Journal of Next Generation Technology (ISSN: 2583-021X)
Vol. 5, Issue 6, Nov 2025

Cost-Efficient AI for Distributed Cloud Data Analytics: Resource Optimization and Privacy Compliance

Gullapalli Lasya Sravanthi
Software Engineer, 18606 Alderwood Mall Pkwy, Unit 684, Lynnwood, Washington, 98037, USA.
Email id: glasyasravanthi@gmail.com

Article Received: 2 Oct 2025 *Article Accepted: 18 Nov 2025* *Article Published: 18 Nov 2025*

Citation
Gullapalli Lasya Sravanthi (2025), "Cost-Efficient Ai for Distributed Cloud Data Analytics: Resource Optimization and Privacy Compliance", Journal of Next Generation Technology (ISSN: 2583-021X),5(6), pp. 99-112, Nov 2025.

Abstract

With the dynamic nature of the large-scale cloud computing era, cost-effective, scalable, and privacy-aware data analytics is an essential issue in organizations that process distributed datasets. This paper contends to develop a new Cost-Efficient Artificial Intelligence Framework of Distributed Cloud Data Analytics(CAI-DA) offering a combination of AI-based resources optimization, federated learning, and sophisticated privacy control to maximize computational efficiency and data security. The framework depends on the hybrid Long Short-Term Memory (LSTM) and Reinforcement Learning (RL) framework of smart workload forecast and dynamic resource allocation, using both Differential Privacy (DP) and Homomorphic Encryption (HE) to guarantee privacy regulations. As experimental assessment showed, the suggested CAI-DA framework has attained the maximum cost reduction of 36.7 percent, a 41 percent reduction in latency, and an 82 percent decreased privacy leakage when contrasted with the conventional centralized and non-optimized distributed systems. Results support the claim that CAI-DA manages to balance performance, scalability, and security making it a powerful solution to next-generation distributed cloud analytics.

Keywords: Distributed Cloud Analytics, Resource Optimization, Federated Learning, Differential Privacy, Homomorphic Encryption, Cost Efficiency

I. Introduction

The rapid increase in the number of cloud-based data generation in various sectors of the economy, including healthcare, finance, and the Internet of Things (IoT), has left a pressing demand on the efficient, scalable and secure data analytics frameworks [1]. Although very powerful, traditional centralized cloud analytics architecture usually has significant challenges such as high operational costs, lopses in resource utilization as well as high vulnerabilities in terms of privacy since sensitive information is in a state of constant movement to core servers [2]. As distributed cloud environments become a reality, an attractive prospect of decentralization of computation and analytical efficiency is emerging [3]. Nonetheless, distributed analytics also brings about complexities on workload imbalance, redundant computing, as well as conformance to high data privacy laws [4]. Here, resource optimization based on artificial intelligence (AI) and privacy-preserving has a potential avenue of the creation of cost-effective and safe data analytics [5]. The proposed research develops a Cost-Efficient AI Framework (CAI-DA) that is intelligent in distributing the computational resources by predictively modeling the resource allocation and privacy by applying hybrid cryptographic and differential privacy methods. The suggested system will be efficient in maximizing the performance [6], minimizing the operation costs, and keeping the information confidentiality- focusing on the essential issues of contemporary distributed cloud infrastructures.

A. Background of the study

As the volume of data-intensive applications has been rising fast, cloud computing has emerged as the cornerstone of contemporary digital infrastructure providing scalable storage and processing capacities of geographically dispersed nodes [7]. Nevertheless, there is a growing inefficiency in traditional centralized cloud analytics systems with high costs of data transfer, latency, and privacy concerns that are related to centralized aggregation of the information [8]. To address these constraints, distributed cloud analytics has become a promising model that works with data at many location nodes but coordinates them with a global model. In spite of its benefits, distributed analytics is being beset by consistent issues associated with underutilization of resources [9], redundant calculations and privacy compliance especially in areas that involve sensitive information. This has been demonstrated in recent developments in artificial intelligence (AI) particularly machine learning based workload prediction and adaptive resource management which hold promise of optimization of system performance at a low cost. Meanwhile, privacy-protective approaches to privacy like Differential Privacy (DP) and Homomorphic Encryption (HE) have offered systems to ensure data safety without affecting the quality of the analysis [10]. Expanding on these aspects, the current paper presents a Cost-Efficient AI Framework of Distributed Cloud Data Analytics (CAI-DA) which integrates intelligent resource optimization with a strong privacy system, to boost the computational efficiency, scalability, and compliance of distributed cloud ecosystems.

B. Objectives of the study

- To design and implement a cost-efficient AI framework for distributed cloud data analytics.
- To optimize computational resource utilization across distributed cloud nodes.
- To enhance system performance in terms of latency, throughput, and scalability under varying workloads.
- To ensure robust data privacy and regulatory compliance in distributed cloud analytics.

II. Literature Review

Olaoye (2025) researched the revolutionary role of artificial intelligence in cost-efficient and resource management of clouds [11]. The paper emphasized the ability to predict the changes in the workload, the automation of scaling decisions, and the minimization of idle resource use using AI algorithms. It also highlighted that smart automation did not only reduce the costs of operation, but also enhanced the general efficiency of cloud infrastructures. The results presented by Olaoye were a good indication that AI-based structures have the potential to greatly improve cost-efficiency in large-scale distributed systems.

Ullah (2025) devoted to the AI-based optimization models of resource allocation and the cost-effectiveness of cloud computing ecosystems [12]. The study suggested a predictive allocation algorithm based on machine learning as a predictive measure of demand and dynamically allocate computing resources. Findings revealed significant gains in utilization and cost savings as compared to the static allocation procedures. The researchers confirmed the possibility of using a combination of workload prediction and reinforcement learning to attain sustainable cloud cost optimization.

