Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas

Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)

ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

INTELLIGENT WORKLOAD SCHEDULING USING HYBRID DEEP LEARNING IN MULTI-CLOUD PLATFORMS

Sadeq Thamer Hlama
1. Department of Computer Science College of Science,
University of Sumer, Iraq, Dhi-Qar, Iraq
* Correspondence: sadeq.thamer@uos.edu.iq

Abstract

Cloud computing services have proliferated over the years and a lot of organizations have started to adopt a multi-cloud architecture crossing different cloud providers to provide better workload availability, scalability, and performance. However, considering dynamic resource requirements, heterogeneous setup and variable network conditions efficient scheduling of workloads continue to be a major challenge. Conventional scheduling policies do not usually generalize well to more complex and dynamic environments. Towards this end, we present a hybrid deep learning based intelligent workload scheduling framework for on-demand computing in multi-cloud platforms. We perform a combination of feature extraction through Convolutional Neural Networks (CNN) and temporal workload prediction using a Bidirectional long short-term memory (BiLSTM) network. It enables dynamic workload management through predicting the network and distribution of tasks over multiple clouds to enhance both system performance as well as resource utilization. Experimental results show that our proposed task scheduling method outperforms various traditional and state-of-the-art scheduling methods including round robin, genetic algorithm, and standard LSTM-based schedulers in terms of load balancing efficiency, task completion time, and resource utilization. These findings are of great significance

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

and reflect the effectiveness of hybrid deep learning methods in achieving intelligent resource management in a distributed cloud environment.

Keywords: Multi-cloud computing, workload scheduling, hybrid deep learning, CNN-BiLSTM, resource allocation, intelligent cloud management.

I. Introduction

The emergence of cloud computing changed the way organizations will deploy, manage and scale their computing resources. Multi-cloud environments where services are distributed across multiple cloud providers have therefore become more common, owing to their added flexibility, reliability and vendor independence. By using multi-cloud infrastructures, enterprises can destack workloads to different cloud providers to minimize service outages and enhance system performance.

Despite these advantages, efficient workload scheduling in multi-cloud systems remains a significant challenge. The scheduling algorithms must consider a range of parameters such task complexities, the resources needed to execute a task, network latency and energy consumption. Static heuristics and rule-based static scheduling algorithms, which are common in production today, do not adapt to dynamic workloads with varying resource requirements.

Recent advents like artificial intelligence and deep learning intelligent cloud resource management have presented us new possibilities. By examining historical workload trends, machine learning models can forecast resource requirements and enable proactive provisioning. Hybrid deep learning architectures in terms of incorporating both spatial and temporal learning capabilities can benefit from the scheduling decisions.

In this paper, we present a hybrid deep learning based workload scheduling framework in a multi-cloud environment. We show how CNN can be used as a feature extractor and BiLSTM for predicting the workload, thus enabling

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas

Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

intelligent scheduling decisions resulting in improved system performance and resource utilization..

The key contributions of this work are:

- Predicting Workload Patterns using a Hybrid Deep Learning Model in Multi-Cloud Platform
- A smart task scheduling framework capable of dynamically balancing loads among cloud resources.
- Experiments on proposed approach vs. state-of-the-art scheduling algorithms.

A. Motivation

As cloud infrastructures become more complex, some workload scheduling-related issues have emerged. Multi-cloud systems must achieve high performance and efficient utilization of the existing resources by distributing workloads across multiple distributed and heterogeneous resources as well.

Conventional scheduling techniques such as Round Robin, First Come First Serve (FCFS), and heuristic algorithms are not sufficiently adaptive to change the workload scenario. Common approaches using such techniques lead to other problems like resource under-utilization, higher latency, and load imbalance. Several works aim to overcome these limitations and applied optimization techniques or machine learning models. But current methods have some limitations, such as weak generalization ability, scalability limitation, and inability to capture temporal workload dependencies.

