J. Electrical Systems 19-2 (2023): 68-77

[1]Torana Kamble

[2]Dr. Sanjivani Deokar

[3]Vinod S. Wadne

[4]Devendra P. Gadekar

[5]Hrishikesh Bhanudas Vanjari

[6]Dr. Purva Mange

Predictive Resource Allocation Strategies for Cloud Computing Environments Using Machine Learning

Journal of Electrical Systems

Abstract: - Cloud computing revolutionizes fast-changing technology. Companies' computational resource use is changing. Businesses can quickly adapt to changing market conditions and operational needs with cloud-based solutions' adaptability, scalability, and cost-efficiency. IT operations and service delivery have changed due to widespread computational resource access. Cloud computing efficiently allocates resources in cloud environments, making it crucial to this transformation. Resource allocation impacts efficiency, cost, performance, and SLAs. Users and providers can allocate cloud resources based on workloads using elasticity, scalability, and on-demand provisioning. IT economics and operational effectiveness have changed due to rapid and flexible resource allocation. Proactive versus reactive resource allocation is key to understanding cloud resource management challenges and opportunities. Reactive strategies allocate resources only when shortages or surpluses occur at demand. This responsive strategy often leads to inefficiencies like over- or under-allocation, which raises costs and lowers performance. Predictive analysis and workload forecasting predict resource needs in proactive resource allocation. Optimize resource use to avoid shortages and over-provisioning. Attention has been drawn to proactive predictive resource allocation. These methods predict resource needs using historical data, machine learning, and predictive analytics. Predictive strategies optimize resource allocation by considering future decisions. Reduced bottlenecks boost user satisfaction and lower operational costs. Matching resource distribution to workloads optimizes cloud resource management. Resource allocation prediction improves with deep learning. CNN, LSTM, and Transformer cloud resource forecasting algorithms are promising. New tools for accurate and flexible workload predictions have come from their ability to spot intricate patterns in historical data. This paper compares CNN, LSTM, and Transformer deep learning algorithms for cloud computing resource allocation forecasting. This study determines the best predictive accuracy and workload ada[1]ptability algorithm using Google Cluster Data (GCD). The study evaluates upgrading cloud computing resource allocation with the Transformer model. This study advances predictive resource allocation strategies, which can help cloud service providers and organizations improve resource utilization, cost-effectiveness, and performance in the face of rapid technological change.

General Terms: Cloud Computing, Deep Learning

Keywords: Resource Allocation, Cloud Computing, Predictive Strategies, Deep Learning Algorithms, Transformer Model.

I. INTRODUCTION

The emergence of cloud computing in the current dynamic digital environment represents a groundbreaking achievement that fundamentally transforms the methods by which organizations oversee and utilize computational resources. The shift from traditional, physical infrastructure to cloud-based solutions has provided businesses and enterprises with an exceptional level of adaptability, expandability, and cost-effectiveness. The advent of cloud computing has brought about an era characterized by the widespread availability of computational resources, leading

[1]Assistant Professor, Bharati Vidyapeeth College of Engineering, Navi Mumbai, Maharashtra, India, torana.kamble@gmail.com

[2]Department of Computer Engineering, Lokmanya Tilak College of Engineering, Mumbai University, Maharashtra, India, sanjivanideokar@gmail.com

[3]JSPM's Imperial College of Engineering and Research, Pune, Maharashtra, India, vinods1111@gmail.com

[4]Mastercard Technology, Pune, Maharashtra, India, devendrapgadekar@gmail.com

[5]Department of Electronics & Telecommunication, Bharati Vidyapeeth College of engineering, Lavale, Pune, Maharastra, India, hrishikesh@outlook.in

[6]Associate Professor, Symbiosis School of Planning Architecture and Design, Symbiosis International University, Pune, Maharashtra, India, purva.mange@gmail.com

*Correspondence: vinods1111@gmail.com

Copyright © JES 2023 on-line : journal.esrgroups.org

to increased reliance on data-driven decision-making, advancements in application development, and a constant drive for resource optimization. Cloud computing holds a crucial role in modern technological strategies, fundamentally reshaping the principles of IT infrastructure and service provision [1], [2].

