iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VIII **Month of publication:** August 2026

DOI: https://doi.org/10.22214/ijraset.2026.84593

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VIII Aug 2026- Available at www.ijraset.com

Intelligent Portable Edge-Cloud Computing Architecture for Secure Data Analysis and Adaptive Resource Optimization Using AI-Driven Resource Scheduling

Pradeep Kachakayala[1], Akshith Kachakayala[2]
[1]*Cloud Engagement Manager, 58 Eddington Ln, Monroe, New Jersey, United States of America (USA),08831.*
[2]*Bachelors in Data Science, Pennsylvania State University , Montgomery County, Pennsylvania , 16802.*

Abstract: ***In recent years the growth of cloud computing, Internet of Things (IoT), artificial intelligence (AI) and edge intelligence has been increasing, and with it the need for portable, scalable and secure computing infrastructures that can process vast amounts of data that is dispersed, and has very low latency. Traditional cloud infrastructures are typically based on central server deployments which can be costly to deploy, immobile, have potentially greater communication latency, and waste resources in dynamic workload environments. In this paper, we introduced an Intelligent Portable Edge – Cloud Computing Architecture (IPECA) that combines the portable computing hardware, AI-based workload prediction, adaptive resource optimization, container-based virtualization and secure edge-cloud collaboration into a single computing architecture. In conventional architectures, there is no intelligent resource orchestration mechanism, which can provide flexible allocation of computational resources according to the property of workload, thermal status, energy consumption, network availability and so on. The architecture also features an adaptive security layer leveraging multiple layers of authentication, secure communication protocols, blockchain for integrity verification and on-the-fly system health monitoring to enhance cyber resilience. Simulations are conducted with varying workloads to gauge the effectiveness of the proposed architecture, and compared to traditional cloud and edge-cloud architectures with the metrics of latency, throughput, CPU utilization, response time, energy consumption, thermal efficiency, and resource utilization. Experiments demonstrate significant energy savings, scalability, responsiveness of the system and efficiency of computations using secure distributed processing. The suggested architecture is viable for the coming intelligent cloud infrastructures that are essential for smart city, industrial IoT, digital healthcare, education and enterprise computing.***

Keywords: Edge Computing, Cloud Computing, Resource Optimization, Artificial Intelligence, Secure Data Analytics, Edge AI, Distributed Computing.

I. INTRODUCTION

Yes, the rapid pace of the development of Cloud Computing, Artificial Intelligence (AI), Internet of Things (IoT), Big Data Analytics, Edge Intelligence and Industry 5.0 have transformed the modern computational infrastructures. Intelligent cloud services are increasingly being used to manage and analyse vast quantities of data spread across various organisations in real-time, across every sector of the economy, from healthcare, manufacturing, education, finance, transportation, agriculture and defense. The growing number of connected devices and the demands for low latency applications have radically changed the way we perceive computation from centralized cloud setups to decentralized intelligent computing environments with high computational efficiency, scalability, security and mobility[1],[2]. The key elements of the traditional cloud computing architecture include large-scale servers, virtualization systems, distributed storage solutions, and centralized data centers. While they possess virtually limitless computational power, they also have a number of practical problems like high deployment costs, higher communication latency, high energy consumption, high maintenance requirements, poor portability and need to be connected to a high-speed network at all times. Such a constraint is even more pronounced when applied to applications that require real-time analytics, mission critical decision making, disaster recovery, military use, industrial automation, remote healthcare or smart city deployments. To satisfy these requirements, the paradigm came into being – Edge Computing, which performs the calculations near the sources of data generation and does not carry all data to the centralized cloud servers. Edge computing can help to reduce network congestion and communication delay, which improves response time, and minimize bandwidth usage.

Edge devices, on the other hand, tend to have very few computational resources, storage space, thermal management, and resource scheduling mechanisms, which limits their effectiveness when dealing with heavy workloads.

Recently, the focus is on Edge–Cloud Collaborative Computing in which computation is dynamically scheduled from the edge sites to the centralized cloud system. This co-operative approach is a hybrid of edge computing and cloud computing that brings together the near-instant response time of edge computing and virtually limitless computing power of cloud computing. However, there are still some technical challenges to be addressed in the existing edge-cloud systems: inefficient workload scheduling, sub-optimal thermal management, limited portability, inadequate security mechanisms, underutilization of resources and lack of support for intelligent adaptive resource allocation.