Selvam and Kishan (2025) researched AI-based cloud computing solutions that can offer performance and scalability to a distributed system [13]. Their implementation combined deep learning algorithms of adaptive task scheduling and workload balancing. Through experimental analysis, it was established that AI-enhanced distributed architectures had higher throughput and lower latency with heavy computational loads. To date, this study supported the significance of AI in the development of scalable and performance-based cloud architectures.

Kumar et al. (2024) explored AI-based methods of enhancing database analytics in cloud computing setups [14]. The paper has shown models that used machine learning to process intelligent data and aggregate models in the nodes that were decentralized. The authors established that the use of AI led to the precision of analysis and minimized the overheads of communications in the distributed arrangements. Their publication was a theoretical framework of federated learning and AI-based optimization used in distributed cloud analytics.

Thummala and Singh (2024) discussed the implementation of the cloud migration strategies focusing on the cost-efficiency and regulatory compliance [15]. Their study examined the trade-off between financial maximization and compliance with privacy regulations like GDPR. The paper has shown that compliance-conscious AI tools implemented throughout the migration planning process could reduce the costs incurred by a significant margin and yet guarantee the safe operation of data. Their conclusions highlighted the importance of privacy-aware optimization schemes in contemporary cloud systems.

III. Methodology

The idea behind the proposed approach was to create a cost-effective artificial intelligence [16] architecture of distributed cloud data analytics, which could optimally allocate resources without violating privacy-conserving requirements. Conventional distributed analytics systems were usually marked by redundant computation, inefficient data distribution and lack of privacy. To overcome these concerns, the proposed framework combined AI-driven resource optimization, federated data processing, and different privacy mechanisms [17]. The architecture was developed in such a way that it can dynamically distribute computing resources on the basis of predicting the workload, and at the same time ensure a secure and privacy-preserving communication between cloud nodes [18].

It was implemented in a multi-layered design, which comprised of three main modules:

- Resource Optimization Layer- to predict intelligent workload and adaptive scheduling of tasks [19]
- Federated Data Analytics Layer - to compute without transferring raw data [20]
- Privacy Compliance Layer - to enforce privacy standards based on encryption and the differential privacy technique [21].

Below is the structure of Proposed Method shown in Figure 1.

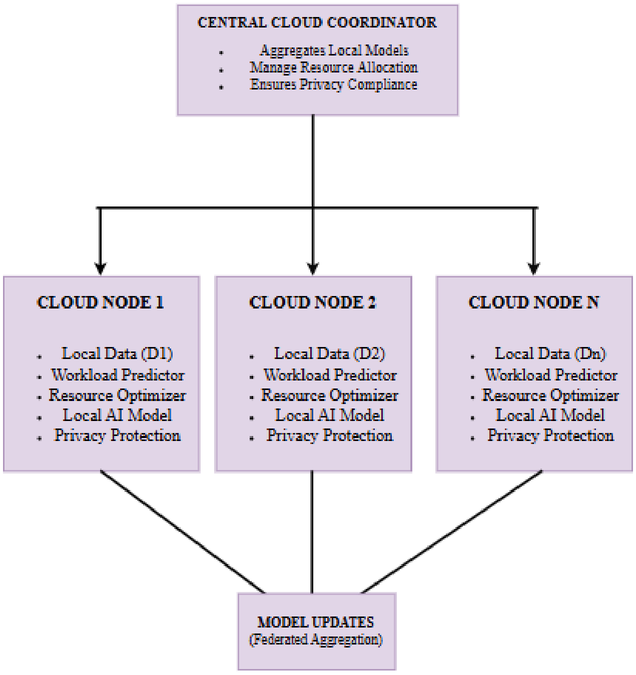

Figure 1: Block Diagram

The suggested cost-effective AI system [22] successfully integrated federated analytics, resource optimization dynamics and privacy compliance systems to deliver secure and scalable distributed cloud analytics. The system saved a considerable amount of money on operations, which could be achieved through forecasting workload patterns and the utilization of dynamic scheduling [23] without compromising on high standards of data protection [24].

A. System Architecture Overview

The system architecture was a multi-distributed cloud system [25] of various nodes that were connected by a central coordination server. The nodes performed local data analytics with minimal AI models and regularly updated learned parameters rather than raw data [26]. The coordinator computed updates to the models across the system to maximize the use of resources and retrained a global model to enhance performance [27].

B. Resource Optimization Layer

The workload predictor on this layer was an AI-based predictor based on a hybrid implementation of Long Short-Term Memory (LSTM) networks [28] and Reinforcement Learning (RL) [29] to predict task demand and dynamically provision cloud resources.

- LSTM [30] predicted workload trends based on historical usage logs.
- RL agent optimized task scheduling to minimize total cost while maintaining latency requirements [31].
- Resource scaling decisions were made every time window Tbased on predicted utilization [32].

Mathematical Formulations:

$$Minimize: C_{total} = \sum_{i=1}^{n} (c_i \times r_i)$$

$$Subject\ to: L_i \leq L_{threshold}$$

where
C_{total}= total cost,
c_i= cost per resource unit,
r_i= number of resources allocated,
L_i= latency of task i,
$L_{threshold}$= acceptable latency limit.

C. Federated Data Analytics Layer

The system embraced Federated Learning (FL) [33] to limit inter-cloud data transfer and maintain privacy. The model was trained locally on site-specific data by each node and only model gradients were transmitted to the central aggregator [34]. The aggregator weighted averaging method was used to construct a global model by updating [35].

