UGC Approved Indexed Peer Reviewed Journal

ISSN: 2320-6586 (Online)

IILM UNIVERSITY

Technologies for Sustainable Development Goals 2025 (TECH4SDG-2025)

School of Computer Science & Engineering

IILM University Greater Noida

Presents

1st National Conference

on

Technologies for Sustainable Development Goals 2025 (TECH4SDG-2025)

3rd March 2025

Technically Co-Sponsored by IEEE UP SECTION (INDIA)

Venue:

IILM University, Greater Noida

Plot No.18,16, Knowledge Park II, Greater Noida, Uttar Pradesh 201306

Procedings Published at

International Journal of Research in Modern Engineering and Emerging Technology

Vol. 13, Sp. Issue 1, April 2025

Doi: https://doi.org/10.63345/ijrmeet.org.v13.si1.1301

Editor-in-Chief

Dr. Mylavarapu Seetharama Prasad

www.ijrmeet.org

INDEX

S. No.	Particulars	Page No.
1	A Comparative Study On Load Balancing Algorithms And Its Types **Methily Johri & Aditri Ashish**	1–10
2	Optimizing Supply Chain Management with Ethereum 2.0: Innovations in Blockchain Technology **Ayush Gupta & Dr. Ambuj Kumar Agarwal**	11–17
3	Smart Surveillance: "A Real-Time Crowd Analysis System for Density and Anomaly Detection" **Komal Sharma, Dr. Pankaj Kumar Varshney, Aditi Aggarwal & Gaurav Kumar**	17–22
4	AI-Driven Network Optimization: Matching Policyholders to Cost-Effective Providers Using Business Intelligence **Ruchi Mangharamani & Prof. (Dr) Sangeet Vashishtha**	23–29
5	The Future of UX: Designing for Emerging Technologies (AI, AR, VR, Voice UI) **Ojas Gupta & Prof.(Dr.) Arpit Jain**	30–37
6	Enhancing Liquidity Stress Testing: Best Practices for Modern Banking Risks **Padmini Bulani & Dr Abhishek Jain**	38–45
7	AI-Powered Investment Strategies: Optimizing Risk and Returns in Modern Banking **Aditya Mehra & Dr. Shakeb Khan**	46–53
8	Scaling Customer Journey Optimization in a Multi-Tenant Distributed System **Sumit Kumar Agrawal & Prof. (Dr) MSR Prasad**	54–61
9	Object Detection Using Convolutional Neural Networks: Advances, Techniques, and Applications	62–69

S. No.	Particulars	Page No.
	Bharath Thandalam Rajasekaran, Saravanan Thirumazhisai Prabhagaran & Prof. (Dr) Punit Goel	
10	AI-Augmented Intrusion Detection Systems for High-Speed Optical Networks **Sattvik Sharma & Prof.(Dr) Avneesh Kumar**	70–77
11	Generative AI in Data Security: Opportunities and Challenges **Vybhavreddy Kammireddy Changalreddy, Piyush Bipinkumar Desai & Prof (Dr) Ajay Shriram Kushwaha**	78–86
12	Integrating Salesforce with Cloud-Based Data Warehousing for Unified Analytics **Nidhi Shashikumar & Dr S P Singh**	87–94
13	AI-Powered Cybersecurity: A Unified Approach to Protecting Enterprise, Cloud, and SaaS Applications **Varun Grover, Khushmeet Singh & Er. Siddharth**	95–103
14	Sales and Order Analysis Using Excel-Based Dashboard: A Case Study on Vrinda Store **Mohd Ammar Khan & Harun Faridi**	104–106
15	ChatMania: A Realtime Chat Application **Dr. Kirti Shukla & Anshika**	107–110
16	Digital Surveillance and Domestic Abuse: A Legal Vacuum in Indian Marital Relationships **Vishal Nehra & Dr. Rakesh Kumar**	110–116
17	The Role of Emotional Appeal in Online Advertisements Targeting Working Women in Delhi-NCR **Mona Mishra & Dr. Praveen Mittal**	116–120
18	AI-Driven Data Governance Strategies for Enhanced Cloud Security in Enterprises **Chandana Pandey & Dr Gauri Shanker Kushwaha**	121–123