At the same time, the modern enterprise environment puts greater and greater demands on portable cloud infrastructures that can provide cloud services from compact computing infrastructures that are easily deployed in geographically distributed locations. These include disaster management centres, emergency medical services, military communication units, industrial automation systems, smart transportation systems, educational laboratories, scientific research organisations and remote scientific expeditions. These application domains require computing platforms that are easily portable, modular, intelligent resource scheduling, offer secure communication, are efficient in cooling and easily scalable to integrate with cloud.

This newfound requirement has motivated this research to propose a Secure Data Analysis and Adaptive Resource Optimization Intelligent Portable Edge–Cloud Computing Architecture. The proposed approach is to integrate the capabilities of portable computing hardware, intelligent workload prediction, adaptive resource allocation, virtualization, container orchestration, thermal-aware scheduling, as well as multi-layer cybersecurity into a single computational platform unlike the traditional cloud servers. The entire architecture consists of AI which continuously analyzes the characteristics of a workload and monitors how the hardware is being used, and predicts computational needs and optimally allocates available resources as per the priorities of the applications and system conditions.

The proposed hardware components of the proposed framework are based on the UK Registered Design of the authors "Cloud Computing Device for Data Analysis and Management" (No. 6519300) which proposes an enclosed, modular portable hardware platform with optimized air-flow, integrated cooling systems and industrial mobility for a cloud-based computational environment. Apart from the industrial design, the current research also features a comprehensive and intelligent software architecture for adaptive resource optimization, secure distributed computing, virtualization, AI-enabled scheduling and performance-conscious workload management. The current study is intended to offer the smart computation models for future edge-cloud systems, while the registered design will deliver the computing platform.

The proposed architecture will include multiple smart software layers including an adaptive Artificial Intelligence Scheduler, a virtualisation manager for containers, a secure authentication system, an encrypted communication system, a thermal-aware resource optimiser, a distributed monitoring engine and a cloud orchestration module. These elements come together to properly use resources, low latency, scalability, high avaibility, energy efficient, and secure distributed computing.

There are a couple of interesting aspects to the proposed architecture, when compared to the more traditional cloud architectures. One is the unique hardware flexibility which makes it portable, reducing deployment complexity and enabling rapid installation in diverse operational situations. Secondly, the AI scheduling system continuously predicts workload changes and dynamically assigns computational resources to maximize the performance of the system. Thirdly, the framework for thermal-aware management intelligently distributes computational load to avoid the over heating of hardware and improve the reliability of hardware. Last, the multi-layered cybersecurity measures like multi-factor authentication, encrypted communication protocols, blockchain supported integrity checks, intrusion detection, and the zero-trust approach guarantee processing of sensitive data in distributed computing environments without compromising security.

A good amount of experimental testing is conducted on the proposed architecture to measure its performance in terms of widely used performance metrics like computational delay, throughput, CPU usage, memory usage, energy utilization, thermal stability, utilization of resources, scalability, response time and reliability of the system. The proposed intelligent architecture's capability in enabling modern distributed AI applications is illustrated by comparing it to the conventional cloud computing, standalone edge computing, and the existing edge-cloud architectures.

The proposed Intelligent Portable Edge–Cloud Computing Architecture will therefore offer an integrated architecture which will satisfy the growing computational demand for next generation intelligent systems, and improve portability, scalability, security, computational efficiency and adaptive resource management. The proposed research offers a realistic development trajectory for future portable cloud infrastructures in various fields such as Industry 5.0, smart manufacturing, digital health, autonomous transportation, intelligent surveillance, and large-scale distributed AI services.