The transformative power of cloud computing is rooted in its exceptional ability to effectively manage and distribute resources. Efficient resource allocation in cloud environments is not just operationally convenient, but also crucial for controlling costs, ensuring strong performance, and meeting strict service level agreements. The cloud's inherent flexibility, ability to scale, and capability to provision resources on demand provide cloud service providers and users with a dynamic set of tools for allocating resources as workloads fluctuate. The capacity to allocate resources rapidly and smoothly has brought about a new level of operational effectiveness, fundamentally reshaping the economics of IT. Consequently, the distribution of resources in cloud environments has become a significant problem, requiring the creation of sophisticated resource allocation techniques to adapt to the constantly changing and dynamic cloud environment [3], [4].

Cloud computing has revolutionized the way organizations handle and utilize computational resources in the current digital environment. The transition from conventional, physically located infrastructure to cloud-based solutions has provided unparalleled adaptability, scalability, and cost-efficiency, empowering businesses to adjust to swiftly evolving demands and market dynamics. The widespread availability of computational resources has caused a significant change in the efficiency and flexibility of IT operations, leading to a paradigm shift. Cloud computing has significantly accelerated the era of making decisions based on data, fostering the development of new applications, and optimizing resource utilization. It has become a crucial component in modern technology strategies [5], [6].

Allocating resources in cloud computing environments is a complex and multi-dimensional challenge. It requires distributing computing resources, such as “CPU, memory, storage, and network bandwidth” to fulfill the needs of running applications, services, and various workloads. The complexity of this process ranges from the dynamic and unpredictable nature of workloads in cloud environments. The issue of resource allocation has been addressed using reactive methods, which involve reacting to fluctuations in resource demand as they arise. This strategy results in inefficiencies, such as an “excessive allocation” or “an insufficient allocation”, which can lead to increased expenses and reduction in performance.

The major point in understanding the benefits and drawbacks of cloud resource management is understanding the distinction between proactive and reactive approaches in allocating resources in cloud computing environment. In reactive strategies, the distribution of resources is based on the current demand and adjustments are made only when there is a shortage. To determine appropriate resource allocation, proactive strategies examine historical data and make predictions about the future demands. Where as in proactive resource allocation try to minimize both under and over demand of resources so that they can be used to their full potential while keeping operational costs minimum. In recent time proactive allocation methods have become important area of study in the field of cloud computing for better resource allocations.

In cloud computing predictive resource allocation strategies are playing crucial role. Predictive resource allocation strategies use historical data, deep learning models, and predictive analytics to forecast future resource allocation requirements. Main aim of predictive strategies is to optimize resource allocation by making informed decisions in advance. The main goal is to ensure the proper availability of required amount of resources when required. This helps to prevent performance blockages, reduce operational costs and improve user satisfaction.

DL models are used in predictive resource allocation strategies. DL algorithms like Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Transformer model have shown promising advancement in various domains [7].

Many research has investigated different facets of forecasting resource allocation in cloud computing. These studies have evaluated the efficacy of various algorithms and strategies, providing insights into their advantages and limitations. Understanding these findings is essential for expanding on current understanding and creating more efficient methods for distributing resources. Achieving optimal efficiency, cost-effectiveness, and better performance in cloud computing environments remains difficult despite advancements in predictive resource allocation [8] – [11].

The main objective of the presented work is

- To evaluate the efficiency of DL algorithms like CNN, LSTM, and Transformer model for resource allocation in cloud computing environments efficiently.
- The study's goal is to learn more about the algorithm that shows the best predictive accuracy and flexibility under changing workload conditions.
- This study specifically evaluates predictive resource allocation strategies, using DL algorithms using Google Cluster Data (GCD) dataset.
- The study assesses the effectiveness of these algorithms for improving resource allocation in cloud computing environments.

II. LITERATURE REVIEW

The literature review offers a thorough synopsis of research papers that contribute to the comprehension of resource allocation strategies in cloud computing environments. This text emphasizes the importance of effective resource management in the cloud and the different methods, frameworks, and models suggested to tackle the difficulties in this ever-changing field.

S. R. Swain et al. [12] examine the effective allocation and utilization of resources in cloud environments. The authors analyze different resource allocation strategies with the goal of maximizing resource utilization for cost-efficiency and high-performance cloud services. The study focuses on the details of managing cloud resources. M. Asad Arfeen et al. [13] presented a comprehensive framework for resource allocation strategies in cloud computing environments. The proposed approach is to tackle the complex challenges related to resource allocation and management in the cloud environment. S. Chouliaras et al. [14] introduced an adaptive auto-scaling framework for cloud resource, with a focus on the dynamic allocation of cloud workloads. The framework allocate resources based on fluctuating workload patterns and hence improved the efficiency of cloud resource management and ensuring optimal performance under various level of demand.