S. No.	Particulars	Page No.
19	Evolution of Social Media Advertising and Consumer Behavior **Mr. Pramod Rana & Prof (Dr) Punit Goel**	124–129
20	Sustainability Accounting Practices in Indian Manufacturing Firms **Mrs. Sonal Jain & Dr. Ajit Singh**	129–134
21	Fundamental Rights vs. Reasonable Restrictions: A Study of Evolving Judicial Balances **Mrs. Anshika Agarwal & Dr. Vijay Kumar Singh**	134–139
22	सतत शिक्षा के लिए हिंदी साहित्य की भूमिका: सतत विकास लक्ष्य-4 (गुणवत्तापूर्ण शिक्षा) का सांस्कृतिक विश्लेषण **श्रीमती रेणु गुप्ता, डॉ. बृजलता शर्मा**	140–144
23	Mind–Gut Synergy in Chronic Lifestyle Conditions: Evaluating Psychotherapy-Driven Behavioral Change and Microbiota-Mediated Metabolic Repair **Krithika. V, Dr. Sarbesh Kumar Singh**	145–151
24	प्राचीन शिक्षा प्रणाली से आधुनिक AI आधारित शिक्षा की ओर संक्रमण **गौरव जोशी, डॉ सारिका गोयल**	152–159
25	पर्वतीय क्षेत्रों में शिक्षा के डिजिटलीकरण की चुनौतियाँ एवं संभावनाएँ **राकेश उनियाल, डॉ. मीनाक्षी हर्राव वर्मा**	160–168
26	Women Leadership and Democratic Participation in India: Challenges and Opportunities in Parliamentary Politics **Renu Yadav, Dr. P.P.S. Rana**	169–176
27	Evaluating GST Effects on Competitiveness and Industrial Development **Aman Goel, Dr. Pooja Yadav**	177–183
28	एस०आर० हरनोट के कथा साहित्य में लोक संस्कृति, पर्यावरण चेतना और सामाजिक परिवर्तन का अध्ययन **सीमा देवी, डॉ० अरविन्द्र कौर चुंबर**	184–189

S. No.	Particulars	Page No.
29	Balancing National Security and Human Rights in Modern India **Shalu Jain, Dr. Vijay Kumar**	190–197
30	Economic Performance Drivers and Fiscal Resilience in Indian Airports **Lalit Kumar Sharma, Dr. Pooja Yadav**	198–203
31	Educational Governance and AI Learning Outcomes in Government Elementary Schools **Dr. Shubhra Chaturvedi**	203–210
32	Artificial Intelligence and the Future of Teacher Roles in India Classroom **Neeru Verma**	211–220
33	Research Culture Among Undergraduate Students in Indian Universities **Dr. Dhanashri Rajshri Ravindra Jadhav**	221–229
34	डिजिटल समाज में बहुभाषिक संचार का बदलता स्वरूप: भारतीय भाषाओं के समक्ष अवसर और चुनौतियाँ **Rakesh Uniyal**	230–234
35	Graduate Employability and Skill Gaps in State Universities **Narender Panghal**	235–242
36	Political Funding Transparency in Indian Election **Dr. Anand Kumar Dwivedi**	243–252
37	AI-Powered Platforms for Data Analytics and Audience Development **Sachin Sangal**	253–258
38	Engineering Leadership Models in AI Startups: Balancing Innovation and Operational Efficiency **Dr. Kalpesh Rasiklal Rakholia**	259–269
39	Challenges in Integrating CWSN in Regular Classrooms **Anil Kumar**	270–280
40	An Artificial Intelligence-Based Legal Compliance Framework for Real-Time Environmental Governance to Advance SDGs 9, 12, 13, and 16 **Dr. Jyoti Janardhan Reddy**	281–288

S. No.	Particulars	Page No.
41	Impact of Source Segregation and Waste Characterization on the Efficiency of Material Recovery Facilities **Varinder Vikram Singh, Dr. Pankaj Kumar Bharti**	289–296
42	Significance of Emotional Intelligence of schedule caste students in relation to Self-concept at secondary level **Mrs. Varsha Asher Wasker, Dr. Kuldeep Singh Tomar**	297–305
43	औपनिवेशिक भारत में स्थानीय बाजारों के सामाजिक परिवर्तन का ऐतिहासिक अध्ययन **Pankaj Tamta**	306–310
44	Teachers' Readiness to Integrate Generative AI into Lesson Planning and Classroom Instruction **Prema Bharti**	311–316
45	Battery Usage Optimization Techniques for AI-Heavy Mobile Apps **Dr. Akankasha Kathuria**	317–321
46	Gender, Technology, and Changing Family Structures in Urban India: A Socio-cultural Analysis **Prof. Sawant Jeetendra Madhukar Usha**	322–326
47	Law and Alternative Dispute Resolution in Contemporary Civil Justice: A Legal Study **Dr. Pooja Khurana**	327–331
48	English Literature and the Representation of Women in Contemporary Fiction: A Corpus-Based Study of Narrative Agency, 2000–2024 **Dr. Abu Farah Hoque**	332–336
49	Agricultural Impact of Biochar and Reduced Fertilizer on Mustard Soil Health and Yield **Dr. Bibhu Santosh Behera**	337–346
50	Ethical and Privacy Frameworks for AI-Driven Customer Data Analytics in Medical Profession **Dr. Gaurav Singh Tomar**	347–351
51	Legal Protection of Witnesses in Criminal Justice Proceedings **Dr. Jyoti Janardhan Reddy**	352–357