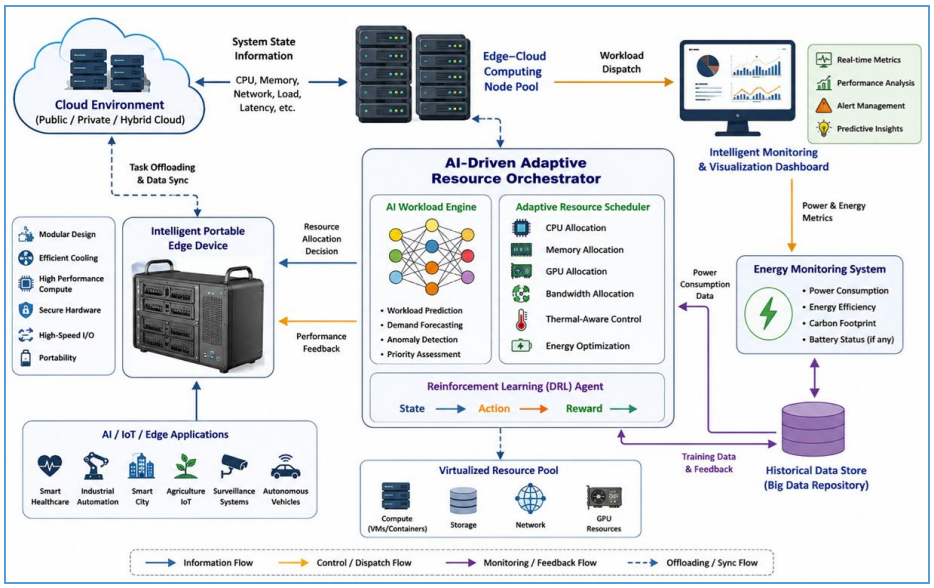


Fig. 1. Overall architecture of the proposed Intelligent Portable Edge–Cloud Computing Architecture

Fig. 1. Architecture for an Intelligent Portable Edge–Cloud Computing Architecture with AI workload prediction, deep reinforcement learning resource scheduling, secure edge-cloud orchestration, containerized virtualization, thermal management and adaptive resource optimization for scalable and energy-efficient distributed data analytics.

A. Motivation

Today's cloud infrastructures are facing unprecedented demands due to the exponential growth of AI applications, Internet of Things (IoT) devices, autonomous systems and distributed data analytics[3]. Traditional approaches to cloud computing heavily depend on centralized data centers, which can be expensive to set up, experience communication delays, congestion and lack portability in geographically distributed deployments. In addition, many mission-critical applications, such as disaster response, defense communications, remote healthcare, industrial automation, and smart city services, have computational platforms that must provide cloud services with minimal delays, while operating in a location where a traditional server infrastructure is impractical[4]. These challenges are at least partially overcome by existing edge-cloud solutions that bring computation closer to the data sources; however, they often fail to include intelligent orchestration of resources, adaptive scheduling of workloads, built-in thermal management, and holistic security mechanisms. Furthermore, the hardware platforms of commercially available portable computing systems are typically designed without intelligent software frameworks that can optimally allocate computational resources based on the constantly varying nature of their workloads and their environment. The proposed Intelligent Portable Edge – Cloud Computing Architecture was developed to address these challenges, with a focus on: AI-based workload prediction, Adaptive resource optimization, Secure distributed computing, and Thermal-aware scheduling. The proposed framework involves the intelligent software orchestration, which is integrated with the modular portable hardware platform inspired by the authors' UK Registered Design, with the intention of enhancing computational efficiency, scalability, energy use, and deployment flexibility of next-generation cloud computing applications.

II. RELATED WORK

Cloud computing has matured and transformed distributed computing with its provision of scalable compute power, virtualized infrastructure and smart service delivery. In recent years, the development of Artificial Intelligence (AI), Edge Computing[5], Internet of Things (IoT), Deep Reinforcement Learning (DRL) and container-based virtualization, among others, has driven research toward resource management solutions that are intelligent and also work well in applications that are latency-sensitive and computation-intensive. Although many of the challenges of cloud resource scheduling and edge collaboration have been overcome, there are practical challenges that remain: How to predict workloads in a manner that adapts to the cloud, how to move cloud resources to other locations, how to build an infrastructure that is thermal-aware, how to embed cybersecurity into the cloud and how to optimize resources for energy efficiency[6].

This section explores the research insights gained from AI-driven resources scheduling, edge-cloud collaboration, workload optimization, security and intelligent cloud infrastructures in detail and critically. The latest literature from 2024-2026 is critically reviewed, covering topics of resource scheduling, edge-cloud collaboration [7]-[8], workload optimization, security and intelligent cloud infrastructures, all enabled with Artificial Intelligence tools.

