Grenze International Journal of Engineering and Technology, June Issue

System-Level Performance Evaluation of Cloud-based Deep Learning Training using AWS Sage-maker

Swathi Mithanthaya[1] and Sampath Kini[2]
[1-2]Nitte (Deemed to be University) NMAM Institute of Technology/CSE Department, Nitte, Karkala, India
Email: swathi.mithanthaya@gmail.com, sampath@nitte.edu.in

***Abstract*— Artificial intelligence systems have become a key enabler for the modern computational thinking technologies, especially for training the deep learning models such as Convolutional Neural Networks (CNNs), which are computationally intensive models. This paper introduces a benchmarking study of CNN training of Image Classification on AWS Sage Maker. The experiment is based on the popular Cats vs. Dogs dataset that has 25,000 labelled images. Two configurations of the SageMaker instance, that is, large and extra-large instances were evaluated for binary image classification problems and to monitor several system-level performance indicators. The metrics analysed are CPU-utilization, memory, training time and total cost of execution. Experimental results lead to meaningful information on the im-pact of cloud resources on deep learning workloads. In particular, compute-optimized in-stances substantially reduced the time for training, although the classification accuracy was comparable for the two configurations. In addition to this, in these instances, improved cost efficiency without sacrificing model performance also. The results of the above experiment can provide standardization to any small-medium businesses or explorers in the field of cog-nitive computing, assisting in decision making on cloud resources to scale and to cost-optimize deep learning applications.**

***Index Terms*— Deep Learning Benchmarking, AWS SageMaker, Cost-Efficient AI Systems, ML Infrastructure Optimization, Cloud-based Machine Learning, Cognitive Computing.**

I. INTRODUCTION

Deep learning has impacted how cognitive systems process and interpret raw data as major information in fields like image processing, NLP, and other autonomous systems. Among the variety of deep learning training models, Convolutional Neural Networks (CNNs) is one of the old, popular and one of the efficient models for image classification. However, training CNN requires high computation power and, depending on the application, it requires high CPU and
memory.

Cloud computing helps solve these issues by offering flexible and scalable computational environments accessible as per requirement. Managed platforms like the AWS SageMaker simplify the process of developing and deploying the machine learning workflows by minimizing the need for infrastructure management and providing tools for training, monitoring, and deploying models.

Grenze ID: 01.GIJET.12.2.91_23

However, deciding which configuration of cloud is best is still a challenging task because of the wide range of cloud instance types apparent and how they all vary in performance characteristics and cost structure. Selecting a wrong configuration can result in unnecessary expenses or a suboptimal performance.

For researchers and practitioners working under the constraints of time and money, finding a good balance between com-putational performance and cost is critical. Although there has been extensive research done in improving CNN architec-tures with higher accuracy, comparatively less work has gone into understanding the performance of the system-level of the cloud-based training environment. In particular, empirical studies on how the choice of the cloud instances affects the efficiency of its training, the utilization of resources, and the total expenditure related to the practical task of image classi-fication are still scarce.

To overcome the above gap, the current study focuses on the CNN training performance on AWS SageMaker using the popular Cats vs Dogs dataset featuring 25,000 labelled images. Different SageMaker instance types are analysed on the basis of the key performance parameters like CPU usage, Memory usage, Training time and cost. By correlating the per-formance results and financial expenditure information, this work gives actionable insights to optimize resource choice inside the cloud and provides the comprehension of cognitive computing systems which are cloud based.

II. Problem Definition

Despite more and more application providers moving to cloud platforms that offer machine learning, deep learning model training remains fairly problematic in cognitive computing environments. CNN training workloads demand a lot of com-putational resources, and even though our AWS SageMaker platform makes it very easy to execute through a managed service, the range of instance configurations is very broad and complicates the role of the resource. This complexity often leads to the inefficient use or higher costs of operations.

From the benchmarking point of view, there are relatively few experimental works that systematically analyse the influence of cloud instance types on the training performance in realistic image classification scenarios. Although SageMaker offers managed execution frameworks, there is limited advice on how to choose the configurations for the training instances in order to balance between training speed, cost and resource efficiency - especially in the case of medium dataset sizes, where adding more resources does not necessarily lead to a proportional rise in performance.