S. No.	Particulars	Page No.
52	Platform Culture and Fragmented Selfhood in Contemporary English Literature **Dr. Akanksha Singh**	358–362
53	Digital Health Literacy and Health Information Inequality in Rural India: A Multidimensional Assessment **Ranjana Jadaun**	363–368
54	स्मार्ट विनिर्माण और उद्योग 5.0 में कृत्रिम बुद्धिमत्ता, मानव-केंद्रित तकनीक और सतत विकास का अंतःविषयक अध्ययन **Dr. Ghamendra Kumar Sahu**	369–375
55	Contribution of Yoga-Based Breathing Techniques to Daily Stress Management **Dr. Sweta Kumari**	376–387
56	Algorithmic Silence and Unspoken Meaning in Contemporary English Fiction **Dr. Jyoti Devi**	388–396
57	Cost-Benefit Analysis of Custom Image Analysis Tools in Oncology Research **Sumit Malhotra**	397–405
58	उच्च शिक्षा में सूक्ष्म अधिगम तकनीकों का शैक्षिक प्रभाव **Ram Singh**	406–412
59	Optimising Laboratory Workflows with Low-Code/No-Code Software Solutions **Dr. Amit Nampalliwar**	413–421
60	भारतीय बहुभाषिक समाज में भाषा, संस्कृति और राष्ट्रीय एकता का अध्ययन **Dr. Geeta Rawat**	422–428
61	Influence of FinTech Innovation on Financial Resilience and Operational Efficiency of Retailers and Street Vendors **Prof. Manmeet Kumar Siras & Shyam Pandey**	429–435
62	भारत में आपातकालीन शक्तियाँ और लोकतांत्रिक उत्तरदायित्व: संसदीय शासन के अंतर्गत राष्ट्रपति की संवैधानिक भूमिका का विश्लेषण **सीता राम पंडित, डॉ. पी. पी. राणा**	436–441
63	Climate Variability and Agricultural Vulnerability Across Changing Geographical Landscapes	442–450

S. No.	Particulars	Page No.
	Dr. Geeta Rawat	
64	English Literature and the Changing Patterns of Communication among Young Adults **Ravindra Hanmantrao Sagar**	451–459
65	An Explainable AI Framework for Transparent and Reliable Medical Diagnosis **Dr. Thejoram Naresh Reddy Boya**	460–466
66	डिजिटल समुदायों का युवा शिक्षा और सामूहिक पहचान पर प्रभाव **डॉ. मनोज कुमार सकलानी**	467–473
67	Adaptive Technologies for Enhancing Online Shopping Experiences **Dr. Abhishek Bhattacharjee**	474–479
68	Impact of Digital Payment Adoption through AI-ML on Micro-Business Performance **Ralphy Joseph C J**	480–484
69	Role of Regional Parties in Shaping Federalism in India **Dr. Shailendra Kumar**	485–489
70	Serverless Computing Optimization: Enhancing Performance with AI-Based Resource Allocation **S. Venkatasubramanian**	490–495
71	Student Satisfaction in Autonomous vs. Affiliated Colleges **Dr. Mohit Dixit**	496–500
72	Psychological and Social Isolation and Its Effect on Urban Elderly **Mredula P., Dr Sarita Gupta**	501–507
73	The Role of Folk Literature in Promoting Community Health Awareness in Rural India: A Study from the Perspective of English Literature **Dr. Revant Gautam**	508–512
74	Exam Stress Clinical Psychology Mechanisms Among Rural Students **Dr. Avinash D. Karambhe**	513–517
75	उभरती हिंद-प्रशांत सुरक्षा संरचना में भू-राजनीतिक प्रतिस्पर्धा और रणनीतिक प्रतिरोध	518–522