S. A. Murad et al. [15] conducted an analysis of job scheduling methods in cloud computing. It presents a sophisticated framework that utilizes priority rules to schedule jobs and allocate resources. The study explores the complexities of job scheduling in cloud environments and highlights the importance of intelligent frameworks in optimizing resource allocation. K. Raghavendar et al. [16] focus on optimizing the allocation of resources in cloud-based Internet of Things (IoT) environments. More precisely, it tackles difficulties associated with uneven distribution of data and the speed at which data is consumed. The study presents a resilient resource allocation model that provides solutions for cloud-based IoT applications that require significant resources. Lin et al [17]. propose a dynamic resource allocation scheme for cloud computing environments, which is based on threshold values. The approach emphasizes the significance of adapting resource allocation in response to changing workload demands. The study seeks to improve the efficiency of resource provisioning in cloud environments by implementing threshold-based strategies.

Z. Xiao et al. [18] presents a novel method for allocating resources in cloud computing environments by utilizing virtual machines. The study highlights the adaptability and effectiveness of resource allocation through virtualization, with the goal of maximizing resource usage while maintaining strong performance. G. Suvarna Kumar et al. [19] examine the optimization of Customer Relationship Management (CRM) services in cloud computing through the implementation of service scheduling and task allocation techniques. The study emphasizes the significance of efficient resource allocation in enhancing customer experiences and the quality of CRM services in cloud environments. Mahenge et al. [20] examined the concept of energy-efficient task offloading in mobile edge computing. They specifically concentrated on enhancing resource allocation for mobile applications that require a significant amount of resources. The study investigates methods for optimizing resource allocation in mobile edge environments.

Belgacem et al. [21] present an advanced multi-agent reinforcement learning model designed for efficient resource allocation in cloud computing environments. The study highlights the importance of intelligent agents in improving resource allocation, providing valuable insights into novel methods for enhancing resource management. G. Marques et al. [22] explore proactive resource management strategies in cloud-of-services environments. It highlights the significance of proactive strategies for allocating and managing resources, with the goal of improving the effectiveness and performance of cloud services through forward-looking resource management. G. Turin et al.

[23] examine the forecasting of resource utilization in Kubernetes container systems. The study presents resource models as predictive instruments for optimizing resource allocation in containerized environments, providing solutions to improve efficiency and performance.

III. DATA COLLECTION AND PREPROCESSING

The proposed system architecture is shown in Figure 1. The architecture follows several steps for predicting resources allocation which include, gathering data, data cleaning, feature selection, and applying ML, DL and Transformer models.

3.1 Dataset

Workload traces from Google's production cluster are collected and made available to the public in the form of the Google Cluster Data (GCD) dataset. It includes information on the utilization of the central processing unit (CPU), memory, and disk space for a variety of workloads, such as web search, machine learning, and data analytics [24].

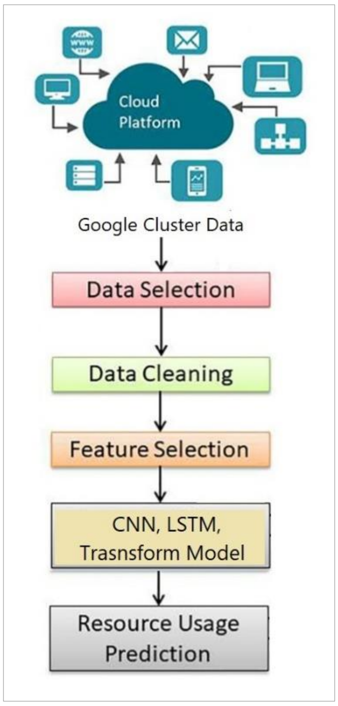

Figure 1: Proposed System Architecture

3.2 Data pre-processing

Data cleaning:

Data cleaning ensure cleaning duplicate entries or outliers, that are dealt with to ensure the dataset quality and uniformity.