Existing deep learning research mainly focuses on algorithmic improvements and accuracy improvements with relatively less focus on understanding the performance implication of cloud infrastructure choices. Consequently, the standardized benchmarks at the system-level, which quantify the cost-performance trade-offs in distributed machine learning environ-ments on the cloud, are lacking. This is not clear and brings uncertainty for researchers trying to design efficient and economical cognitive computing pipelines.

The central issue tackled in the work in this paper is the lack of good system-level benchmarks to assess the impact of AWS SageMaker instance choice on CNN training performance and cost. Through controlled experimentation, this research would aim to support informed decision making in infrastructure development and provide part of the development of scalable, cost-effective cloud-based cognitive computing systems.

III. Literature Review

Cloud computing has become a basis to implement machine learning and intelligent systems thanks to its elastic resource provisioning and operational flexibility. Prior research shows that machine learning workloads can have a wide range of performance variations across cloud environments that operate with similar hardware specifications. Differences in the processor behaviour, memory management and overhead by virtualization frequently contribute to the training efficiency and stability. Existing comparative studies have often focused on the general-purpose computing benchmarks rather than learning intensive workload.

Although managed machine learning services help to reduce the development complexity, there may also be some trade-offs in terms of execution speed and fine-grained system control consideration. Cloud based cognitive services: They make it possible to perform perception and decision-making tasks, depending on the infrastructure configuration choices.

Research is also suggestive of the lack of consistent superiority of one cloud platform on others for all workload types, highlighting the need for workload-specific evaluation. These results support the need for targeted approaches to bench marking, with a focus on cognitive computing applications. This work identifies a gap of evaluating cloud infrastructure performance systematically with deep learning-based cognitive workloads.

TABLE I: LITERATURE REVIEW

Sl. No.	Authors (with citation no.)	Title of the Paper	Methodology	Key Findings
1	H. Jamal [1]	Performance Analy-sis of Machine Learn-ing Algorithm on Cloud Platforms: AWS vs Azure vs GCP	The study experimentally evaluates machine learning workloads exe-cuted on standard virtual machines across AWS, Azure, and GCP, focus-ing on system resource usage and ex-ecution efficiency	The results show noticeable differences in performance and cost across cloud providers, indicating that platform choice significantly impacts ML efficiency, although broader workload diversity is not considered
2	M. Townsend, H. Wimmer & J. Du [2]	Barriers and Drivers to Adoption of Cloud Infrastructure Services: A Security Perspective	This research relies on qualitative methods, including literature synthe-sis and stakeholder perspectives, to examine cloud adoption challenges	Findings emphasize that concerns related to security, trust, and regulatory compliance slow adoption, while cost savings and flexibility motivate organizations
3	S. Shekhar, P. Pandey & O. Goel [3]	Evaluating Scalable Solutions: A Comparative Study of AWS, Azure, and GCP	The authors conduct a comparative assessment of scalability features of-fered by leading cloud providers based on observed service characteristics	The study highlights the lack of extensive real-world experimentation and stresses the need for updated and large-scale benchmarking studies
4	D. Arrizki, S. Kosim & U. Yuhana [4]	A Comparative Performance Analysis and Cost Efficiency Between AWS and GCP Services	Performance and pricing metrics of AWS and GCP are evaluated using common software development workloads	Results indicate that cost efficiency and performance advantages depend on specific service configurations, with limited insight into managed ML services
5	P. Kaushik, A. Rao, D. Singh, S. Vashisht, and S. Gupta [5]	Cloud Computing and Comparison Based on Service and Performance	A controlled test environment is used to compare latency and throughput characteristics of IaaS and PaaS ser-vices	The analysis reveals differences in service maturity and performance consistency, though detailed ML workload analysis is missing
6	R. Gohil & H. Patel [6]	Comparative Analy-sis of Cloud Platform – A Review	The paper synthesizes vendor docu-mentation and prior studies to review cloud platform capabilities	It concludes that each provider offers distinct strengths, but rapid service evolution limits the long-term relevance of reviews
7	B. Gupta, P. Mittal & T. Mufti [7]	A Review on AWS, Azure, and GCP Services	The authors organize cloud services into functional categories through an extensive review of existing platforms	While useful for understanding service offerings, the study does not provide empirical comparisons across workloads
8	V. Nerella [8]	Automated Compli-ance Enforcement in Multi-Cloud Data-base Environments	A feature-level comparison is conducted on data governance tools across multiple cloud providers	Despite architectural and functional differences, the tools serve similar compliance objectives, exposing a gap in multi-cloud governance research
9	W. Choi, T. Choi & S. Heo [9]	Comparative Study of Automated Ma-chine Learning Platforms	AutoML platforms are evaluated ex-perimentally using a domain-specific anthropometric dataset	The study demonstrates that automation choices affect accuracy and training time, though conclusions are limited to a single application domain
10	A. Alkhatib, A. Shaheen & R. Albustanji [10]	Comparative Analy-sis of Cloud Compu-ting Services	The research combines literature re-view with limited experimental ob-servations	While functional differences among platforms are discussed, the absence of large-scale experiments reduces reproducibility
11	M. Omurgon-ulsen et al. [11]	Cloud Computing: A Systematic Literature Review and Future Agenda	A structured review of cloud compu-ting literature published over a five-year period is conducted	The study identifies dominant research themes, existing gaps, and proposes future directions for organizational adoption
12	R. Dhiman et al. [12]	Artificial Intelligence and Sustainability—A Review	The paper critically reviews AI-related sustainability research from en-vironmental and socio-economic perspectives	It notes a growing focus on sustainability but highlights weak integration with global development goals
13	S. Tripathi [13]	Assessing the Current Landscape of AI and Sustainability Litera-ture	A thematic examination of recent AI and sustainability studies is performed	The findings emphasize ethical concerns and the need for collaborative AI–human decision frameworks
14	C. Gohr [14]	Artificial Intelligence in Sustainable Development Research	A large-scale review of published re-search is conducted with respect to UN SDGs	The analysis shows that AI remains underutilized for sustainability applications, with uneven global research distri-bution
15	M. Enugula, S. K. Sahoo & S. Goswami [15]	Cloud Computing for Sustainable Development	The authors analyse sustainability benefits of cloud adoption across economic, environmental, and social dimensions	Cloud computing is shown to support efficiency and accessibility, though challenges must be addressed to maximize benefits