A. Cloud Computing: AI Based Resource Scheduling

AI is one of the critical technologies that is enabling intelligent cloud resource management[1],[2]. Recently, machine learning, reinforcement learning, deep neural networks, and graph learning algorithms have been employed to dynamically allocate computational resources according to the dynamic nature and changing demand of workloads. Recent systematic review Computing (Springer, 2026) is available in detail and explains in detail the technology of resource allocation used in modern cloud computing environments using AI. The authors identified the following methodologies that have been previously developed: supervised learning, reinforcement learning, deep reinforcement learning, evolutionary optimization, and hybrid intelligent scheduling. Their review has identified that although AI has significant impact on the accuracy of scheduling and computational efficiency, there are no standard evaluation frameworks in the scheduling field, sustainability-oriented optimization approaches, and package real-time deployment frameworks that can be deployed in production cloud environments. Furthermore, most of the existing approaches are not portable and do not consider thermal considerations and resource optimization. Similarly, in the recent past, few cloud-native scheduling frameworks have introduced Kubernetes-aware orchestration capabilities and infrastructure-aware scheduling capabilities to reduce energy consumption and improve workload balancing. While these techniques prove to be more efficient in terms of computing, they are mostly about container orchestration and not about portable cloud platforms that incorporate hardware-aware intelligence. This section explores the use of deep reinforcement learning for optimization of edge–cloud resources.

B. Deep Reinforcement Learning (DRL) for edge–cloud resource optimization

One of the most promising techniques for intelligent resource scheduling is Deep Reinforcement Learning (DRL) which has the ability to learn optimal scheduling policies under dynamic environments. One of the most comprehensive surveys was presented by Ismail et al., on the resource scheduling problem of Multi-access Edge Computing (MEC) using DRL[3],[4]. They reviewed over a hundred research papers related to computation offloading, content caching and intelligent resource management. The authors believe that Deep Q-Network (DQN) is the most widely applied DRL algorithm because it can optimize latency, energy consumption and quality of service. The survey also noted important research challenges, such as mobility awareness, workload balancing, privacy preservation, scalable scheduling, and integrated security mechanisms, which mean that the available solutions are fragmented and not comprehensive. Recently, a multi-objective task scheduling approach for edge-cloud computing based on a DRL was proposed in Scientific Reports (2026) that optimizes the three factors, namely latency, energy consumption, and service-level agreement (SLA) violation. Experimental analysis indicated that significant improvement was gained over conventional scheduling methods. However, the framework primarily targeted scheduling optimization and failed to mention portable architectures for hardware, virtualization, thermal management, and a complete cyber security mechanism. Similarly, For geo-distributed cloud computing system, an adaptive scheduling strategy for a DRL was proposed, which is considering the bandwidth-aware and cost-efficient scheduling strategy. The proposed framework was an improvement in the response time, in the cost efficiency but it was assumed that the cloud data centers are geographically distributed and not small portable cloud infrastructures to support intelligent edge deployment.

C. Edge–Cloud Collaborative Computing

With the growing trend of intelligent applications, the shift from centralized cloud computing to collaborative edge-cloud computing has grown even faster. To improve responsiveness and bandwidth efficiency in computation, edge-cloud collaboration can enable computation at the network edge, with cloud-based centralized resources for large-scale analytics, to execute latency sensitive computation at the network edge[5],[6].According to recent surveys, the top areas of focus for modern edge-cloud architectures are task offloading, adaptive resource placement, service migration and intelligent orchestration. But, most of the current systems do not take into account the integrated hardware portability, intelligent thermal management and uniform security architectures. In addition, the interoperability among heterogeneous edge devices is still a big research challenge. Scheduling framework for intelligent edge-cloud scheduling in next generation communications networks has been published in paper in Scientific Reports (2025).Paper of an intelligent edge-cloud scheduling framework for next generation communication networks published in Scientific Reports (2025). The framework had a positive effect on real-time scheduling and workload optimization, but it didn't take portable deployments and hardware-aware adaptive optimization into account.

D. Intelligent Resource Management in Fog and Edge Computing

Fog computing brings cloud computing closer to end devices by introducing intermediate processing layers in between. To distribute workloads, move services and offload computations[7], AI methods are becoming more common in the field of resource management in a fog computing environment. In addition, Reinforcement learning methods and evolutionary optimization and deep learning were found to improve significantly the latency and computational efficiency of fog computing resource management methods in a comprehensive review published on IEEE Access (2024). Regardless of this, it was recognized that there has been little research on the intersection of intelligent resource management and thermal-aware optimization, hardware portability and unified cybersecurity frameworks. Likewise, IEEE Access surveys on resource scheduling in edge computing highlight multi-objective optimisation with respect to latency, fairness and load balancing as well as green computing, intelligent orchestration and adaptive resource management as key future research avenues.