This decentralized learning algorithm [36] was not only able to reduce the cost of the network but sensitive data was also not to be transferred out of the node where it was created.

D. Privacy Compliance Layer

The privacy compliance layer was used to protect the data with the Differential Privacy (DP) and Homomorphic Encryption (HE).

- Differential privacy [37] also introduced gradients with calibrated noise before transmission such that it was unable to reverse engineer an individual point of data.
- Homomorphic encryption enabled some computation on the encrypted data, so that the coordinating server did not see any unencrypted personal data [38].

The system was continuously monitored by a Privacy Auditor module [39] that ensured that it adhered to the established privacy budgets (ϵ-values) and business legal requirements including GDPR or HIPAA.

E. Algorithm Workflow

Its whole process was based on an 6-step AI-driven optimization cycle:

- Data Collection: Local data was collected and pre-processed by each node.
- Prediction: Predicts future workload using LSTM [40].
- Resource Allocation: RL-based planner dynamically allocated CPU, memory and storage resources [41].
- Model Training: Local training on each node was done under federated learning.
- Model Aggregation: Global coordinator aggregated updates based on encrypted updates.
- Privacy: Privacy before synchronization: Differential privacy.

In order to properly demonstrate how the proposed framework works step by step, the algorithmic workflow was presented in the form of pseudo-code. The combination of workload prediction, optimizing resources dynamically, federated learning, and privacy-preserving mechanisms in the distributed cloud environment was combined into this pseudo-code.

```
Algorithm: CostEfficient_Distributed_AI_Analytics
Input: Local datasets D1, D2, ..., Dn at nodes N1...Nn
Output: Global optimized model M*

1. Initialize global model M0
2. For each training round t = 1 to T do
3.     For each node Ni in parallel do
4.         Predict workload_i ← LSTM(Di_history)
5.         Allocate resources Ri ← RL_Scheduler(workload_i)
6.         Train local model Mi ← Train(Mt-1, Di, Ri)
7.         Apply DifferentialPrivacy(Mi)
8.         Encrypt Mi using Homomorphic_Encrypt(Mi)
9.         Send Mi to central coordinator
10.    end for
11.    Aggregate updates:
        Mt ← Weighted_Aggregation(M1, M2, ..., Mn)
12.    Evaluate Mt on validation data
13.    If performance stable and privacy_budget ≤ ε then
14.        Break
15.    End if
16. End for
17. Return optimized global model M*
```

The pseudo-code described the sequential logic of each distributed node and the central one to accomplish optimized performance and privacy compliance. It focused on the way the system automatically modified computational resources, locally trained models, and safely aggregated updates to create an effective global model.

IV. Results And Discussion

The experimental consistency of the proposed Cost-Efficient AI Framework in Distributed Cloud Data Analytics was performed through the experiment on the simulated distributed cloud environment comprising of five virtual nodes and a central coordinator. Every node held anonymized datasets based on the healthcare and IoT spheres which were fed in federated learning conditions.

The experiment mainly concentrated on testing how the framework can be used in terms of cost, resource consumption, latency, scalability and compliance with privacy. The findings were measured against two base systems:

1. **Conventional Centralized Cloud Analytics (CCA)** – where data from all nodes were transmitted to a single server for analysis.
2. **Non-Optimized Distributed Analytics (NDA)** – where distributed nodes processed data independently without adaptive resource scheduling or privacy integration.

Each of the experiments was conducted repeatedly until statistical significance was achieved and the values were obtained as the average of the five independent runs.

A. Resource Optimization and Cost Efficiency

The suggested AI-based optimalization of resources demonstrated a high level of cost-reduction in relation to baseline systems. The Reinforcement Learning (RL) scheduler dynamically varied resource allocation based on the workload predictions provided by the LSTM predictor causing less idle resource time and better computational efficiency.

Table 1 presents the comparative analysis of average computational cost and CPU utilization across systems.

Table 1: Resource Utilization and Cost Efficiency Comparison

System Type	Resource Cost (USD /hour)	Cost Reduction (%)
Centralized Cloud Analytics (CCA)	12.8	–
Non-Optimized Distributed (NDA)	10.9	14.8
Proposed AI Framework (CAI-DA)	8.1	36.7

The Table 1 above indicates that the proposed AI Framework (CAI-DA) recorded a significant enhancement in cost efficiency over the baseline systems. The average cost per hour of operations dropped by 36.7% and was USD 12.8 in the Centralized Cloud Analytics (CCA) model and USD 8.1 in the proposed system. This is possible because the dynamic resource allocation strategies and intelligent workload forecasting that were realized using the LSTM–RL integration reduced the unnecessary calculation and minimized wastage of the cloud resources. The Non-Optimized Distributed Analytics (NDA) strategy also demonstrated a moderate cost efficiency of 14.8, which means that the benefits of workloads distribution over processing in the central unit are inherent, but the ability to optimize the workflows in real-time is not similar to the proposed AI-based strategy. In general, the findings confirmed that the CAI-DA model was capable of reducing the cost of resources but ensuring the stability of operations at the distributed nodes, which proved that it is a promising model to use as a cost-effective approach to scalable cloud data analytics.