16	P. S. H. Darius [16]	Comparative Analy-sis of Cognitive Ser-vices in Popular Cloud Platforms	Cognitive APIs from major cloud providers are examined based on fea-tures and pricing models	The study reveals notable variation in service capabilities, influencing adop-tion decisions
17	M. Chen, F. Her-rera & K. Hwang [17]	Cognitive Compu-ting: Architecture, Technologies and Ap-plications	A conceptual framework is presented linking AI, cloud computing, and cognitive systems	The paper highlights the role of cogni-tive computing in enabling intelligent, human-centric applications
18	P. Pandit [18]	A Case Study of Am-azon Web Services	AWS services are reviewed through a descriptive case study approach	The study concludes that AWS offers scalable and cost-efficient infrastructure compared to traditional IT setups
19	J. Hao, T. Jiang, W. Wang, and I. K. Kim [19]	Empirical Analysis of VM Startup Times	VM boot times are measured across multiple public IaaS providers through empirical observation	Significant variability is observed, sug-gesting unpredictability in cloud startup performance
20	S. Ahuja et al. [20]	Performance Varia-bility in Public Clouds	The authors assess performance sta-bility across cloud VMs using controlled experiments	Results confirm that virtualization does not fully eliminate performance fluctua-tions
21	M. Cakir [21]	IaaS VM Perfor-mance Comparison	General-purpose VM instances are benchmarked across AWS, Azure, and GCP	The findings indicate that workload type strongly influences provider suitability
22	Anonymous [22]	Comparison and Analysis of Leading CSPs	A comparative review of cloud archi- tectures, pricing, and availability is conducted	The study emphasizes that cloud selec-tion must align with specific application requirements
23	N. V. Gadekar & A. Gadekar [23]	Comparing Major Cloud Providers for AI/ML Workloads	AI/ML services are reviewed across cloud platforms	Each provider demonstrates strengths in different AI dimensions
24	J. Jiang [24]	Serverless Machine Learning Training	Serverless platforms are experimen-tally evaluated for ML training tasks	The study identifies efficiency and scala-bility constraints in serverless ML work-flows
25	I. Panella et al. [25]	Deep Learning Cog-nitive Architecture	Existing cognitive architectures are analysed and a new DL-based model is proposed	The proposed architecture aims to en-hance reasoning and learning in autono-mous systems
26	S. Kaur [26]	AI in Cloud Compu-ting	The paper conceptually explores AI integration within cloud environ-ments	It highlights scalability benefits along-side future optimization challenges
27	J. Tharwani & A. Purkayastha [27]	Cost-Performance Evaluation of Com-pute Instances	Standardized cloud instances are compared based on cost and perfor-mance metrics	ARM-based instances offer cost ad-vantages, while Intel instances better serve enterprise needs
28	A. W. Damtew & Y. Y. Goshu [28]	Impacts of AWS on Supply Chain Resili-ence	AWS-based digital frameworks are examined in manufacturing contexts	Cloud adoption improves operational re-silience and supports sustainable prac-tices
29	P. Patel & S. Gajjar [29]	Cloud Wars	An industry-focused comparative study of major cloud providers is conducted	The study reinforces that cloud choice depends on organizational priorities
30	F. Mare, F. Mo-hamed & A. Mo-hammed [30]	AWS vs Microsoft Azure	Service offerings of AWS and Azure are compared across key dimensions	Both platforms provide strong capabili-ties, with suitability depending on user requirements

