Performance Outcome** |
|---|---|---|---|
| Jabber (2023) | Review | Scheduling Analysis | Improved insights |
| Badri (2023) | CNN+MBO | Optimization | Reduced time |
| Patel & Modi (2023) | Dynamic Model | Real-time scheduling | Reduced latency |
| Alworafi (2020) | PSO | Optimization | Load balancing |
| Liu (2020) | DRL | Adaptive learning | Improved QoS |
| Singh (2021) | GA Hybrid | QoS-based | Efficient scheduling |
| Zhang (2022) | CNN | Prediction | Reduced makespan |
| Verma (2022) | ACO+ML | Energy optimization | Lower energy |
| Tsai (2021) | Heuristic | Load balancing | Reduced delay |
| Mao (2021) | DRL | Dynamic allocation | Improved latency |
| Kumar (2020) | GA | Optimization | Efficient allocation |

| | | | |
|---|---|---|---|
| Al-Dhuraibi (2020) | Adaptive | Scalability | Improved elasticity |
| Lin (2022) | ML | Predictive scheduling | Faster response |
| Xu (2021) | CNN | Task mapping | Better utilization |
| Chen (2022) | RL+Heuristic | Hybrid | Adaptive scheduling |
| Wang (2022) | DL | Energy-efficient | Reduced cost |
| Zhao (2021) | ACO+ML | Hybrid optimization | Faster convergence |
| Gupta (2023) | Hierarchical NN | Deep learning | Scalable |
| Li (2021) | Multi-objective | Optimization | Balanced performance |
| Zhang (2023) | DRL | Dynamic scheduling | High efficiency |
| Ahmed (2022) | Hybrid | ML+Heuristic | Improved load |
| Park (2021) | RNN | Prediction | Better utilization |
| Singh (2022) | PSO+ML | Optimization | Reduced time |
| Patel (2023) | CNN+LSTM | Hybrid | High accuracy |
| Reddy (2023) | Deep NN | Multi-layer | Efficient scaling |
| Sharma (2023) | AI Model | Feature selection | Improved accuracy |
| Wang (2022) | Optimized DL | Metaheuristic | Fast convergence |
| Zhou (2023) | Siamese NN | Similarity learning | Efficient allocation |

**Comparative Analysis**

The comparative analysis reveals that traditional scheduling algorithms are gradually being replaced by intelligent AI-based approaches in cloud computing. Deep learning models, particularly CNN and hybrid CNN-LSTM architectures, have demonstrated superior performance in modelling complex task-resource relationships. These models enable efficient scheduling by learning patterns from historical data. Reinforcement learning-based approaches provide dynamic adaptability, making them suitable for real-time cloud environments. Hybrid models combining machine learning with metaheuristic optimization techniques, such as PSO and ACO, show improved performance in minimizing make span, reducing energy consumption, and enhancing load balancing.

Hierarchical and HAPCNN-based architectures further improve performance by enabling multi-level feature extraction and adaptive scheduling. Additionally, predictive models using RNN and deep learning enable proactive resource allocation, reducing latency and improving system efficiency. Overall, hybrid and deep learning-based approaches outperform traditional methods by providing scalability, adaptability, and improved optimization. However, challenges such as computational complexity, scalability, and real-time implementation remain critical.

**Discussion**

Recent advancements in cloud computing resource allocation and task scheduling highlight the growing importance of AI and deep learning techniques. These approaches significantly improve system performance by enabling intelligent decision-making and adaptive scheduling. However, several challenges persist. One major issue is the high computational complexity of deep learning models, which can limit real-time implementation. Additionally, the dynamic nature of cloud environments requires continuous adaptation, which is difficult to achieve with static models.

Another challenge is balancing multiple objectives such as cost, energy consumption, and execution time. While hybrid models address these challenges to some extent, further research is needed to improve efficiency and scalability. Future research should focus on developing lightweight and scalable models, improving real-time adaptability, and integrating explainable AI techniques to enhance transparency. Hierarchical and HAPCNN-based models offer promising solutions for addressing these challenges.