3.2.1 Data Transformation:

Data Transformation is the subsequent phase, wherein data is prepared to be more amenable for analysis. This involves converting categorical variables, such as job types or resource categories, into numerical representations. Converting the variable “job Type” into numerical values enables its incorporation into predictive models. It may be necessary to aggregate data over various time intervals, such as hourly or daily, in order to gain insights into resource trends. Converting “Resource Usage” data into daily averages allows for a more comprehensive understanding of resource consumption patterns.

3.2.2 Feature Selection:

Feature selection is a crucial aspect of the preprocessing process. The process entails identifying and preserving the most pertinent characteristics for predictive modeling. When it comes to GCD, it is crucial to choose features that directly affect the allocation of resources. Examples of highly relevant features include “CPU Utilization” and “Memory Usage” since they directly impact the allocation of computing resources. By prioritizing these characteristics, the analysis becomes more efficient and in line with the research goals.

3.2.3 Data normalization:

Data normalization is crucial for standardizing various features to a uniform scale. This is especially important in the context of cloud computing; as different resource metrics can vary greatly in their ranges. For example, “Network Bandwidth” and “CPU Utilization” are normalized to prevent employing excessive effect on the predictive model compared to the other. Normalizing data ensures that all relevant attributes are given fair consideration when predicting resource allocation.

IV. DEEP LEARNING ALGORITHM

4.1 CNN

Convolutional Neural Networks (CNNs) have a distinct function in the field of Predictive Resource Allocation Strategies for Cloud Computing Environments. CNNs are widely recognized for their proficiency in image processing and computer vision applications. However, they have also demonstrated their utility in cloud computing by effectively analyzing historical data on resource usage. The utilization of Convolutional Neural Networks (CNNs) proves advantageous in analyzing data, particularly when it is presented in the form of time series or multi-dimensional information, due to their ability to recognize patterns. Within cloud environments, these neural networks are capable of effectively recognizing repetitive resource usage patterns, conducting automatic feature extraction, and identifying anomalies. Hence, Convolutional Neural Networks (CNNs) play a crucial role in extracting valuable knowledge from historical resource data, thereby facilitating the forecasting of future resource requirements.

4.2 LSTM

LSTM networks play a crucial role in predicting resource allocation in cloud computing environments, especially for workloads that are time-dependent. The significance of LSTM's relevance lies in its adeptness at capturing sequential dependencies in resource usage data. LSTMs are particularly effective in situations where it is crucial to forecast future resource needs, as they are able to accurately capture the temporal connections between different data points related to resources. They possess an exceptional ability to recall previous conditions, enabling them to incorporate extended connections in workload patterns. LSTM models demonstrate flexibility by accommodating variations in workload characteristics, thereby improving the accuracy and adaptability of predictions for resource allocation.

4.3 Transformer Model (BERT)

The Transformer model BERT [7], originally prominent in natural language processing tasks, has also emerged as a valuable tool for sequence-to-sequence modeling, which has direct implications for predicting resource allocation in cloud computing. The Transformer's proficiency in processing sequence data, combined with its self-attention mechanism, makes it highly suitable for comprehending temporal dependencies in resource allocation. Moreover, the model's ability to process multiple tasks simultaneously and its capacity to adjust effectively to different lengths of sequences in historical resource data make it highly adaptable and efficient. The Transformer model's attributes make it an appealing option for predicting resource allocation, as it considers the sequence of past resource usage data and its associated features.

V. EVALUATION PARAMETERS

Mean Absolute Error (MAE):

MAE is a direct measurement that computes the average absolute deviation between the predicted and actual values. MAE measure precisely and conclusively to prediction model performs where lower MAE values represents higher model performance. Eq.1 represents MAE.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \quad \dots 1$$

Where n= “data points”, $\hat{y}_i$= “predicted values”, y_i= “actual values”.

Root Mean Squared Error (RMSE):

RMSE calculates the square root of the mean of the squared differences between predicted and actual values. RMSE is more accurate to predicting errors as compared to MAE. This makes it a suitable tool for detecting outliers or significant discrepancies. Eq.2 represents RMSE.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad \ldots 2$$

Where n= “data points”, $\hat{y}_i$= “predicted values”, y_i= “actual values”.