IV. Methodology

This study explores the methodology of using Convolutional Neural Networks (CNNs) for image classification using a cloud-based cognitive computing environment (in this case, AWS SageMaker AI Studio). SageMaker is a managed service for machine learning, a seamless platform to integrate data, computation, and monitoring facilitating the development of scalable AI applications. A CNN was chosen for this analysis because of its efficient task of extracting hierarchical features from images and can perform accurate pattern recognition. The dataset that was used taken 25000 images of cat and dog images from various datasets such as Kaggle and Pinterest. As part of preprocessing, duplicate images were removed, data was divided into different class folders and the dataset has been divided into training, validation and test set with ratio 70:10:30.

As it is a cloud deployment, a SageMaker domain is set up and two different types of instances are used: ml.t3.large: A general purpose, burstable CPU instance with 2 vCPUs and 8 GB RAM which can be used for initial experiments and for small size models. ml.m5.2xlarge: A compute optimized instance with 8 vCPUs and 32 GB RAM available that is useful for training larger models and also for heavy computational workload. The CNN was trained for five epochs, using Ten-sorFlow and Keras, with Python libraries, such as Boto3 and CloudWatch, being used to track important metrics such as the training time, CPU and memory usage, cost, and validation accuracy. Each of the instance types were tested separately and the collected data were visualized to analyse the efficiency, the utilization of resources, cost-effectiveness and the performance of the model. The process for implementation, shown in Fig. 1, shows a clear sequence: The CNN-based model is developed and is responsible for classifying the images in a cloud-based setup, the data set is cleaned and prepared, the training process takes place on different cloud instances and is used to check the impact of the hardware on the performance, and finally the results are analysed and the efficiency of the training and overall behaviour of the model is analysed. Even though this analysis provides experimental analysis of two exemplary Amazon SageMaker instances (ml.t3.large and ml.m5.2xlarge), Amazon AWS encompasses a broader category of instances such as GPU instances (ml.g4dn, ml.p3), Memory optimized instances (ml.r5), and Graviton2 instances based on ARM architectures. Such a broad analysis of in-stances considering larger datasets and complex neural networks is a subject of further analysis in this area due to resource and economic limitations.