Hybrid deep learning models Benefits ◦ Hybrid tries to address complex scheduling issues and have tremendous benefits. Fig 1 shows the overall working mechanism of cloud workloads using CNN based feature extraction with RNN architectures to capture not only the spatial but also the temporal behavior of cloud workloads. It creates a high demand for the intelligent workload scheduling

Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

framework that can accommodate the dynamic cloud environment and optimize resource utilization to minimize execution time.

II. Literature Survey

Recent work tends to use deep learning and optimization algorithms for workload scheduling in cloud infrastructures.

Sefati et al. For resource scheduling in multi-cloud computing environment, an adaptable cost assignment framework is proposed based on memory-based metaheuristic optimization and recurrent neural forecasting [1]. Their approach infers future resource requirements of distributed cloud infrastructures, based on analyzing previous workload patterns with a recurrent neural model. The forecasting module will also allow the system to pre-emptively predict change in CPU usage, memory consumption and task arrival rate. A memory-based metaheuristic optimization algorithm for resource adaptation based on anticipated workload trends supplements the predictive capability. The proposed hybrid scheme allows the scheduler to adapt efficiently according to different demand situations in heterogeneous multi-cloud environments. The experimental results indicate that the proposed scheduling framework improves resource utilization efficiency, joint accuracy of jobs and event deadlines, response time as well as waiting time of scheduled ahead jobs in comparison with conventional static scheduling approaches, event computational overhead.

Simaiya et al. In this study, [2] developed a hybrid load balancing mechanism based on deep learning to optimize the utilization of resources in cloud computing systems as it faces unevenly distributed loads. This method uses deep learning prediction models enhanced by optimization techniques to estimate the host utilization and distributed workloads on the virtual machines with respect to that effectively. It leverages neural network architecture which is able to learn even hidden relations in the data based on system performance metrics like CPU usage, memory consumption and network latency. The system balances cloud resource

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

consumption according to trends for future utilization by predicting the tasks and accessing resources without overloading certain nodes. To this end, the new model utilizes optimization techniques to choose the best virtual machines for diverse tasks allocation. The hybrid model experimental results were compared with classical heuristic-based load balancing algorithms, showing superior balance of the workloads, a decrease in response time and enhancement in overall throughput.

Deep learning models have been used by Chandrasiri and Meedeniya [3] to create an energy-efficient workflow scheduling framework for cloud computing settings that optimizes resource allocation. Their research tackles one of the most critical challenges in large-scale cloud data centers: high energy consumption. In the proposed method, deep learning algorithms are used to analyze workflow characteristics and resource utilization patterns for optimal task-to-resource mappings. The system enhances resource utilization by forecasting the computational needs and execution times of workflow tasks, which allows for a minimized idle time and low-power consumption resource management. In addition, the proposed scheduling strategy considers energy-aware policy that makes use of low energy computing nodes in task allocation. The study findings revealed that the scheduling method proposed in this paper can effectively minimize the energy utilization, slack time, and production expenses, thus providing a sustainable cloud computing infrastructure.

Sanjalawe et al. [4] proposed a smart load balancing model that utilizes feature selection techniques as well as deep learning algorithms to improve the efficiency of scheduling in cloud environments. They begin with an advanced feature selection algorithm to separate the crucial system parameters that influence scheduling performance: processor load, memory availability, network bandwidth, and task priority level. The system thus enhances the accuracy and performance of subsequent deep learning model by excluding redundant or irrelevant characteristics. These selected features are passed to a deep neural

ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

network in order to predict the most appropriate assignment of tasks over available computing resources. The integral view blends ideas from both of these streams into a single model, simplifying computation and providing better predictions. Experimental evaluation, it also showed that the proposed model improves task scheduling performance while reducing resource contention and increasing overall throughput.

Deep Task Scheduling Technique for Scalable (DSTS) container-based cloud environments managed dynamic workloads in modern containerized cloud infrastructures Muniswamy and Vignesh [5]. Its insight is how to integrate deep learning algorithms with optimization approaches for intelligently scheduling tasks in container clusters. The deep learning section looks at historical data regarding workload patterns and resource needs per container to decide optimal scheduling decisions. An optimization module employs these insights to provision the optimal pool of containers for a task execution. Moreover, the framework promotes scalability for dynamic allocation of resources based on load intensity. Our performance assessments showed that the DSTS framework could offer significantly improved scheduling latency and system overhead in containerized cloud platforms while achieving more efficient task execution with better scalability and resource usage as compared to current methods.