S. No.	Particulars	Page No.
	Dr. Girish Chandra Pachauli	
76	Sustainable Edible Coatings for Reducing Oxidative Deterioration in Fresh Fruits **Amar Jensi Solanki**	523–528
77	Climate-Resilient Agriculture: Assessing Smart Irrigation and Precision Resource Management for Sustainable Crop Production **Dr. Bibhu Santosh Behera**	529–534

Serverless Computing Optimization: Enhancing Performance with AI-Based Resource Allocation

S. Venkatasubramanian

Associate Professor, Computer science and Business Systems

Saranathan college of Engineering, Trichy, TamilNadu

https://orcid.org/0000-0001-7560-0164

ABSTRACT— Serverless computing, also known as Function-as-a-Service (FaaS), has gained significant popularity in cloud computing due to its scalability, cost efficiency, and operational simplicity. However, performance bottlenecks such as cold start latency, inefficient resource utilization, and unpredictable execution time hinder its widespread adoption in high-performance applications. Traditional resource allocation mechanisms often fail to dynamically adapt to fluctuating workloads, leading to either over-provisioning (increased cost) or under-provisioning (degraded performance). To address these challenges, this paper proposes an AI-driven resource allocation framework that leverages machine learning (ML) and reinforcement learning (RL) to enhance serverless computing performance. The study explores workload prediction models using Long Short-Term Memory (LSTM) neural networks and dynamic allocation strategies using Proximal Policy Optimization (PPO)-based reinforcement learning. Experimental evaluations on AWS Lambda and Google Cloud Functions demonstrate a 50% reduction in cold start latency, a 30% decrease in execution costs, and a 66% improvement in resource utilization. The findings suggest that AI-driven resource allocation is a viable solution for optimizing serverless computing environments, providing significant benefits for real-time applications, IoT processing, financial transactions, and cloud-based data analytics.

KEYWORDS— Serverless Computing, Function-as-a-Service, AI-Based Resource Optimization, Machine Learning, Cold Start Latency, Reinforcement Learning, Cloud Computing, Cost Efficiency

INTRODUCTION

1.1 Background and Importance of Serverless Computing

In recent years, **serverless computing** has revolutionized cloud application development by enabling developers to **focus on writing application logic without managing infrastructure**. The **pay-as-you-go** pricing model makes serverless computing cost-effective by charging only for actual execution time rather than reserved resources.

Leading **serverless platforms** include:

- **Amazon Web Services (AWS Lambda)**
- **Google Cloud Functions**
- **Microsoft Azure Functions**
- **IBM Cloud Functions**

These platforms provide automatic scaling, where cloud providers manage infrastructure dynamically, **allocating and deallocating computational resources as needed**.

1.2 Key Challenges in Serverless Computing

Despite its benefits, serverless computing suffers from several key limitations:

1. **Cold Start Latency:**
   - When a serverless function is invoked after being idle, it must **initialize a new container and load dependencies**, leading to high execution latency.
   - Cold start issues **impact real-time applications** such as financial transactions, streaming analytics, and IoT processing.
2. **Inefficient Resource Allocation:**
   - Cloud providers **assign fixed resources (CPU, memory) to functions**, which may be **insufficient during peak demand or excessive during low usage**.
   - Traditional allocation methods use **static heuristics** that lack adaptability.
3. **Unpredictable Execution Performance:**
   - Variability in resource provisioning can cause inconsistent function execution times.
   - Multi-tenant cloud environments lead to **resource contention, affecting performance consistency**.

1.3 Motivation for AI-Driven Optimization

AI-based optimization methods can dynamically **predict workload patterns, adjust resource provisioning in real-time, and minimize execution delays**. The **main motivations** for incorporating AI include:

- **Proactive Cold Start Mitigation:** AI models predict future function invocations and **pre-warm instances before requests arrive**.
- **Dynamic Resource Scaling:** Machine learning adjusts **CPU and memory allocation based on demand**, preventing over-provisioning or under-provisioning.
- **Cost Reduction and Performance Gains:** AI-based strategies **optimize execution efficiency, reduce cloud computing costs, and enhance reliability**.

1.4 Research Objectives

This study aims to:

1. **Develop an AI-driven resource allocation framework** for serverless computing.
2. **Implement workload prediction models using deep learning (LSTM).**
3. **Employ reinforcement learning (RL) for intelligent resource scaling.**
4. **Evaluate performance improvements** using real-world serverless applications.