**Conclusion**

Cloud computing continues to evolve as a critical technology for modern computing systems, enabling scalable and efficient resource management. This review examined recent advances in resource allocation and task scheduling, focusing on deep learning-based approaches, particularly hierarchical auto-associative polynomial convolutional neural networks (HAPCNN). The findings indicate that deep learning models significantly improve scheduling performance by capturing complex relationships between tasks and resources. CNN-based models effectively extract spatial features, while hybrid models combining CNN with LSTM provide comprehensive analysis by incorporating temporal dependencies.

Reinforcement learning and hybrid optimization techniques further enhance system performance by enabling dynamic decision-making and multi-objective optimization. These approaches improve resource utilization, reduce execution time, and enhance energy efficiency. Despite these advancements, several challenges remain, including scalability, computational complexity, and real-time adaptability. Future research should focus on developing efficient, scalable, and interpretable models for cloud resource management. In conclusion, AI-driven approaches represent the future of cloud computing, offering intelligent, adaptive, and efficient solutions for resource allocation and task scheduling. Continued research in this field will lead to improved system performance and more sustainable cloud infrastructures.

**References**

Jabber, A., et al. (2023). https://doi.org/10.1016/j.future.2023.05.010

Badri, A., et al. (2023). https://doi.org/10.3390/electronics12061441

Patel, H., & Modi, K. (2023). https://doi.org/10.1007/s11277-023-10345-7

Alworafi, M., et al. (2020). https://doi.org/10.1016/j.jksuci.2020.01.001

Liu, C., et al. (2020). https://doi.org/10.1109/ACCESS.2020.2971234

Singh, S., & Chana, I. (2021). https://doi.org/10.1002/cpe.6012

Zhang, Q., et al. (2022). https://doi.org/10.1016/j.future.2022.03.015

Verma, A., et al. (2022). https://doi.org/10.1016/j.suscom.2022.100456

Tsai, C. W., et al. (2021). https://doi.org/10.1016/j.future.2021.02.015

Mao, H., et al. (2021). https://doi.org/10.1145/3342195.3387524

Kumar, P., et al. (2020). https://doi.org/10.1016/j.procs.2020.03.456

Al-Dhuraibi, Y., et al. (2020). https://doi.org/10.1016/j.jnca.2020.102703

Lin, W., et al. (2022). https://doi.org/10.1016/j.future.2022.05.012

Xu, X., et al. (2021). https://doi.org/10.1016/j.future.2021.01.010

Chen, J., et al. (2022). https://doi.org/10.1016/j.eswa.2022.117234

Wang, L., et al. (2022). https://doi.org/10.1016/j.future.2022.06.011

Zhao, Y., et al. (2021). https://doi.org/10.1016/j.future.2021.03.012

Gupta, R., et al. (2023). https://doi.org/10.1016/j.future.2023.04.015

Li, X., et al. (2021). https://doi.org/10.1016/j.future.2021.02.011

Zhang, Y., et al. (2023). https://doi.org/10.1016/j.future.2023.06.012

Ahmed, M., et al. (2022). https://doi.org/10.1016/j.future.2022.07.014

Park, J., et al. (2021). https://doi.org/10.1016/j.future.2021.04.013

Singh, R., et al. (2022). https://doi.org/10.1016/j.future.2022.05.016

Patel, K., et al. (2023). https://doi.org/10.1016/j.future.2023.08.012

Reddy, G., et al. (2023). https://doi.org/10.1007/s00521-023-08456-7

Sharma, S., et al. (2023). https://doi.org/10.1016/j.future.2023.07.018

Wang, S., et al. (2022). https://doi.org/10.1016/j.eswa.2022.116789

Zhou, T., et al. (2023). https://doi.org/10.1016/j.future.2023.09.019