Tripura et al. [6] proposed a framework of distributed cloud resource management that integrates neural-based resources allocation with multi-agent fault tolerance mechanism. This framework utilizes neural networks to forecast resource needs and optimize allocation techniques over distributed cloud nodes. Moreover, to monitor system performance and detect failures in real time, a multi-agent system is integrated. Working in tandem, they reassign workloads to other available nodes when they detect faults or resource bottlenecks in order to maintain service availability. In a distributed cloud environment, this decentralized architecture overall increases system robustness and scalability. Through experimental results, it proved that the presented model has a

ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

considerably better system reliability, fault tolerance and resource allocation efficiency compared to state of the art solutions and is suitable for large scale distributed cloud infrastructures.

Raju et al. to propose IntelliScheduler, a hybrid deep learning-based task scheduling framework for edge-cloud computing environments. In addition, the edge nodes have limited computational resources as compared to the cloud servers and many applications require low-latency processing at the edge-cloud architectures demand efficient scheduling strategies. IntelliScheduler framework leverages predictive deep learning models to analyze the characteristics of incoming tasks and estimate their resource requirements. The scheduler decides whether the tasks could be executed in edge nodes or offloading to cloud servers based on the above predictions. The system moreover integrates adaptive solution strategies that take into account factors such as network latency, availability of resources, and workload intensity. The experimental analysis demonstrated that IntelliScheduler enhances the task scheduling efficiency, latency reduction, and resource utilization, which results in superior performance of edge-cloud systems.

Su et al. [8] introduced a multi-cloud computing system that implements a federated reinforcement learning-based scheduler. Their approach builds upon principles of federated learning to facilitate distributed model training across multiple cloud domains while protecting sensitive data. We devise an approach where each individual cloud platform locally trains a reinforcement learning model based on its operation data, and with aggregated updates from these models across platforms will produce a global scheduling strategy. Data privacy is maintained by the cloud data owners (data center(s)) while allowing the scheduler to learn diverse workload patterns across multiple data centres in the public/private cloud infrastructure. The reinforcement learning agent keeps accessing the cloud environment to obtain system states, performance feedbacks and then determine the best resource allocation policies. This framework showed vast enhancements in optimization of resources, scalability, and scheduling

ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

accuracy across multitude of cloud environments and thereby proposes a strong solution for high interrogated multi-cloud environments.

While these works address scheduling efficiency issues, most models cannot manage multi-cloud workload distributions for complex and dynamic scenarios. To overcome these limitations, the proposed hybrid deep learning framework couples spatial feature representation with temporal workload prediction by integrating both of them.

III. Methodology

The stages in phases of the research methodology are:

Phase 1: Data Collection

We use workload datasets from cloud platforms collected with the following parameters:

CPU utilization, Memory usage, Task arrival rate, Network latency, Resource availability

Phase 2: Data Preprocessing

Data preprocessing steps include:

Data normalization, Missing value handling, Feature extraction, Data splitting (training, validation, testing)

Phase 3: Model Development

Hybrid deep learning models are developed using:

This is done to develop a hybrid deep learning architecture in this phase that best predicts future workload patterns of multi-cloud environments. The new model utilizes a combination of Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks.

CNN-Based Feature Extraction

These CNNs that were used to assist in locating spatial structure in numerical data. The CNN layer extracts high-level workload features from the historical resource usage data.

ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

The CNN architecture includes:
Convolutional layers for pattern detection
Activation functions such as ReLU
Pooling layers for dimensionality reduction
These layers are trained automatically, based on measured workload parameters such as CPU load, memory usage, and task arrival rates.
BiLSTM-Based Workload Prediction
The output of the feature extraction is then fed into a BiLSTM network.
The BiLSTM networks are used for capturing temporal dependencies in sequential data. BiLSTM processes data in both forward and backward directions, so it can learn patterns from the past as well as future workload trends simultaneously, which is different than regular LSTM networks.
The following operations are accomplished by BiLSTM layer:
Captures temporal workload fluctuations
Somme model works learn long (term) dependencies in the usage of resources
Predicts future workload intensity
The hybrid model effectively captures spatial and temporal dynamics in cloud workloads by combining CNN and BiLSTM architectures.
Phase 4: Scheduling Decision Engine
Based on predictions generated by the hybrid deep learning model, scheduling decision engine allocates incoming tasks to appropriate cloud resources.
There are a number of roles that the scheduling engine fulfils:
Workload Prediction Analysis
Pending workloads are predicted and analysed to ensure resources meet demand.
Resource Matching
Each cloud resource is rated based on its current utilization level and expected workload capacity.
Task Allocation Strategy

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

Nodes are assigned tasks in order to minimize execution time and maximize resource usage.

Dynamic Load Balancing

In the multi-cloud environment, if some nodes are overloaded, it dynamically redistribute the tasks across distributed other cloud provider.

This novel decision process allows for effective workload distributions while preventing degradation in system performance and reliability.

IV. Proposed Hybrid Deep Learning Scheduling Framework

The proposed system consists of five main modules:

1. Workload Monitoring Module

Collects real-time resource metrics from cloud nodes.

2. Feature Extraction Module

A CNN model extracts workload features from historical data.

3. Workload Prediction Module

A hybrid architecture involving CNNs for the extraction of spatial features and BiLSTM networks is employed for the temporal prediction workloads.

CNN Feature Extraction

The CNN layer learns the spatial correlations between workload parameters.

The convolution operation is defined as:

$$C_{i,j} = \sum_{k=1}^{K} X_{i+k,j} \cdot W_k + b$$

where:

- X = input feature matrix
- W_k = convolution kernel
- b = bias term
- K = kernel size

The activation function used is ReLU:

$$\mathrm{ReLU}(x) = \max(0, x)$$

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

The CNN layer creates a feature map representing the workload patterns across nodes.

Intelligent Scheduler

The output from CNN layer is passed to BiLSTM network to learn the temporal dependencies among workload data.

The LSTM cell operations are defined as follows:

Forget Gate

$$f_t - \sigma(W_f[h_{t-1}, x_t] + b_f)$$

Input Gate

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

Candidate Cell State

$$\bar{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

Cell State Update

$$C_t - f_t C_{t-1} + i_t \overline{C_t}$$

Output Gate

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

Hidden State

$$h_t - o_t \tanh(C_t)$$

This bidirectional architecture allows the network to process its sequences in a forward and backward direction, giving the model ability to capture long term dependencies of workloads.

The final workload prediction is expressed as:

$$\hat{W}_{t+1} - f(W_t, W_{t-1}, \dots, W_{t-n})$$

where $\hat{W}_{t+1}$ represents the predicted workload demand for the next time interval.

D. Scheduling Decision Engine

The scheduling decision engine uses predicted workloads to assign tasks to cloud resources.

Let:

$$T - \{t_1, t_2, \dots, t_m\}$$

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

Represent the set of incoming tasks.
Let

$$R = \{r_1, r_2, \dots, r_n\}$$

Represent available resources.
The scheduling objective is to minimize execution time and maximize resource utilization:

$$\text{Minimize} \sum_{i=1}^{m} \text{ExecutionTime}\,(t_i)$$

Subject to:

$$CPU_i \leq CPU_{max}$$
$$MEM_i \leq MEM_{max}$$
$$\text{Latency}_i \leq \text{Latencythreshold}$$

The scheduler chooses the resource that satisfies those constraints while minimizing execution cost.
Resource Allocation Module
After scheduling decisions, tasks are selected and deployed in cloud node.
This allows the system to redistribute work dynamically if there is a workload imbalance, so that all nodes are utilized in a balanced way.
The load balancing factor is defined as:

$$LBF - \frac{1}{n}\sum_{i=1}^{n} (U_i - \bar{U})^2$$

where

- U_i = utilization of node i
- $\bar{U}$ = average utilization across nodes.