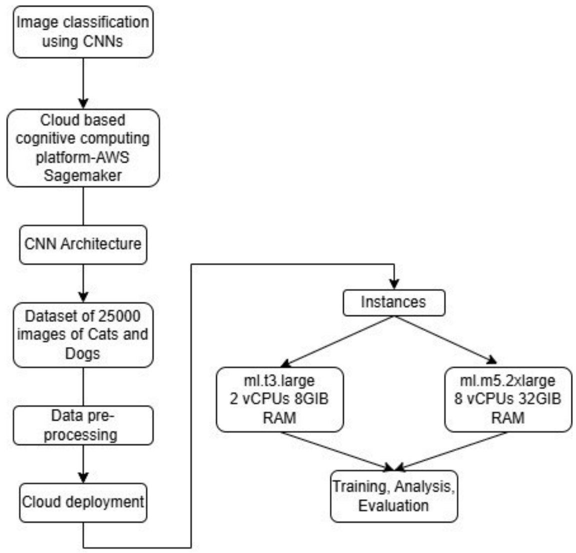


Figure 1. Implementation Flowchart

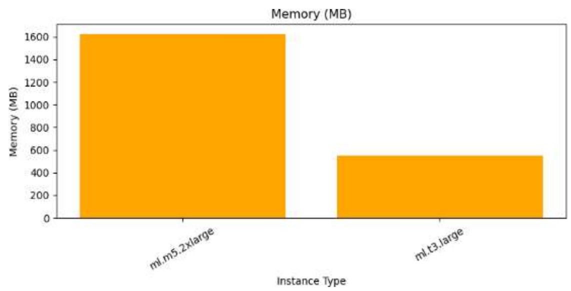
V. RESULTS AND DISCUSSIONS

This study to critique the performance of a Convolution Neural Network (CNN) used to classify 25,000 images of cats and dogs on AWS SageMaker. Two types of instances were tested: the general-purpose ml.t3.large, and the compute optimized ml.m5.2xlarge. After five training rounds, the CNN achieved a validation accuracy of about 90.7% on both of them which indicates effective learning and convergence without the problem of overfitting, this was demonstrated by the descendants of loss of training and validation supervision. This reflects the power of CNNs in learned hierarchical image features for image recognition applications.

Large differences were found in terms of computational efficiency and resource consumption between the two types of instances. Training using the ml.m5.2xlarge compute optimized took only 1,720 seconds, while the general-purpose ml.t3.large took 7,436 seconds, thus making the use of high-capacity computing cloud instances an advantage for deep learning tasks. Interestingly, CPU usage was higher on t3.large (66.7%) than on m5.2xlarge (49.2%), while memory con-sumption was higher from m5.2xlarge (1,552 MB vs. 522 MB), a reflection of the improved processing of the large datasets by the optimized instance. Cost analysis also was in the favor of the compute optimized instance with the cost of training on m5.2xlarge being $0.022 vs $0.207 on t3.large. Detailed metrics are presented in Table I and visual comparisons are shown in Figures 2 - 5.

Overall, this study serves as a proof-of-concept of how computational resources from the cloud can be used to improve the deployment of deep learning models, both in terms of high accuracy and cost efficiency. By carefully considering the right type of instances, it is possible to achieve an optimized process in the performance and resource use, laying the foundation for scalable and cost-effective AI systems. These findings are of particular relevance for cognitive computing and autono-mous applications where efficient to use cloud-based image recognition is a key for real-time decision-making situations and intelligent system operation.

From the scalability point of view, it was observed that when more computational resources were added, the training time decreased considerably without impacting the accuracy of the model. The scalability nature of the instance 'ml.m5.2xlarge' was better as it was able to process the same workload more efficiently compared to the general-purpose instance. For larger amounts of data or a deeper CNN model, it would be beneficial to implement vertical or horizontal scaling using GPU-based instances for better results. The AWS SageMaker tool also supports these scenarios with its capabilities of automatic resource allocation during the training process of models.



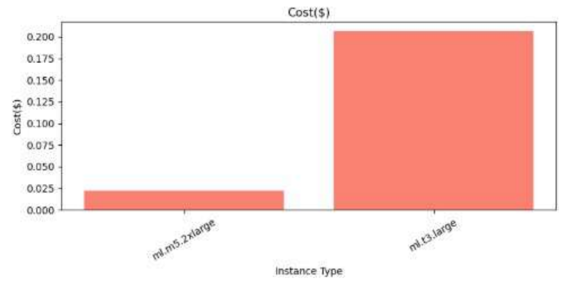
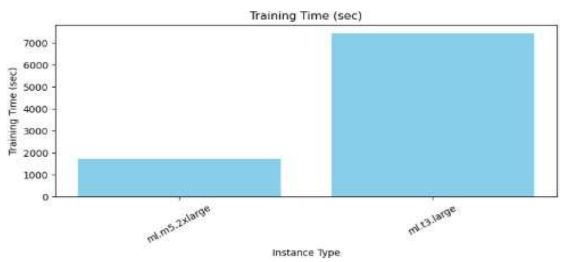