Normalized Mean Absolute Error (NMAE):

NMAE measures the average absolute difference between predicted and actual values which is normalized to a specific range. It is derived from the MAE metric. The NMAE provides a comparative assessment of the accuracy of predictions, making it appropriate for comparing models across various. Eq.3 represents NMAE.

$$NMAE = \frac{MAE}{\max(y) - \max(x)} \quad \ldots 3$$

Normalized Root Mean Squared Error (NRMSE):

NRMSE is a normalized version of the Root Mean Squared Error (RMSE), similar to the NMAE. The method divides the RMSE by the range of actual values, resulting in a relative measure of prediction accuracy that can be compared between different datasets. Eq.4 represents NRMSE.

$$NRMSE = \frac{RMSE}{\max(y) - \max(x)} \quad \ldots 4$$

Hit Rate (%):

The represents the ratio of accurate predictions made by the model within the framework of resource allocation. This can act as a metric for evaluating the accuracy of the model in predicting resource requirements. Eq.5 represents Hit Ratio.

$$Hit\ Rate = \frac{No.of\ correct\ predictions}{Total\ Predictions} \times 100\% \quad \ldots 5$$

False Alarm Rate (%):

It measures the ratio of incorrect predictions made when the actual outcome was negative. In resource allocation, this refers to situations where resources were allocated unnecessarily. Eq.6 represents False alarm rate.

$$False\ Alarm\ Rate = \frac{No.of\ False\ Alarm}{Total\ Actual\ Negatives} \times 100\% \quad \ldots 6$$

Miss Rate (%):

This measures the proportion of predictions that were inaccurate when the actual result was positive. This denotes situations where resources were not allocated at the required time. Eq.7 represents Miss rate.

$$Miss\ Rate = \frac{No.of\ Misses}{Total\ Actual\ Positives} \times 100\% \quad \ldots 7$$

Resource Utilization (%):

Resource Utilization measures the effectiveness of resource allocation. It computes the proportion of resources that were efficiently utilized from the overall allocated resources. High resource utilization indicates effective allocation, whereas low utilization may suggest inefficiency and wastage. Eq.8 represents Resource utilization.

$$Resource\ Utilization = \frac{Used\ resources}{Allocated\ Resources} \times 100\% \quad \ldots 8$$

VI. RESULTS AND OUTPUTS

Table 1. Evaluation parameters comparison

Algorithm	MAE	RMSE	NMAE	NRMSE
CNN	0.13	0.23	0.1	0.13
LSTM	0.09	0.15	0.07	0.1
Transformer	0.03	0.1	0.03	0.07

Table 2. Evaluation Metrics

Algorithm	Hit Rate (%)	False Alarm Rate (%)	Miss Rate (%)	Resource Utilization (%)
CNN	89	5	5	88
LSTM	94	2.5	2.5	93
Transformer	99	0.5	0.5	96