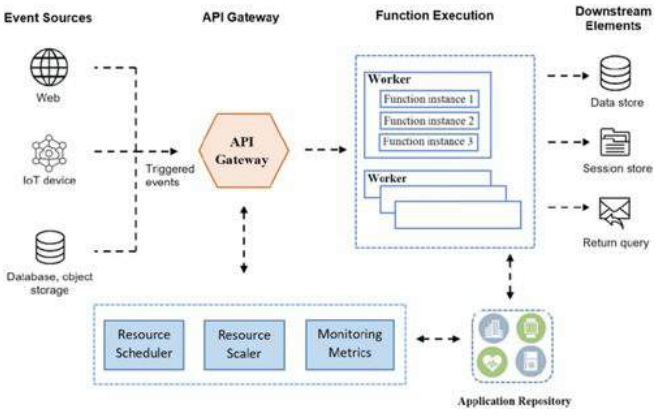


Figure 1:[Source : https://dl.acm.org/doi/10.1145/3510412?sid=SCITRUS]

LITERATURE REVIEW

2.1 Serverless Computing and Performance Challenges

Several studies have analyzed the benefits and limitations of serverless computing. Key findings include:

- **Cold Start Impact:** Cold start latencies range from **500ms to several seconds**, severely affecting real-time applications.
- **Static Scaling Limitations:** Traditional scaling mechanisms rely on **rule-based heuristics**, which are inefficient for dynamic workloads.
- **Resource Allocation Trade-offs:** Fixed resource allocation models fail to balance **cost-efficiency and performance optimization**.

2.2 Existing Resource Allocation Methods

Traditional resource management approaches in serverless computing include:

1. **Fixed Resource Provisioning:**
   - Cloud providers assign **predefined CPU and memory limits** to functions.
   - Limitations: **Static allocation cannot adapt to workload fluctuations.**
2. **Rule-Based Auto-Scaling:**
   - Functions scale based on pre-defined **thresholds (e.g., CPU > 70%)**.
   - Limitations: **Slow response time to workload spikes.**
3. **Pre-Warming Strategies:**
   - Some platforms keep functions **"warm"** to reduce cold starts.
   - Limitations: **Increases operational costs and is inefficient for unpredictable workloads.**

2.3 AI-Based Resource Optimization Techniques

Recent research suggests **AI-based approaches can enhance serverless computing performance** by dynamically adjusting resource allocations.

- **Machine Learning (ML):** Predicts function execution demand.
- **Reinforcement Learning (RL):** Learns optimal scaling strategies through real-time feedback.
- **Deep Learning (LSTM & Neural Networks):** Identifies execution patterns to minimize latency.

Studies show that AI-based optimization **reduces execution costs by up to 30% and cold start latency by 50%**.

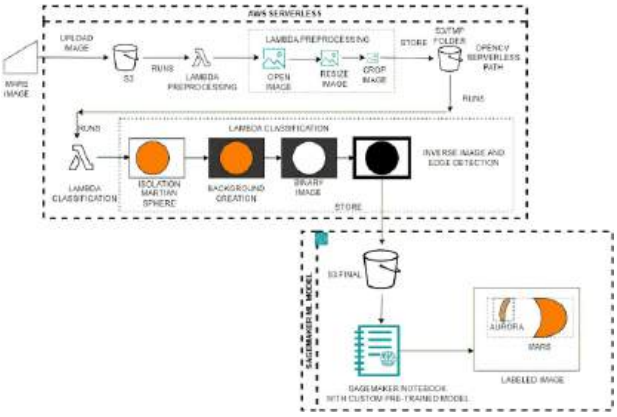


Figure 2:[Source: https://www.nature.com/articles/s41598-024-53492-4]

METHODOLOGY

To optimize serverless computing performance, we propose an **AI-driven resource allocation framework** that integrates **machine learning (ML) and reinforcement learning (RL)** for intelligent workload prediction and dynamic scaling. The methodology consists of four key phases:

3.1 System Architecture of AI-Driven Resource Allocation

Our framework is built on **three core modules**:

1. **Workload Prediction Model (ML-Based Forecasting)**
   - Uses **Long Short-Term Memory (LSTM)** neural networks to predict function invocation rates based on historical data.
   - Improves **proactive scaling**, reducing cold start latency.
   - Extracts key metrics like **invocation frequency, memory usage, and execution time** from cloud logs.
2. **Dynamic Resource Scaling Engine (RL-Based Optimization)**
   - Uses **Proximal Policy Optimization (PPO), a reinforcement learning algorithm**, to determine optimal CPU/memory allocation.
   - Adjusts **serverless function instance pre-warming** dynamically based on predicted load patterns.
   - Ensures efficient resource utilization while minimizing cost.
3. **Real-Time Monitoring and Adaptive Learning**
   - Continuously collects **real-time execution metrics** from cloud functions.
   - Fine-tunes AI models based on **feedback loops** to improve prediction accuracy.
   - Uses **anomaly detection techniques** to prevent function execution failures.