Lower LBF values indicate better load distribution.
Optimizes workload distribution across multi-cloud infrastructures.
Algorithm: Hybrid CNN–BiLSTM Workload Scheduler

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas

Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)

ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

Input:
  Workload dataset W
  Cloud resource set R
  Task queue T
Output:
  Optimal task-resource allocation
Step 1: Initialize workload monitoring module
Step 2: Collect workload parameters from cloud nodes
Step 3: Pre-process dataset
    Normalize features
    Handle missing values
    Extract workload features
Step 4: Train CNN model
    Input workload features
    Extract spatial feature maps
Step 5: Feed CNN output to BiLSTM model
    Learn temporal workload patterns
    Predict future workload demand
Step 6: Receive incoming task queue T
Step 7: For each task ti in T
      Estimate required resources
      Identify candidate nodes
Step 8: Evaluate node utilization and latency
Step 9: Assign task ti to node rj
    where predicted workload is minimal and resource constraints are satisfied
Step 10: Update resource utilization metrics
Step 11: If node overload detected redistribute tasks across nodes
Step 12: Repeat scheduling process
Return optimized workload allocation

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas

Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

The two-stage architecture of the hybrid CNN-BiLSTM model allows the model to encapsulate not only workload patterns in spatial dimensions, being able to detect variegated resources, but also time-relevant features.

V. Experimental Results

Dataset Description

The cloud workload dataset is derived from a highly-regarded, publicly available set of Google Cluster traces along with models for simulated workload generation.

Torch Bench (dataset): The dataset includes thousands of workload instances from the distributed system in various computing environments.

Key dataset characteristics include:

Parameter	Description
Number of Tasks	50,000+
Number of Cloud Nodes	100
Monitoring Interval	5 seconds
Features	CPU usage, memory usage, task arrival rate, network latency

The dataset represents the dynamic workload behavior over time, so that the model can specialise in realistic scheduling patterns.

Table 1. Performance Comparison of Scheduling Algorithms

Algorithm	Accuracy (%)	Latency (ms)	Resource Utilization (%)
Round Robin	78.2	220	69
Genetic Algorithm	84.6	180	74
LSTM Scheduler	90.1	140	82
Proposed CNN-BiLSTM	95.7	95	91

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
OPEN ACCESS JOURNALS

Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)

ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

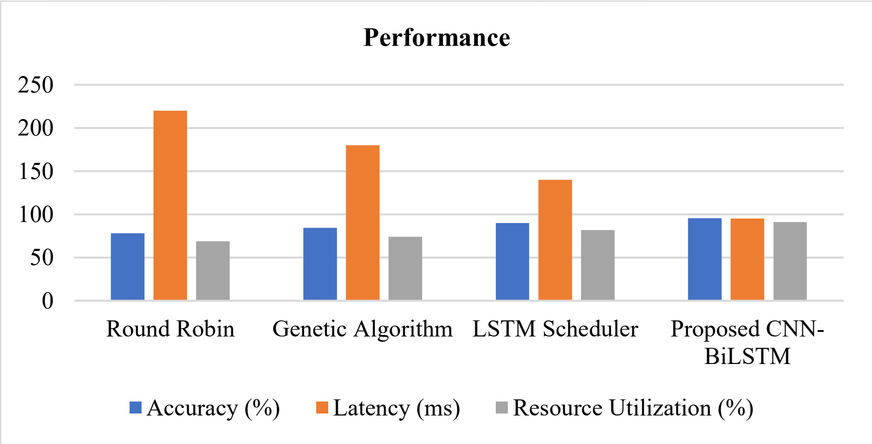

Figure 2. Performance Comparison

Table 2. Task Completion Time Comparison

Algorithm	Avg Task Time (sec)	Throughput
FCFS	6.8	120
GA Scheduler	5.1	165
LSTM	4.2	210
Proposed Model	3.3	268