Figure 2. Memory vs Instance Type Figure 3. Cost vs Instance Type



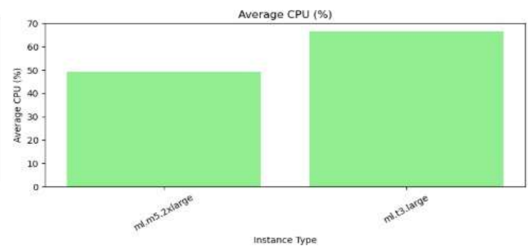


Figure 5 Training Time vs Instance Type Figure 4 Average CPU vs Instance Type

TABLE II. AWS PERFORMANCE

	Instance Type	Training_Time (seconds)	CPU usage (%)	Memory_used (MB)	Estimated Cost ($)	Validation Accuracy
1	ml.t3.large	7436.092	66.7	522.7	0.207	90.6976759
2	ml.m5.2xlarge	1720.016	49.2	1552.1	0.022	90.6976759

VI. CONCLUSIONS

This paper shows how Convolutional Neural Network (CNN) can be applied successfully for image classification in a cognitive computing environment based on AWS SageMaker cloud. While both tested instances had a good validation accuracy of over 90.7% the cost-optimized ml.m5.2xlarge was clearly superior to the general-purpose ml.t3.large in terms of training speed, memory usage, and overall cost. This reinforces the significance of selecting the optimal cloud instance to maximize the efficiency in deep learning workflows.

The research aligns with the field of Cognitive Computing and Autonomous Systems, and provides tried-and-tested per-formance numerical of computation and cost. It gives an elaborated side-by-side comparison of 2 most widely-used AWS SageMaker instances in small-to medium workload for CNN-based image classification-based applications. The training time and the CPU and memory usage and the total cost of cloud-based AI systems are also analysed in this research. The results reveal that it is feasible to reach high accuracy without excessive computational resources and thus cost-effective and sustainable AI solutions. Furthermore, this work also provides the insights for the researchers and practitioners trying to optimize scalability, efficiency, and cost when deploying the machine learning models on the cloud.

Although the present research is primarily concentrating on single instance training, the results indicate its scalability pos-sibilities on cloud deep learning infrastructure. The possible extensions of the present research could comprise multi-in-stance distributed training as well as the assessment of GPU SageMaker instances regarding scalability on realistic massive processing tasks. Looking forward, the methodology could be expanded to the other cloud environments, such as the Google Cloud Vertex AI, making it possible to compare the performance across different platforms. Such an approach would allow a broader framework for employing the deep learning model in autonomous and cognitive computing application domains, for real time, scalable, and cost-effective artificial intelligence (AI) deployments. Moreover, research can be further ex-tended to the scope with other SageMaker.

REFERENCES

[1] H. Jamal, "Performance analysis of machine learning algorithm on cloud platforms: AWS vs Azure vs GCP," in Information Technologies and Intelligent Decision-Making Systems, Springer Nature Switzerland, 2023, pp. 43–60.

[2] M. Townsend, H. Wimmer, and J. Du, "Barriers and drivers to adoption of cloud infrastructure services: A security perspective," in Proc. 2020 IEEE International IOT, Electronics (IEMTRONICS), 2020, pp. 1–7.

[3] S. Shekhar, P. Pandey, and O. Goel, "Evaluating scalable solutions: A comparative study of AWS, Azure, and GCP," International Journal of Novel Research and Development, vol. 9, no. 8, pp. 20–33, Aug. 2024.

[4] D. Arrizki, S. Kosim, and U. Yuhana, "A comparative performance analysis and cost efficiency between AWS and GCP services in cloud-based software development," in Proc. 2024 2nd International Conference on Software Engineer-ing and Information Technology (ICoSEIT), 2024, pp. 149–154.

[5] P. Kaushik, A. Rao, D. Singh, S. Vashisht, and S. Gupta, "Cloud computing and comparison based on service and performance between Amazon AWS, Microsoft Azure, and Google Cloud," in Proc. 2021 International Conference on Tech-nological Advancements and Innovations (ICTAI), 2021, pp. 268–273.

[6] R. Gohil and H. Patel, "Comparative analysis of cloud platform: Amazon Web Service, Microsoft Azure, and Google Cloud provider—A review," in Proc. 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2024, pp. 1–5.

[7] B. Gupta, P. Mittal, and T. Mufti, "A review on Amazon Web Service (AWS), Microsoft Azure, and Google Cloud Platform (GCP) services," 2021.

[8] V. Nerella, "Automated compliance enforcement in multi-cloud database environments: A comparative study of Azure Purview, AWS Macie, and GCP DLP," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 11, pp. 270–283, 2025.