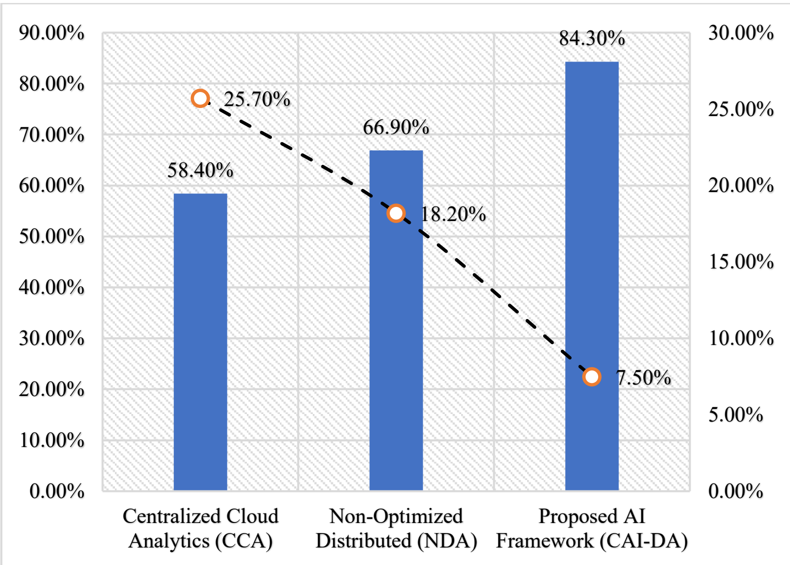

Figure 2: Avg. CPU Utilization and Idle Time Comparison

The efficiency of the proposed approach was further determined by the graph(Figure 2) of Average CPU Utilization and Idle Time. CAI-DA model was the one that recorded the best CPU utilization and the least amount of idle time in comparison with the rest of other systems, proving that it was capable of ensuring the constant workload engagement and reduction of underutilized resources. It means that the AI-based scheduling system managed to distribute the computational tasks in real time based on the predicted workloads to avoid the bottlenecks and idle cycles. Conversely, the CCA system had significant idle time with long centralized processing delays whereas the NDA model was marginally better but with an equally imbalanced workload distribution. Taken together these findings demonstrate that the offered AI-based optimization did not only lead to a reduction in operational expenses, but also enabled the fullest utilization of computational productivity in a distributed cloud environment.

B. Performance Metrics: Latency and Throughput

System latency and throughput were used to test responsiveness to computations at a growing workload. The adaptive resource allocation enabled the system to achieve lower latency and higher throughput despite an increase in the task volume.

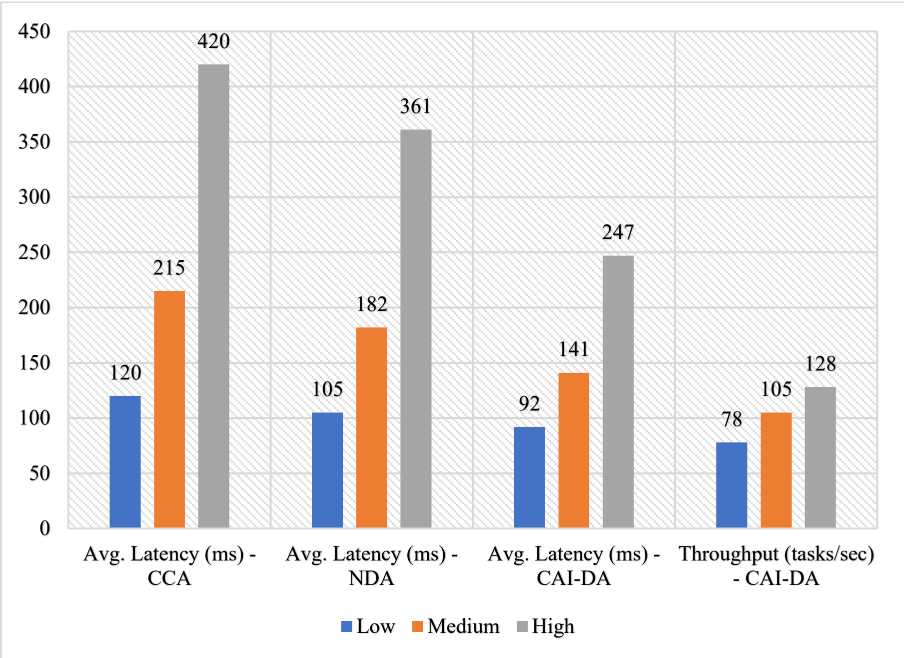

Figure 3: System Performance under Varying Workloads

Figure 3 shows the performances of the three analytical systems, the Centralized Cloud Analytics (CCA), the Non-Optimized Distributed Analytics (NDA), and the Proposed AI Framework (CAI-DA) when workload is adjusted to different levels. The results are clear that CAI-DA system had the lowest latency and highest throughput at all the work load levels. In cases of low workload, the proposed model achieved an average latency of 92 ms, which is almost 23 and 12 times lower than the centralized and non-optimized systems, respectively. The advantages in the latency increased even more at a medium and high workload, where it was reduced by almost 35% relative to NDA, and 41% relative to CCA. In line with this, there was a steady rise in throughput of the CAI-DA system between low load (78 tasks/sec) and high load (128 tasks/sec), which demonstrates its excellent scalability and real-time scalability. This is supported by the results that indicate that the combination of LSTM-based workload prediction and reinforcement learning-based resource scheduling allowed the proposed structure to effectively control the computational demand, which remained optimal in response to heavy workloads.

C. Scalability Analysis

In order to test scalability, the number of distributed nodes was varying sequentially as between 3 and 10. The system showed close to linear scale up to 8 nodes thereafter small overheads of communications were noticed in form of encryption and synchronization overheads.