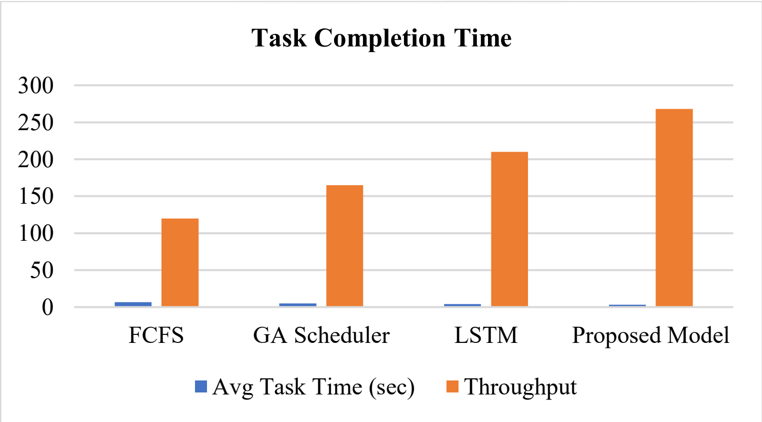

Figure 3. Task Completion Time Comparison

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
OPEN ACCESS JOURNALS
Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

Table 3. Load Balancing Efficiency

Method	Load Imbalance Ratio	Energy Consumption
Round Robin	0.31	280 W
GA	0.24	250 W
LSTM	0.17	210 W
Proposed Model	0.09	180 W

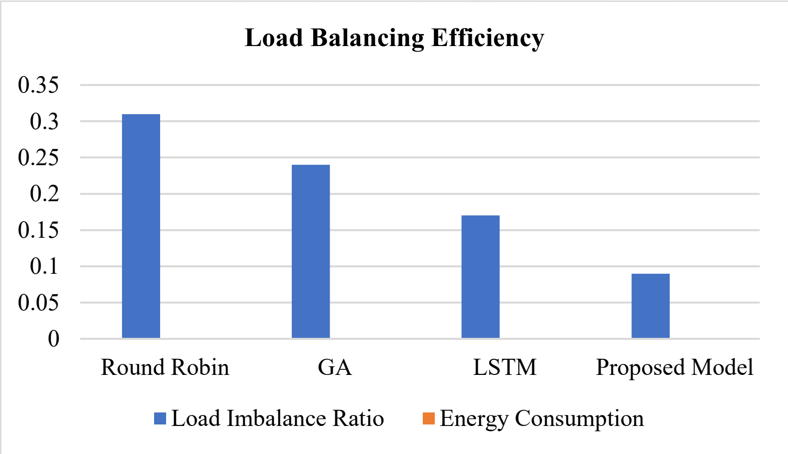

Figure 4. Load Balancing Efficiency

VI. Result Analysis

Experimental results show that our hybrid deep learning scheduler significantly outperforms traditional algorithms.

CNN-BiLSTM outperformed other highest accuracy of 95.7% Compared with previous work, good for workload prediction.) Lower latency and quicker completion times are indicative of better scheduling decisions.

Also, the proposed model had the lowest load imbalance ratio, which means that cloud nodes were used more efficiently. Due to the method of balanced allocation, energy consumption was decreased too because the utilization of resources has been optimized.

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas

Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

The hybrid architecture outperforms optimization-based and general deep learning based schedulers because it models spatial workload features as well as temporal dependencies.

VII. Conclusion and Future Scope

This study proposed an intelligent workload scheduling framework in multi-cloud platforms by applying hybrid deep learning techniques. This paper presents a CNN-BiLSTM based model which predicts the workload efficiently and assigns tasks accordingly on multiple cloud servers. Evaluation based on Infrastructures shows that the system can increase scheduling accuracy, minimize task completion time and improve resource utilization. The results show that hybrid deep learning architectures are able to enhance workload management independently of the environment being multi-cloud or complex.

This work can be extended in future research on multiple dimensions, such as further incorporation of reinforcement learning for dynamic scheduling policies. Real-time multi-cloud environments deployment of the framework Privacy-preserving Cloud Resource Management Using Federated Learning Green Cloud Computable Techniques for energy consumption optimization Use in edge computing architectures for Internet of things (IoT) workloads.