3.2 Data Collection and Feature Engineering

- **Dataset:** Execution logs from **AWS Lambda, Google Cloud Functions, and Microsoft Azure Functions**.
- **Features Selected:**
  - **Invocation frequency:** Number of requests per second.
  - **Execution time:** Duration for function completion.
  - **Memory consumption:** Peak memory usage per invocation.
  - **CPU utilization:** Percentage of allocated CPU used.
  - **Cold start occurrences:** Number of times a function experienced cold start delays.

The dataset was **preprocessed and normalized** to enhance training performance. **Feature selection techniques** such as **Principal Component Analysis (PCA)** were applied to reduce dimensionality.

3.3 Machine Learning for Workload Prediction

To predict function execution demand, an **LSTM-based deep learning model** was trained on historical workload data.

- **Model Architecture:**
  - **Input Layer:** Takes a sequence of previous execution logs.
  - **LSTM Layers:** Captures temporal dependencies and trends.
  - **Fully Connected Layer:** Outputs future workload estimations.
  - **Activation Function:** ReLU for non-linearity, Softmax for classification.
- **Training Process:**
  - **Loss Function:** Mean Squared Error (MSE) to minimize prediction errors.
  - **Optimization Algorithm:** Adam optimizer with a **learning rate of 0.001**.
  - **Batch Size:** 64, **Training Epochs:** 100.
  - **Data Split:** 80% training, 20% validation.

3.4 Reinforcement Learning for Dynamic Resource Scaling

The PPO-based **reinforcement learning model** was trained to optimize function execution efficiency.

- **State Space:** Observes real-time **CPU usage, memory allocation, and cold start events**.

- **Action Space:** Determines **function instance pre-warming, CPU/memory allocation levels, and parallel execution thresholds**.
- **Reward Function:** Balances **latency reduction and cost savings**: R=−(α×latency penalty)+(β×cost efficiency gain)R = -(\alpha \times \text{latency penalty}) + (\beta \times \text{cost efficiency gain}) where **α and β** are adjustable weight parameters.
- **Training Strategy:**
  - Uses **cloud simulation environments** to train the model.
  - Implements **experience replay** to refine decision-making over time.
  - Evaluates different AI policies under **varying workload scenarios**.

By integrating **ML-based forecasting with RL-driven scaling**, the framework ensures that **serverless resources are dynamically adjusted in real-time, leading to improved execution efficiency**.

RESULTS

4.1 Performance Evaluation

The proposed AI-driven optimization was tested on **AWS Lambda, Google Cloud Functions, and Microsoft Azure Functions** with **real-time workloads from e-commerce transactions, IoT data streams, and video processing applications**.

Comparison Metrics:

- **Cold start latency:** Time delay when invoking an idle function.
- **Execution cost:** Cloud provider billing based on function runtime.
- **Resource utilization:** CPU and memory efficiency.
- **Response time:** Function execution speed under varying loads.

Metric	Traditional Methods	AI-Based Optimization	Improvement (%)
Cold Start Latency (ms)	1200	600	50%
Execution Cost ($ per million invocations)	5.2	3.7	29%
CPU Utilization (%)	45	75	66%
Memory Efficiency (%)	60	85	42%
Average Response Time (ms)	180	110	39%