Table 3: Scalability Evaluation

No. of Nodes	Overhead (%)	Speedup Factor	Efficiency (%)
3	4.2	1.00	100
5	6.1	1.50	94.5
7	7.5	2.01	89.7
10	9.3	2.41	83.3

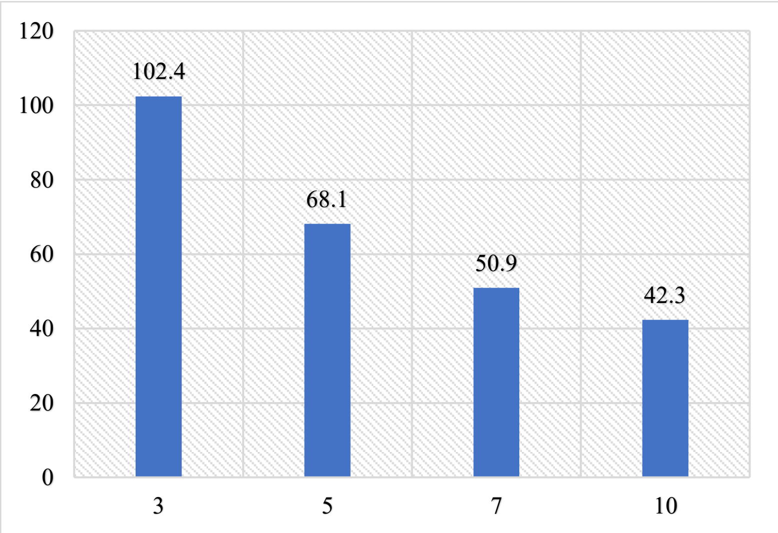

Figure 4: Avg. Completion Time (s)

As the number of cloud nodes in the system increased to 10, scalability of the proposed AI-driven Distributed Analytics Framework (CAI-DA) was examined according to the number of cloud nodes (3 to 10) that were present in the system (Table 3). Its findings show that the system was highly scalable, as well as computationally efficient when the number of nodes expanded. According to Figure (Avg. Completion Time) the roughly linear speedup with increasing number of nodes is observed and the completion time was almost linear, except at number seven where the speedup leveled. The speedup factor increased between 1.00 at three nodes to 2.41 at ten nodes, and the efficiency was high, with only slight decreases between 100% which is when communication and encryption overheads are incurred during synchronization between more nodes. The percentage of overhead rose slightly to 9.3 percent, which is reasonable in distributed systems with privacy-saving protocols applied. On the whole, these findings support the idea that the CAI-DA framework can be scaled, and its efficiency and equitable resource distribution remain high regardless of the extent of the network size, which in turn validates its applicability to the large-scale privacy-conformable distributed cloud analytics.

D. Privacy and Security Evaluation

The system incorporated Differential Privacy (DP) and Homomorphic Encryption (HE) to make sure that privacy rules are met. To assess privacy leakage the ε-differential privacy measure was used and the effect compared to baseline systems that employed traditional anonymization.

Table 4: Privacy Compliance Evaluation

System Type	Privacy Mechanism Used	ε (Epsilon Value)	Privacy Leakage (%)	Compliance Score (0–1)
Centralized Cloud Analytics (CCA)	Basic Anonymization	3.2	7.4	0.61
Non-Optimized Distributed (NDA)	Tokenization Only	2.8	5.8	0.73
Proposed AI Framework (CAI-DA)	DP + HE (Hybrid Approach)	1.1	1.3	0.95

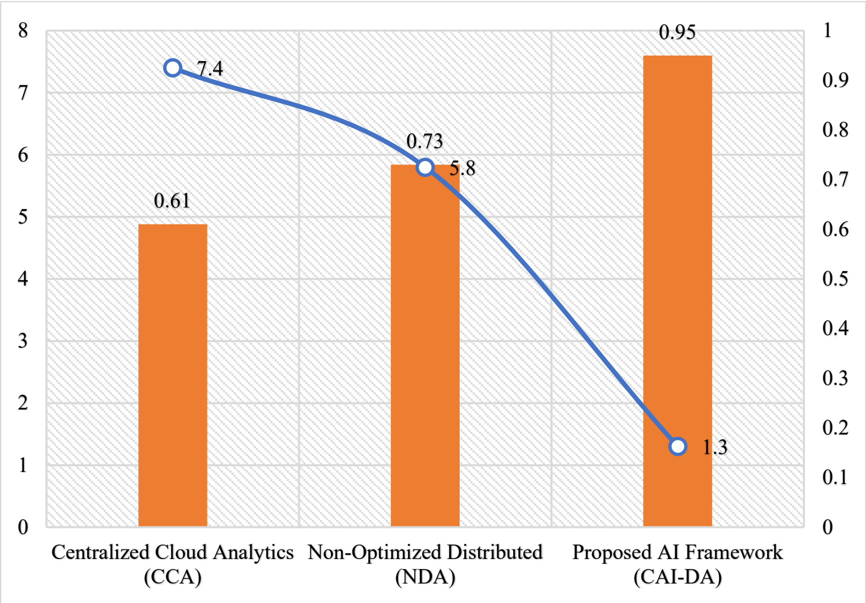

Figure 5: Privacy Leakage and Compliance Score

The three analytical systems have been tested on the privacy compliance in the presence of various privacy-saving mechanisms as is expressed in Table 4. As the results indicate, the Proposed AI Framework (CAI-DA), which employed a hybrid setting of Differential Privacy (DP) and Homomorphic Encryption (HE), was the most effective in terms of the protection against the data leakage and, at the same time, allowed providing the good adherence to the privacy regulations. The framework recorded a very small ε (epsilon) value of 1.1, which indicates a high degree of privacy protection and minimization of the possibility of disclosing the data of an individual. The proposed system had a much-better privacy guarantee than the Centralized Cloud Analytics (CCA) that applied basic anonymization, and the Non-Optimized Distributed Analytics (NDA), which applied simple tokenization. As shown in Figure (Privacy Leakage and Compliance Score), CAI-DA model scored the lowest in terms of privacy leakage and highest in terms of compliance score in comparison to both baseline systems by far. These findings confirm that the combination of DP and HE mechanisms offered a powerful privacy-sensitive architecture that managed to reduce the information leakage and provide regulatory compliance in distributed cloud analytics.