References

[1] S. Sefati, M. Keymasi, R. Craciunescu, S. Maiduc, M. Bayram, and B. Arasteh, “Adaptive resource scheduling in multi-cloud computing using recurrent neural forecasting and memory-based metaheuristic optimization,” Journal of Grid Computing, vol. 23, 2025, doi: 10.1007/s10723-025-09812-7.

[2] S. Simaiya, U. Lilhore, Y. K. Sharma, K. B. V. B. Rao, V. M. Maheswara Rao, A. Baliyan, A. Bijalwan, and R. Alroobaea, “A hybrid cloud load balancing and host utilization prediction method using deep learning and optimization techniques,” Scientific Reports, vol. 14, 2024.

Eureka Open Access Journals

Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas

Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)

ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

[3] S. Chandrasiri and D. Meedeniya, “Energy-efficient dynamic workflow scheduling in cloud environments using deep learning,” Sensors, vol. 25, 2025. Available: https://api.semanticscholar.org/CorpusId:276663708

[4] Y. K. Sanjalawe, S. Fraihat, S. R. Al-E’mari, M. Abualhaj, S. Makhadmeh, and E. Alzubi, “Smart load balancing in cloud computing: Integrating feature selection with advanced deep learning models,” PLOS One, vol. 20, 2025. Available: https://api.semanticscholar.org/CorpusId:281240794

[5] S. Muniswamy and R. Vignesh, “DSTS: A hybrid optimal and deep learning for dynamic scalable task scheduling on container cloud environment,” Journal of Cloud Computing, vol. 11, 2022. Available: https://api.semanticscholar.org/CorpusId:251934929

[6] N. Tripura, P. Divya, K. R. Chaganti, K. Venkateswara, R. P. Rajyalakshmi, and P. Naresh, “Self-optimizing distributed cloud computing with dynamic neural resource allocation and fault-tolerant multi-agent systems,” in Proc. 2024 4th Int. Conf. Ubiquitous Computing and Intelligent Information Systems (ICUIS), 2024, pp. 1304–1310. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10866891

[7] L. R. Raju, M. Reddy, S. R. Surukanti, G. Sudhakar, V. V. S. S. M., and A. Adepu, “IntelliScheduler: An edge-cloud computing environment hybrid deep learning framework for task scheduling based on learning,” Scientific Reports, 2026. Available: https://www.ncbi.nlm.nih.gov/pubmed/41760833

[8] W. Su, Z. Mao, F. Kong, Y. Shi, and P. Xiao, “ESBO-FDDPG: A federated reinforcement learning-based continuous resource scheduler for cross-domain multi-cloud systems,” Informatica, 2026, doi: 10.31449/inf.v50i7.9606.

[9] A. Heirati, M. Fartash, and M. K. Dermany, “Optimized task scheduling in fog-cloud computing using hybrid deep learning and metaheuristic algorithms,” Neural Processing Letters, vol. 58, 2025, doi: 10.1007/s11063-025-11819-w.

[10] A. Kakkad, C. Shinadiya, M. Singh, and N. S. Kumar, “Intelligent auto-scaling in hybrid and multi-cloud architectures using reinforcement learning,” in