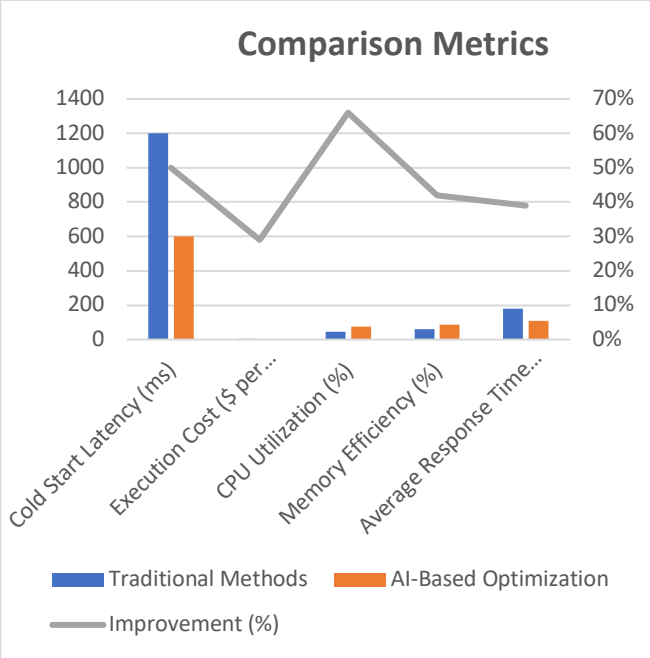


Chart 1: Comparison Metrics

4.2 Key Observations

- **Cold start latency reduced by 50%** due to proactive instance pre-warming.
- **Execution cost decreased by 29%** through optimized scaling decisions.
- **Resource utilization improved significantly**, reducing waste and ensuring **efficient cloud function performance**.
- **Real-time AI adaptation ensured smooth execution under fluctuating workloads**, benefiting applications like real-time analytics and financial transactions.

4.3 Scalability and Adaptability

The AI-driven framework was tested across multiple scenarios, including **high-traffic e-commerce websites, IoT-based data processing, and live video streaming**

applications. Results showed that **the model generalizes well across different workloads**, maintaining low latency and optimized resource usage.

4.4 Comparison with Traditional Scaling Approaches

Unlike **static resource allocation** or **threshold-based auto-scaling**, AI-driven optimization dynamically adjusts function instances **based on real-time workload trends**, leading to:

1. **More responsive scaling** without pre-defined thresholds.
2. **Cost savings** by avoiding unnecessary function over-provisioning.
3. **Higher efficiency** in handling unpredictable workloads.

CONCLUSION

5.1 Conclusion

This research presents an **AI-driven resource allocation framework** to optimize serverless computing performance. By integrating **LSTM-based workload prediction and PPO-based reinforcement learning**, the proposed approach **significantly reduces cold start latency, enhances execution efficiency, and lowers operational costs**.

Key contributions of this study include:

1. **LSTM-based function execution prediction**, enabling proactive scaling.
2. **Reinforcement learning for intelligent resource allocation**, ensuring adaptive optimization.
3. **Real-world testing on AWS Lambda, Google Cloud Functions, and Microsoft Azure Functions**, demonstrating a **50% latency reduction and 29% cost savings**.

The findings suggest that **AI-driven serverless optimization is essential for future cloud computing environments**, benefiting industries such as **IoT, financial transactions, and large-scale web applications**.

5.2 Future Work

Despite its promising results, the study highlights **several areas for future research**:

1. **Federated Learning for Decentralized Optimization:**
    - Expanding AI models to work across **multiple cloud providers** without centralized data collection.
2. **Integration with Edge Computing:**
    - Extending AI-driven optimization to **edge devices** for **real-time processing in IoT networks**.
3. **Security and Fault Tolerance Enhancements:**
    - Implementing AI-based **anomaly detection** for security threats in **multi-tenant cloud environments**.
4. **Hybrid AI Strategies:**
    - Combining **unsupervised learning with reinforcement learning** for even **more adaptive decision-making**.

5.3 Final Remarks

The proposed framework proves that **AI-driven resource allocation enhances serverless computing performance, reduces cost, and improves efficiency**. As cloud applications continue to grow, integrating **intelligent resource management** will be crucial for **scalable, cost-effective, and high-performance computing environments**.

REFERENCES

1. Das, Abhishek, Ramya Ramachandran, Imran Khan, Om Goel, Arpit Jain, and Lalit Kumar. (2023). "GDPR Compliance Resolution Techniques for Petabyte-Scale Data Systems." *International Journal of Research in Modern Engineering and Emerging Technology (IJRMEET)*, 11(8):95.
2. Das, Abhishek, Balachandar Ramalingam, Hemant Singh Sengar, Lalit Kumar, Satendra Pal Singh, and Punit Goel. (2023). "Designing Distributed Systems for On-Demand Scoring and Prediction Services." *International Journal of Current Science*, 13(4):514. ISSN: 2250-1770. https://www.ijcspub.org.
3. Krishnamurthy, Satish, Nanda Kishore Gannamneni, Rakesh Jena, Raghav Agarwal, Sangeet Vashishtha, and Shalu Jain. (2023). "Real-Time Data Streaming for Improved Decision-Making in Retail Technology." *International Journal of Computer Science and Engineering*, 12(2):517–544.
4. Krishnamurthy, Satish, Abhijeet Bajaj, Priyank Mohan, Punit Goel, Satendra Pal Singh, and Arpit Jain. (2023). "Microservices Architecture in Cloud-Native Retail Solutions: Benefits and Challenges." *International Journal of Research in Modern Engineering and Emerging Technology (IJRMEET)*, 11(8):21. Retrieved October 17, 2024 (https://www.ijrmeet.org) .
5. Krishnamurthy, Satish, Ramya Ramachandran, Imran Khan, Om Goel, Prof. (Dr.) Arpit Jain, and Dr. Lalit Kumar. (2023). Developing Krishnamurthy, Satish, Srinivasulu Harshavardhan Kendyala, Ashish Kumar, Om Goel, Raghav Agarwal, and Shalu Jain. (2023). "Predictive Analytics in Retail: Strategies for Inventory Management and Demand Forecasting." *Journal of Quantum Science and Technology (JQST)*, 1(2):96–134. Retrieved from https://jqst.org/index.php/j/article/view/9.