[9] W. Choi, T. Choi, and S. Heo, "A comparative study of automated machine learning platforms for exercise anthropometry-based typology analysis," Bioengineering, vol. 10, no. 8, 2023.

[10] A. Alkhatib, A. Shaheen, and R. Albustanji, "A comparative analysis of cloud computing services: AWS, Azure, and GCP," International Journal of Computing and Digital Systems, vol. 18, pp. 1–15, 2025.

[11] M. Omurgonulsen, M. Ibis, Y. Kazançoğlu, and P. Singla, "Cloud computing: A systematic literature review and fu-ture agenda," Journal of Global Information Management, vol. 29, pp. 1–25, 2021, doi: 10.4018/JGIM. 20211101.oa40.

[12] R. Dhiman, S. Miteff, Y. Wang, S.-C. Ma, R. Amirikas, and B. Fabian, "Artificial intelligence and sustainability—A review," 2023. [Online]. Available: https://doi.org/10.20944/preprints202312.2117.v1

[13] S. Tripathi, "Assessing the current landscape of AI and sustainability literature: Identifying key trends, addressing gaps and challenges," Journal of Big Data, vol. 11, art. no. 65, 2024.

[14] C. Gohr, "Artificial intelligence in sustainable development research," Nature Sustainability, vol. 8, pp. 970–978, 2025.

[15] M. Enugula, S. K. Sahoo, and S. Goswami, "Cloud computing for sustainable development: An analysis of environmental, economic and social benefits," Journal of Future Sustainability, vol. 4, pp. 45–60, 2024.

[16] P. S. H. Darius., "Comparative analysis of cognitive services in popular cloud platforms," in Proc. International Conference on Computational Intelligence (ICCI 2022), Springer, 2023.

[17] M. Chen, F. Herrera, and K. Hwang, "Cognitive computing: Architecture, technologies and intelligent applications," IEEE Access, vol. 6, pp. 19774–19783, 2018.

[18] P. Pandit, "A case study of Amazon Web Services," 2021.

[19] J. Hao, T. Jiang, W. Wang, and I. K. Kim, "An empirical analysis of VM startup times in public IaaS clouds," arXiv preprint, 2021.

[20] S. Ahuja, V. H. Lopez Chalacan, and H. Resendez, "Performance variability in public clouds: An empirical assessment," Information, vol. 16, no. 5, art. no. 402, 2025.

[21] M. Cakir, "Cloud computing: IaaS general purpose VM performance comparison between Microsoft Azure, Amazon AWS, and Google Cloud GCP," 2023.

[22] "Comparison and analysis of leading cloud service providers (AWS, Azure and GCP)," International Journal of Advanced Research in Engineering and Technology, vol. 15, no. 3, pp. 266–278, 2024.

[23] N. V. Gadekar and A. Gadekar, "Comparing major cloud providers for AI/ML workloads: AWS vs Azure vs GCP," International Journal of Advanced Research in Education and Technology, vol. 11, no. 3, 2024.

[24] J. Jiang, "Towards demystifying serverless machine learning training," in Proc. ACM SIGMOD/PODS, 2021, pp. 857–871.

[25] I. Panella, L. Zanotti Fragonara, and A. Tsourdos, "A deep learning cognitive architecture: Towards a unified theory of cognition," 2020.

[26] S. Kaur, "Research paper on artificial intelligence in cloud computing," National Research Journal of Information Technology and Information Science, 2025.

[27] J. Tharwani and A. Purkayastha, “Cost-performance evaluation of general compute instances: AWS, Azure, GCP, and OCI,” International Journal of Computer Trends and Technology, vol. 72, no. 11, pp. 248–255, 2024.
[28] A. W. Damtew and Y. Y. Goshu, “The impacts of AWS from digital visions to action for supply chain resilience, performances, and inclusiveness,” International Journal of Advanced Manufacturing Technology, vol. 130, pp. 4821–4834, 2024.
[29] P. Patel and S. Gajjar, “Cloud wars: A comparative study of AWS, Azure and Google Cloud,” International Research Journal of Modernization in Engineering Technology and Science, vol. 5, no. 10, 2023.
[30] F. Mare, F. Mohamed, and A. Mohammed, “Amazon AWS vs. Microsoft Azure: Two prominent cloud service platforms,” 2023.