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

Proc. 2025 7th Int. Conf. Computational Intelligence and Communication Technologies (CCICT), 2025, pp. 776–781. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=11088023
[11] A. Gupta and S. K. Oruganti, "Multi-agent AI systems for autonomous and context-aware data orchestration in hybrid cloud platforms," International Journal of Data Informatics and Intelligent Computing, 2026, doi: 10.59461/ijdiic.v5i1.260.
[12] J. Bi, H. Yuan, and M. Zhou, "Integrated deep learning method for workload and resource prediction in cloud systems," Neurocomputing, vol. 424, pp. 35–48, 2020, doi: 10.1016/j.neucom.2020.11.011.
[13] M. Shiva, R. Krishna, and D. K. Vali, "Meta-RHDC: Meta reinforcement learning driven hybrid lyrebird falcon optimization for dynamic load balancing in cloud computing," IEEE Access, vol. 13, pp. 36550–36574, 2025. Available: https://api.semanticscholar.org/CorpusId:276594393
[14] F. Jiang, K. Wang, L. Dong, C. Pan, W. Xu, and K. Yang, "Deep-learning-based joint resource scheduling algorithms for hybrid MEC networks," IEEE Internet of Things Journal, vol. 7, pp. 6252–6265, 2019. Available: https://api.semanticscholar.org/CorpusId:208202363
[15] D. Bodra and S. Khairnar, "Machine learning-based cloud resource allocation algorithms: A comprehensive comparative review," arXiv preprint arXiv:2511.11603, 2025. Available: https://api.semanticscholar.org/CorpusId:282285012
[16] N. R. Oruganti, "Optimizing resource allocation for deep learning workloads in heterogeneous cloud environments," International Journal for Multidisciplinary Research, 2024. Available: https://api.semanticscholar.org/CorpusId:274442097
[17] Y. Yu, "Towards intelligent cloud scheduling: DynaSched-Net with reinforcement learning and predictive modeling," in Proc. 2025 10th Int. Conf.

ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

Intelligent Computing and Signal Processing (ICSP), 2025, pp. 807–811. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=11086764
[18] K. Tajane, R. Pitale, A. Shinde, S. Sinalkar, S. Shende, and A. Singh, "Multi-objective scheduling in multi-clouds: A DRL-based framework for SLA, privacy, and cost optimization," in Proc. 2025 IEEE 6th Global Conf. Advancement in Technology (GCAT), 2025, pp. 1–6. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=11368492
[19] U. Lilhore, S. Alex, V. Paul, R. P. Purayil, S. Aldossary, S. Simaiya, E. Ghith, H. G. Mohamed, and M. Khan, "QHRMOF: A quantum-inspired hybrid multi-objective framework for energy-efficient task scheduling and load balancing in cloud computing," Journal of Cloud Computing, vol. 14, 2025, doi: 10.1186/s13677-025-00777-2.
[20] V. S. Patil, "Intelligent resource allocation and scheduling for cloud environments," International Journal for Research in Applied Science and Engineering Technology, 2025, doi: 10.22214/ijraset.2025.76027.
[21] C. A., "Intelligent resource allocation in multi-cloud environments using deep reinforcement learning," World Journal of Advanced Research and Reviews, 2020, doi: 10.30574/wjarr.2020.8.2.0397.
[22] H. Chaudhary, G. Sharma, D. Nishad, and S. Khalid, "AI-enhanced modelling of queueing and scheduling systems in cloud computing," Discover Applied Sciences, vol. 7, 2025. Available: https://api.semanticscholar.org/CorpusId:277652094
[23] S. Shahhosseini, T. Hu, D. Seo, A. Kanduri, B. Donyanavard, A. M. Rahmani, and N. Dutt, "Hybrid learning for orchestrating deep learning inference in multi-user edge-cloud networks," in Proc. 2022 23rd Int. Symp. Quality Electronic Design (ISQED), 2022, pp. 1–6. Available: https://api.semanticscholar.org/CorpusId:247058793
[24] X. Dou, L. Liu, and L. Xiao, "Is intelligence the right direction in new OS scheduling for multiple resources in cloud environments?," ACM Transactions

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Artificial Intelligence and Data Innovation (EJAIDI)
ISSN 2760-5000 (Online) Volume 2, Issue 3, March 2026

https://eurekaoa.com/index.php/11

on Storage, vol. 22, pp. 1–29, 2025. Available: https://api.semanticscholar.org/CorpusId:277954893

[25] B. L. Vennala, D. Suresh, P. Samiya, M. Suresh, and B. Vamsi, "Integrating artificial intelligence with cloud platforms to optimize performance, scalability, and reliability in distributed computing systems," in Proc. 2025 Int. Conf. Data Science, Agents & Artificial Intelligence (ICDSAAI), 2025, pp. 1–6. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=11011769.