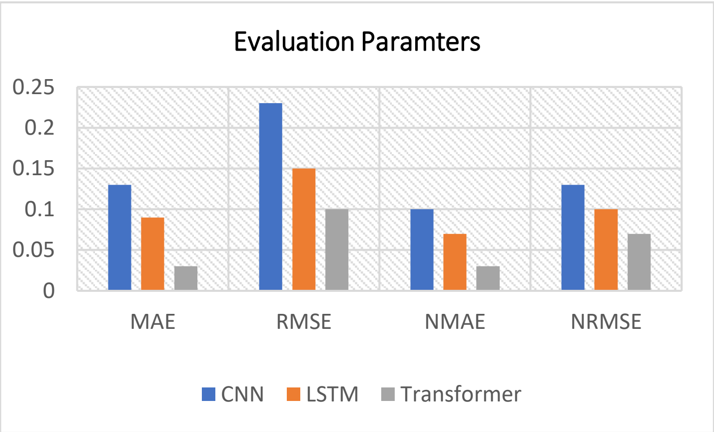

Figure 2: Evaluation parameters comparison graph

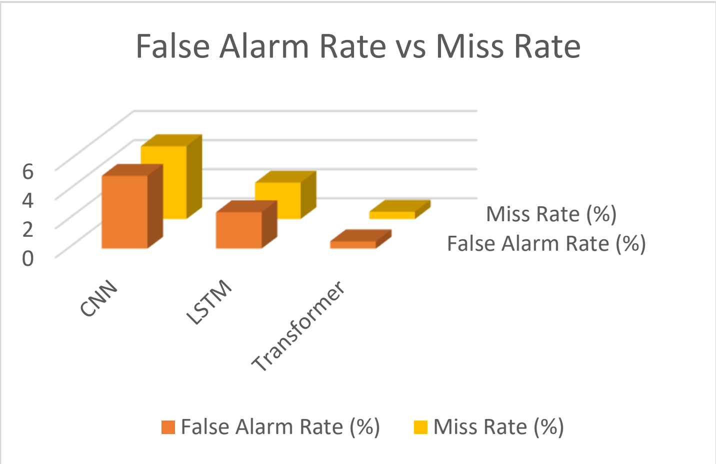

Figure 3: False alarm rate vs Miss rate

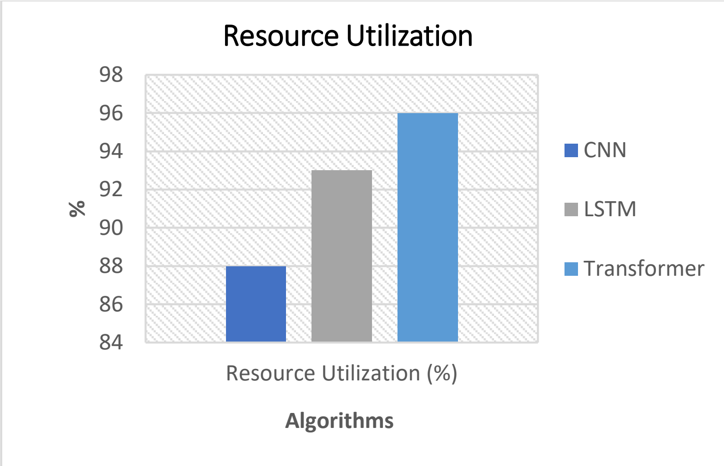

Figure 4: Resource utilization comparison graph

The result summary evaluates CNN, LSTM, and Transformer in cloud computing resource allocation prediction are shown in table-1,2 and figure 2-4. CNN makes good predictions with an MAE of 0.13 and an RMSE of 0.23. The NMAE and NRMSE scores of 0.1 and 0.13 indicate high relative accuracy. It has an 89% Hit Rate, indicating many accurate predictions. However, it has 5% False Alarm and Miss Rates. Resource Utilization is 88%, indicating efficient resource allocation.

The LSTM model has a 0.09 MAE and 0.15 RMSE, indicating a slight prediction accuracy improvement. The NMAE and NRMSE values of 0.07 and 0.1 indicate a higher level of relative accuracy. It attains a remarkable 94% accuracy in identifying targets, while maintaining a low 2.5% rate of false alarms and missed detections. The 93% resource utilization rate suggests good resource allocation.

With an MAE of 0.03 and RMSE of 0.1, the Transformer algorithm is highly accurate. Excellent relative accuracy is shown by the NMAE and NRMSE values of 0.03 and 0.07. The Hit Rate is 99%, with a 0.5% False Alarm Rate and Miss Rate. High Resource Utilization of 96%, indicating efficient resource allocation.

Transformer outperforms LSTM in predictive resource allocation accuracy and efficiency. CNN is adequate, but the Transformer model is better for cloud computing resource allocation. The Transformer model has low prediction errors and high resource utilization.

VII. CONCLUSION AND FUTURE SCOPE

The assessment of three deep learning algorithms—CNN, LSTM, and Transformer—in the context of predictive resource allocation for cloud computing environments has yielded valuable insights regarding their performance and efficacy. The transformer has exhibited remarkable precision and effectiveness in forecasting resource requirements, boasting remarkably low Mean Absolute Error (MAE) and a high Hit Rate. The LSTM model also demonstrated commendable performance, exhibiting enhanced accuracy in prediction and efficient utilization of resources. Although CNN achieved commendable outcomes, Transformer emerges as the foremost selection for enhancing resource allocation in cloud computing, establishing a standard for predictive models in this field. These findings highlight the importance of using advanced deep learning techniques to improve resource allocation strategies, which in turn leads to increased efficiency and reliability in cloud computing environments. Future research in predictive resource allocation for cloud computing will involve the creation of hybrid models, implementation of real-time adaptive strategies, and emphasis on energy-efficient resource allocation. Additionally, there will be an expansion of these strategies to edge and fog computing environments to improve resource management. Furthermore, it is crucial to prioritize the resolution of security and privacy issues and implement automated resource allocation using AutoML techniques in order to significantly influence the development of the field.