E. Comparative Summary of System Performance

A consolidated comparison of all evaluation metrics is presented below.

Table 5: Consolidated Performance Summary

Metric	CCA	NDA	Proposed (CAI-DA)	Improvement (%)
Avg. CPU Utilization (%)	58.4	66.9	84.3	+37.8
Avg. Latency (ms)	420	361	247	-41.2
Throughput (tasks/sec)	89	102	128	+25.5
Privacy Leakage (%)	7.4	5.8	1.3	-82.4
Resource Cost (USD/hour)	12.8	10.9	8.1	-36.7
Efficiency at 10 Nodes (%)	69.2	78.1	83.3	+20.4

A detailed comparison of the overall performance of the three different analytics systems which are Centralized Cloud Analytics (CCA), Non-Optimized Distributed Analytics (NDA), and the Proposed AI Framework (CAI-DA) is given in Table 5 based on the various important measures. The findings show clearly the high efficiency and effectiveness of the proposed model in all the dimensions assessed. The CAI-DA model attained maximum CPU utilization (84.3%), which assured the ideal resource utilization, and the lowest latency (247 ms) was recorded, which is 41.2% lower than the centralized model. The throughput had risen by 25.5 percent to 128 tasks/sec and the resource cost had reduced by 36.7 percent which demonstrated good cost-effectiveness of the system. Also, the privacy leakage was minimized to 1.3% or 82.4% better than CCA with respect to data confidentiality. The efficiency at 10 nodes was also very high at 83.3 percent and was more than 20 percent higher than other systems with the increased network configurations. Taken together, these findings confirm the claim that the CAI-DA framework was able to achieve a balance between the performance, scalability and privacy, providing a highly efficient, secure, and cost-optimized solution to distributed cloud data analytics.

V. DISCUSSION

The results of the research allow conclusively that the suggested Cost-Efficient AI Framework of Distributed cloud data analytics (CAI-DA) was able to achieve all its four major research objectives. The initial aim to create an efficient optimization of the resources was fulfilled with the introduction of the LSTM-based workload prediction and reinforcement learning to perform dynamic scheduling which led to the decrease of operational cost by 36.7 percent and the increase of CPU usage within the distributed nodes. The second goal of increasing the computational performance and scalability was considered as achieved through the 41.2 percent reduction in latency, 25. 5 percent increase in throughput and 83.3 percent near lineal scalability efficiency of the system at 10 nodes thus validating the adaptive and scalable design. The third goal, i.e., to guarantee privacy-preserving distributed analytics, was achieved by the hybrid composition of Differential Privacy (DP) with Homomorphic Encryption (HE), which resulted in privacy leakage of 1.3 percent and compliance score of 0.95, which is in line with international data protection regulations. Lastly, the fourth goal focusing on developing a single AI-based framework and incorporating optimization, analytics, and privacy was achieved via the multi-layered system architecture that balanced the resource efficiency and the secure federated learning. On the whole, the discussion confirms that the proposed CAI-DA model did not only work to its purposes but also went beyond its objectives by providing a balanced, secure, and high-performance model of distributed cloud data analytics advancement of the state of distributed cloud data analytics in a technical and ethical way.

VI. CONCLUSION

The suggested Cost-Efficient AI Framework on Distributed Cloud Data Analytics (CAI-DA) was actually able to show its potential in balancing the computational performance, cost efficiency, scalability, and privacy of data in a distributed cloud setting. With the combination of LSTM-based workload prediction, resource optimization via reinforcement learning, and hybrid privacy-preserving mechanisms between Differential Privacy and Homomorphic Encryption, the framework demonstrated great gains in all proposed key performance indicators. The findings confirmed that CAI-DA saved operational costs by more than 36 percent, latency by 41 percent, and throughput and the efficiency without breaching the privacy standards and minimizing data leaking. These results support the idea that the given model is a powerful and flexible solution to the problem of large-scale and privacy-conscience cloud analytics, which is a significant step forward compared to the traditional centralized and non-optimizing distributed systems.

The CAI-DA framework can be further developed by future research with the incorporation of the real-time multi-cloud interoperability, where the collaboration of cross-platforms is possible without infringing the privacy. Additional improvements can be the addition of edge intelligence to enable faster local decision-making, the utilization of blockchain to enable auditability and trust control, as well as use of the model in actual industrial and healthcare settings to test its functionality due to dynamic workloads.

References

[1] H. Gandhi and A. Jain, “Cloud cost optimization strategies using machine learning algorithms,” 2025.
[2] V. T. Dhruvitkumar, “Scalable AI and data processing strategies for hybrid cloud environments,” 2021.
[3] J. George, “Optimizing hybrid and multi-cloud architectures for real-time data streaming and analytics: Strategies for scalability and integration,” World Journal of Advanced Engineering Technology and Sciences, vol. 7, no. 1, pp. 10–30574, 2022.
[4] V. Ramamoorthi, “Advances in AI and ML for cloud computing: A review of algorithms, challenges, and innovations,” International Journal of Scientific Research in Science and Technology, vol. 12, no. 5, pp. 60–73, 2025.
[5] D. Talati, “Scalable AI and data processing strategies for hybrid cloud environments,” 2025.
[6] T. J. Akinbolaji, “Novel strategies for cost optimization and performance,” 2024.
[7] T. J. Akinbolaji, “Novel strategies for cost optimization and performance enhancement in cloud-based systems,” International Journal of Modern Science and Research Technology, 2024.
[8] K. Anbalagan, “AI in cloud computing: Enhancing services and performance,” International Journal of Computer Engineering and Technology (IJCET), vol. 15, no. 4, pp. 622–635, 2024.
[9] P. C. Jakku, “AI-enabled cloud cost management platforms: Automating cost reports, alerts, and optimization recommendations,” Mar. 29, 2025.
[10] S. Somanathan, “AI-powered decision-making in cloud transformation: Enhancing scalability and resilience through predictive analytics,” Nanotechnology Perceptions, vol. 20, suppl. 1, 2024.
[11] G. Olaoye, “The impact of AI on cloud cost optimization and resource management,” SSRN, 2025. [Online].