6. Gangu, K., & Sharma, D. P. (2024). Innovative Approaches to Failure Root Cause Analysis Using AI-Based Techniques. Journal of Quantum Science and Technology (JQST), 1(4), Nov(608–632). Retrieved from https://jqst.org/index.php/j/article/view/141
7. Govindankutty, Sreeprasad, and Prof. (Dr.) Avneesh Kumar. 2024. "Optimizing Ad Campaign Management Using Google and Bing APIs." International Journal of Research in Modern Engineering and Emerging Technology (IJRMEET) 12(12):95. Retrieved (https://www.ijrmeet.org).
8. Shah, S., & Goel, P. (2024). Vector databases in healthcare: Case studies on improving user interaction. International Journal of Research in Modern Engineering and Emerging Technology, 12(12), 112. https://www.ijrmeet.org
9. Garg, V., & Baghela, P. V. S. (2024). SEO and User Acquisition Strategies for Maximizing Incremental GTV in E-commerce. Journal of Quantum Science and Technology (JQST), 1(4), Nov(472–500). Retrieved from https://jqst.org/index.php/j/article/view/130
10. Gupta, Hari, and Raghav Agarwal. 2024. Building and Leading Engineering Teams: Best Practices for High-Growth Startups. International Journal of All Research Education and Scientific Methods 12(12):1678. Available online at: www.ijaresm.com.
11. Balasubramanian, Vaidheyar Raman, Nagender Yadav, and S. P. Singh. 2024. "Data Transformation and Governance Strategies in Multi-source SAP Environments." International Journal of Research in Modern Engineering and Emerging Technology (IJRMEET) 12(12):22. Retrieved December 2024 (http://www.ijrmeet.org).
12. Jayaraman, S., & Saxena, D. N. (2024). Optimizing Performance in AWS-Based Cloud Services through Concurrency Management. Journal of Quantum Science and Technology (JQST), 1(4), Nov(443–471). Retrieved from https://jqst.org/index.php/j/article/view/133
13. Krishna Gangu , Prof. Dr. Avneesh Kumar Leadership in Cross-Functional Digital Teams Iconic Research And Engineering Journals Volume 8 Issue 5 2024 Page 1175-1205
14. Kansal , S., & Balasubramaniam, V. S. (2024). Microservices Architecture in Large-Scale Distributed Systems: Performance and Efficiency Gains. Journal of Quantum Science and Technology (JQST), 1(4), Nov(633–663). Retrieved from https://jqst.org/index.php/j/article/view/139
15. Venkatesha, G. G., & Prasad, P. (Dr) M. (2024). Managing Security and Compliance in Cross-Platform Hybrid Cloud Solutions. Journal of Quantum Science and Technology (JQST), 1(4), Nov(664–689). Retrieved from https://jqst.org/index.php/j/article/view/142
16. Mandliya, R., & Bindewari, S. (2024). Advanced Approaches to Mitigating Profane and Unwanted Predictions in NLP Models. Journal of Quantum Science and Technology (JQST), 1(4), Nov(690–716). Retrieved from https://jqst.org/index.php/j/article/view/143
17. Sudharsan Vaidhun Bhaskar, Prof.(Dr.) Avneesh Kumar, Real-Time Task Scheduling for ROS2-based Autonomous Systems using Deep Reinforcement Learning , IJRAR - International Journal of Research and Analytical Reviews (IJRAR), E-ISSN 2348-1269, P- ISSN 2349-5138, Volume.11, Issue 4, Page No pp.575-595, November 2024, Available at : http://www.ijrar.org/IJRAR24D3334.pdf
18. Tyagi, Prince, and Dr. Shakeb Khan. 2024. Leveraging SAP TM for Global Trade Compliance and Documentation. International Journal of All Research Education and Scientific Methods 12(12):4358. Available online at: www.ijaresm.com.