REFERENCES

[1] T. Khan, W. Tian, G. Zhou, S. Ilager, M. Gong, and R. Buyya, "Machine learning (ML)-centric resource management in cloud computing: A review and future directions," J. Netw. Comput. Appl., vol. 204, 2022

[2] K. Kumaran and E. Sasikala, “Computational access point selection based on resource allocation optimization to reduce the edge computing latency,” Meas. Sensors, vol. 24, no. September, p. 100444, 2022

[3] P. Pradhan, P. K. Behera, and B. N. B. Ray, “Modified Round Robin Algorithm for Resource Allocation in Cloud Computing,” Procedia Comput. Sci., vol. 85, no. Cms, pp. 878–890, 2016

[4] Z. Sharif, L. Tang Jung, M. Ayaz, M. Yahya, and S. Pitafi, “Priority-based task scheduling and resource allocation in edge computing for health monitoring system,” J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 2, pp. 544–559, 2023, doi: 10.1016/j.jksuci.2023.01.001.

[5] P. Wei, Y. Zeng, B. Yan, J. Zhou, and E. Nikougoftar, “VMP-A3C: Virtual machines placement in cloud computing based on asynchronous advantage actor-critic algorithm,” J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 5, p. 101549, 2023, doi: 10.1016/j.jksuci.2023.04.002.

[6] X. Xiao, M. Zhao, and Y. Zhu, “Multi-stage resource-aware congestion control algorithm in edge computing environment,” Energy Reports, vol. 8, pp. 6321–6331, 2022, doi: 10.1016/j.egyr.2022.04.078.

[7] V. Khetani, Y. Gandhi, S. Bhattacharya, S. N. Ajani, and S. Limkar, “Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse Domains,” Int. J. Intell. Syst. Appl. Eng., vol. 11, pp. 253–262, 2023.

[8] A. Sarah, G. Nencioni, and M. M. I. Khan, “Resource Allocation in Multi-access Edge Computing for 5G-and-beyond networks,” Comput. Networks, vol. 227, no. 308909, p. 109720, 2023

[9] T. Thein, M. M. Myo, S. Parvin, and A. Gawanmeh, “Reinforcement learning based methodology for energy-efficient resource allocation in cloud data centers,” J. King Saud Univ. - Comput. Inf. Sci., vol. 32, no. 10, pp. 1127–1139, 2020, doi: 10.1016/j.jksuci.2018.11.005.

[10] Y. Xie, C. Allen, and M. Ali, “Critical success factor based resource allocation in ERP implementation: A nonlinear programming model,” Heliyon, vol. 8, no. 8, p. e10044, 2022

[11] H. Kaur and A. Anand, “Review and analysis of secure energy efficient resource optimization approaches for virtual machine migration in cloud computing,” Meas. Sensors, vol. 24, no. September, p. 100504, 2022

[12] S. R. Swain, A. K. Singh, and C. N. Lee, “Efficient Resource Management in Cloud Environment,” pp. 1–9, 2022, [Online].

[13] M. Asad Arfeen, K. Pawlikowski, and A. Willig, “A framework for resource allocation strategies in cloud computing environment,” Proc. - Int. Comput. Softw. Appl. Conf., pp. 261–266, 2011, doi: 10.1109/COMPSACW.2011.52.

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

[15] S. A. Murad, A. J. M. Muzahid, Z. R. M. Azmi, M. I. Hoque, and M. Kowsher, “A review on job scheduling technique in cloud computing and priority rule based intelligent framework,” J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 6, pp. 2309–2331, 2022, doi: 10.1016/j.jksuci.2022.03.027.

[16] K. Raghavendar, I. Batra, and A. Malik, “A robust resource allocation model for optimizing data skew and consumption rate in cloud-based IoT environments,” Decis. Anal. J., vol. 7, no. February, p. 100200, 2023

[17] W. Lin, J. Z. Wang, C. Liang, and D. Qi, “A threshold-based dynamic resource allocation scheme for cloud computing,” Procedia Eng., vol. 23, pp. 695–703, 2011, doi: 10.1016/j.proeng.2011.11.2568.