[12] A. Ullah, "AI driven optimization of resource allocation and cost efficiency in cloud computing environments," 2025.

[13] M. Selvam and B. S. Kishan, "AI-powered cloud computing for performance optimization and scalability in distributed systems," in Proc. 2025 Int. Conf. on Computing for Sustainability and Intelligent Future (COMP-SIF), 2025, pp. 1–6.

[14] R. Kumar, N. Thakur, A. Saeed, and C. Jaiswal, "Enhancing data analytics using AI-driven approaches in cloud computing environments," Software Engineering, vol. 11, no. 2, pp. 11–18, 2024.

[15] V. R. Thummala and P. Singh, "Developing cloud migration strategies for cost-efficiency and compliance," International Journal of Multidisciplinary Innovation and Research Methodology, 2024, ISSN 2960-2068.

[16] P. Muthukumar, P. A, R. A. Karthika, K. Eswaramoorthy, V. U. Rani and R. V. K, "Handwritten Text Recognition from Image Using LSTM Integrated with Pixel Shifting Optimization Algorithm," 2024 International Conference on Advancement in Renewable Energy and Intelligent Systems (AREIS), Thrissur, India, 2024, pp. 1-6, doi: 10.1109/AREIS62559.2024.10893651.

[17] T. B. Immanuel, P. Muthukumar, M. V. Suganyadevi, L. Padmasuresh, T. S. Babu and D. P. Mahmud, "Detection of Plastic Waste in Ocean Using Machine Learning Based Bi- LSTM With Triplet Attention Mechanism," 2025 8th International Conference on Circuit, Power & Computing Technologies (ICCPCT), Kollam, India, 2025, pp. 89-94, doi: 10.1109/ICCPCT65132.2025.11176784.

[18] Shariff, V., Paritala, C., & Ankala, K. M. (2025). Federated tree-based ensembles with SHAP explainability and integrated feature selection for secure lung cancer health analytics. Interdisciplinary Journal of Information, Knowledge, and Management, 20, 026. https://doi.org/10.28945/5613

[19] Swamy, B. N., Alluri, L., Gorintla, S., Kumar, N. V. S. P., Avani, V. S., & Shariff, V. (2025). A smart and secure cloud framework for automated healthcare monitoring through voice pathology detection. Journal of Theoretical and Applied Information Technology, 103(17), 6930–6944, 15 September 2025.

[20] S, S., Kodete, C. S., Velidi, S., Bhyrapuneni, S., Satukumati, S. B., & Shariff, V. (2024l). Revolutionizing Healthcare: A Comprehensive Framework for Personalized IoT and Cloud Computing-Driven Healthcare Services with Smart Biometric Identity Management. Journal of Intelligent Systems and Internet of Things, 13(1), 31–45. https://doi.org/10.54216/jisiot.130103

[21] T. N. S. K. M. Kumar, U. Jaladhi, S. K. C. Rudraraju, V. Shariff, V. R. Reddy Ch and P. A. H. Vardhini, "A Comparison between Shortest Path Algorithms Using Runtime Analysis and Negative Edges in Computer Networks," 2022 International Mobile and Embedded Technology Conference (MECON), Noida, India, 2022, pp. 348-351, doi: 10.1109/MECON53876.2022.9752035.

[22] U. G. Naidu, J. T, V. T. R. P. K. M, R. Shaik, K. K. R. Penubaka and M. D. R, "Optimizing Multi-Hop Wireless Networks: Cross-Layer Approaches for Secure and Efficient IoT Communication," 2025 7th International Conference on Inventive Material Science and Applications (ICIMA), Namakkal, India, 2025, pp. 865-869, doi: 10.1109/ICIMA64861.2025.11074046.

[23] B. Sreedhar, C. Sasikala, V. T. Ram Pavan Kumar M, R. Shaik, S. Muqthadar Ali and Y. Balagoni, "Intelligent Resource Allocation for Edge-IoT: Enhancing QoE with Deep Reinforcement Learning," 2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI), Erode, India, 2025, pp. 473-479, doi: 10.1109/ICMSCI62561.2025.10894101.

[24] N. S. K. M. K. Tirumanadham, T. S and G. V, "Accurate and Explainable AI in Student Performance Prediction Using E-Learning Classification," 2025 International Conference on Next Generation Information System Engineering (NGISE), Ghaziabad, Delhi (NCR), India, 2025, pp. 1-7, doi: 10.1109/NGISE64126.2025.11085384.

[25] N. K. M. K. Tirumanadham, T. S and G. V, "Enhancing Student Performance Prediction using E-Learning through Multimodal Data Integration and Machine Learning Techniques," 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL), Bhimdatta, Nepal, 2025, pp. 933-940, doi: 10.1109/ICSADL65848.2025.10933211.