[18] Z. Xiao, W. Song, and Q. Chen, “Dynamic resource allocation using virtual machines for cloud computing environment,” IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 6, pp. 1107–1117, 2013

[19] Virendra Swaroop Sangtani, Bharat Bhushan Jain, Nandkishor Gupta, & Ashish Raj. (2022). Design Simulation and Perfromance Assessment of PMSM Based Motor Charactersitics for Electric Vehicle Applications. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(3), 96–104. https://doi.org/10.17762/ijrmee.v9i3.379

[20] G. Suvarna Kumar, R. Priyadarshini, N. H. Parmenas, H. Tannady, F. Rabbi, and A. Andiyan, “Design of Optimal Service Scheduling based Task Allocation for Improving CRM in Cloud Computing,” 6th Int. Conf. I-SMAC 2022 - Proc., pp. 438–445, 2022

[21] Rai, S. K. ., Rana, D. P. ., & Kashif, D. M. . (2022). Hotel Personnel Retention In Uttar Pradesh: A Study of HYATT Hotels. International Journal of New Practices in Management and Engineering, 11(01), 47–52. https://doi.org/10.17762/ijnpme.v11i01.173

[22] M. P. J. Mahenge, C. Li, and C. A. Sanga, “Energy-efficient task offloading strategy in mobile edge computing for resource-intensive mobile applications,” Digit. Commun. Networks, vol. 8, no. 6, pp. 1048–1058, 2022

[23] A. Belgacem, S. Mahmoudi, and M. Kihl, “Intelligent multi-agent reinforcement learning model for resources allocation in cloud computing,” J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 6, pp. 2391–2404, 2022, doi: 10.1016/j.jksuci.2022.03.016.

[24] G. Marques, C. Senna, S. Sargento, L. Carvalho, L. Pereira, and R. Matos, “Proactive resource management for cloud of services environments,” Futur. Gener. Comput. Syst., vol. 150, pp. 90–102, 2024

[25] G. Turin, A. Borgarelli, S. Donetti, F. Damiani, E. B. Johnsen, and S. L. Tapia Tarifa, "Predicting resource consumption of Kubernetes container systems using resource models," J. Syst. Softw., vol. 203, p. 111750, 2023

[26] Google, "GitHub - google/cluster-data: Borg cluster traces from Google." 2019, [Online]. Available: https://github.com/google/cluster-data.

[27] Lachouri, A., & Ardjouni, A. (2022). Aeroelastic Stability of Combined Plunge-Pitch Mode Shapes in a Linear Compressor Cascade. Advances in the Theory of Nonlinear Analysis and Its Applications, 6(1), 101–117.

[28] Panwar, A., Morwal, R., & Kumar, S. (2022). Fixed points of ρ-nonexpansive mappings using MP iterative process. Advances in the Theory of Nonlinear Analysis and Its Applications, 6(2), 229–245.

[29] Regularization method for the problem of determining the source function using integral conditions. (2021). Advances in the Theory of Nonlinear Analysis and Its Application, 5(3), 351-361. https://atnaea.org/index.php/journal/article/view/208

[30] Saurabh Bhattacharya, Manju Pandey,"Deploying an energy efficient, secure & high-speed sidechain-based TinyML model for soil quality monitoring and management in agriculture", Expert Systems with Applications, Volume 242, 2024, 122735, ISSN 0957-4174.https://doi.org/10.1016/j.eswa.2023.122735.

[31] Khetani, V., Gandhi, Y., Bhattacharya, S., Ajani, S. N., & Limkar, S. (2023). Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse Domains. International Journal of Intelligent Systems and Applications in Engineering, 11(7s), 253-262.

[32] Boutebba, H., Lakhal, H., Slimani, K., & Belhadi, T. (2023). The nontrivial solutions for nonlinear fractional Schrödinger-Poisson system involving new fractional operator. Advances in the Theory of Nonlinear Analysis and Its Applications, 7(1), 121–132.

[33] Shivadekar, S., Kataria, B., Limkar, S., S. Wagh, K., Lavate, S., & Mulla, R. A. (2023). Design of an efficient multimodal engine for preemption and post-treatment recommendations for skin diseases via a deep learning-based hybrid bioinspired process. Soft Computing, 1-19.

© 2023. This work is published under https://creativecommons.org/licenses/by/4.0/legalcode(the“License”). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.