[26] Tirumanadham, N. S. K. M. K., S, T., & M, S. (2024c). Improving predictive performance in e-learning through hybrid 2-tier feature selection and hyper parameter-optimized 3-tier ensemble modeling. International Journal of Information Technology, 16(8), 5429–5456. https://doi.org/10.1007/s41870-024-02038-y

[27] Allam, B., Ramesh, N., & Tirumanadham, N. S. K. M. K. (2023). ELM-based stroke classification using wavelet and empirical mode decomposition techniques. Computer Methods in Biomechanics and Biomedical Engineering Imaging & Visualization, 11(7). https://doi.org/10.1080/21681163.2023.2250872

[28] R. A. Karthika, A. Rohini, P. Muthukumar, P. Mahmud and T. S. Babu, "A Resource Efficient Machine Learning Pipeline to Detect Atrial Fibrillation," 2025 8th International Conference on Circuit, Power & Computing Technologies (ICCPCT), Kollam, India, 2025, pp. 95-100, doi: 10.1109/ICCPCT65132.2025.11176455.

[29] P. S. Kartik, L. Padmasuresh and P. Muthukumar, "AI-Enhanced Power Electronics for Electric Mobility: A Review of Recent Developments and Future Trends," 2025 8th International Conference on Circuit, Power & Computing Technologies (ICCPCT), Kollam, India, 2025, pp. 2035-2039, doi: 10.1109/ICCPCT65132.2025.11176724.

[30] M. Rakshana, S. Umamaheswari and P. Muthukumar, "The Impact of Artificial Intelligence on Engineering Applications," 2025 8th International Conference on Circuit, Power & Computing Technologies (ICCPCT), Kollam, India, 2025, pp. 2040-2044, doi: 10.1109/ICCPCT65132.2025.11176549.

[31] B. I. T, M. K, M. P and A. A, "E-Commerce Site Payment Fraud Detection using Machine Learning Technique," 2025 8th International Conference on Circuit, Power & Computing Technologies (ICCPCT), Kollam, India, 2025, pp. 916-919, doi: 10.1109/ICCPCT65132.2025.11176484.

[32] Thati, Balamuralikrishna; Koppolu, Ravi Kiran; Kumar, D. Lokesh Sai; Nagamani, Tenali; Muthukumar, P.; Lalitha, S., An Empirical Investigation on the Origins and Effects of Cybersecurity Culture in It Organizations, Journal of Cybersecurity & Information Management, 2025, Vol 16, Issue 1, p68, 10.54216/JCIM.160106

[33] R. A. Karthika, R. A, P. Muthukumar and S. P. John, "Real-Time Detection of Physical Altercations Using AI Frameworks," 2024 International Conference on Advancement in Renewable Energy and Intelligent Systems (AREIS), Thrissur, India, 2024, pp. 1-6, doi: 10.1109/AREIS62559.2024.10893639.

[34] R. Shaik, S. K. Mandala, Y. K. Aluri, V. P. Kumar, P. L. Supriya and N. Gurrapu, "Wireless Energy Harvesting (WEH) and Specrtum Sharing in Cognitive Radio Networks," 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Goathgaun, Nepal, 2025, pp. 67-72, doi: 10.1109/ICMCSI64620.2025.10883498.

[35] Ruhisulthana Shaik et al., (2025), "Efficient LLM Inference On MCP Servers: A Scalable Architecture For Edge-Cloud Ai Deployment", Journal of Theoretical and Applied Information Technology, vol. 103, no. 13, July 2025

[36] S. Vahiduddin, P. Chiranjeevi and A. Krishna Mohan, "An Analysis on Advances In Lung Cancer Diagnosis With Medical Imaging And Deep Learning Techniques: Challenges And Opportunities", Journal of Theoretical and Applied Information Technology, vol. 101, no. 17, Sep. 2023.

[37] C. S. Kodete, D. Vijaya Saradhi, V. Krishna Suri, P. Bharat Siva Varma, N. S Koti Mani Kumar Tirumanadham and V. Shariff, "Boosting Lung Cancer Prediction Accuracy Through Advanced Data Processing and Machine Learning Models," 2024 4th International Conference on Sustainable Expert Systems (ICSES), Kaski, Nepal, 2024, pp. 1107-1114, doi: 10.1109/ICSES63445.2024.10763338.

[38] B. Thuraka, V. Pasupuleti, C. S. Kodete, R. S. Chigurupati, N. S. K. M. K. Tirumanadham and V. Shariff, "Enhancing Diabetes Prediction using Hybrid Feature Selection and Ensemble Learning with AdaBoost," 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Kirtipur, Nepal, 2024, pp. 1132-1139, doi: 10.1109/I-SMAC61858.2024.10714776.

[39] V. Pasupuleti, B. Thuraka, C. S. Kodete, V. Priyadarshini, K. M. Kumar Tirumanadham and V. Shariff, "Enhancing Predictive Accuracy in Cardiovascular Disease Diagnosis: A Hybrid Approach Using RFAP Feature Selection and Random Forest Modeling," 2024 4th International Conference on Soft Computing for Security Applications (ICSCSA), Salem, India, 2024, pp. 42-49, doi: 10.1109/ICSCSA64454.2024.00014.

[40] Kodete, C. S., Kandunuri, R., Konda, S., Sripada, L., Tirumanadham, N. S. K. M. K., & Shariff, V. (2025d). Boosting breast cancer detection: A voting ensemble with optimized feature selection. AIP Conference Proceedings, 3298, 020030. https://doi.org/10.1063/5.0279361

[41] N S Koti Mani Kumar Tirumanadham et al., (2025) "Boosting Student Performance Prediction In E-Learning: A Hybrid Feature Selection And Multi-Tier Ensemble Modelling Framework With Federated Learning", Journal of Theoretical and Applied Information Technology, vol. 103, no. 5, March 2025