E. Research Gap Analysis

Although the objective of intelligent cloud resource management has been achieved, cloud resource management is still predominantly studied from a local optimization perspective, including computation offloading, task scheduling, energy efficiency, workload balancing and container orchestration. The integration of a portable cloud infrastructure, workload prediction using AI, deep reinforcement learning scheduling, virtualization through containers, thermal management of resource utilization, multi-layered cyber security, and energy monitoring is a rare mix in previous research found in a single computing system. Further, the commercially available portable cloud servers are mostly geared towards hardware portability and don't offer intelligent software architecture that can offer adaptable scheduling, dynamic workload forecast and secured resource orchestration[9],[10].

The portable hardware platform suitable for cloud computing applications is provided by the previously registered United Kingdom Registered Design for a "Cloud Computing Device for Data Analysis and Management" (Design No. 6519300) of the authors[20]. This paper presents an integrated intelligent edge-cloud architecture to address some of these challenges, and presents a complete framework for workload distribution between services consisting of workload prediction using AI, scheduling based on DRL, secure container orchestration, thermal management, and energy-aware optimisation[8].

III. PROPOSED INTELLIGENT PORTABLE EDGE–CLOUD COMPUTING ARCHITECTURE

Cloud Infrastructures are getting deployed on a larger scale to support new operations like Artificial Intelligence (AI), Internet of Things (IoT), Industry 5.0, autonomous systems and large scale distributed analytics, that require more computational demands from them.

The existing “Centralized” cloud architecture has several drawbacks in terms of latency, bandwidth consumption, deployment flexibility, energy consumption, and real time responsiveness in particular, especially in the case of application deployment in different geographic areas and mission critical applications. One of the current trends is intelligent edge-cloud collaboration, dynamically distributing the computational load between the edge devices and cloud-based systems via adaptive scheduling and intelligent orchestration.

The most recent research on orchestration emphasizes on multi-layer orchestration and flexible deployment of services in edge and cloud resources.

This study aims to overcome these challenges by creating an Intelligent Portable Edge–Cloud Computing Architecture (IPECA), which integrates a portable cloud hardware platform, an AI-based workload prediction system, a Deep Reinforcement Learning (DRL) scheduler, the adaptive resource optimization, secure virtualization, thermal-aware resource management, and cloud-edge orchestration into an intelligent computing system.

Unlike the other cloud computing architectures that mainly focus on software and/or container management, the proposed one combines intelligent software services and the mobile computing platform, which can be easily deployed in the healthcare sector, smart cities, industrial automation, disaster management, military operations, and in educational laboratories and remote scientific research facilities.

The hardware platform is based upon the authors' UK Registered Design "Cloud Computing Device for Data Analysis and Management" (Design No. 6519300) [20] which proposes a compact design, appropriate for portable cloud computing applications. This work did not only cover the registered design but extended it to a fully-fledged computational architecture based on artificial intelligent scheduling system that adapts to the actual workload, artificial intelligent resources orchestration, virtualization, thermal optimization, cyber security and distributed cloud-edge collaboration.

The entire process of the proposed architecture is shown in Figure 2.

A. Overall System Architecture

Proposed Intelligent Portable Edge–Cloud Computing Architecture consists of nine tightly coupled functional layers, which are responsible for performing intelligent computation, adaptive scheduling, secure communication, workload prediction and virtualization and resource optimization. Cloud Environment offers centralized compute services, massive storage, backup services, AI model repository and application hosting services. Depending on the deployments requirement, they can be public, private, hybrid or community cloud.The Edge–Cloud Orchestrator will allow for communication between cloud servers and portable computing devices, and will intelligently offload, synchronize workloads, migrate services, provide resources and manage distributed applications. The orchestrator monitors application characteristics, network conditions and determines whether to perform application computation within the local resources, or centralize it and move it to the cloud-based resources.

B. AI Driven Adaptive Resource Optimization

At the heart of the proposed architecture is the AI-Driven Adaptive Resource Optimizer, which is continually assessing the status of the computing platform with the help of various metrics of hardware and software performance.

The monitoring sub-system collects:

- CPU utilization
- GPU utilization
- Memory usage
- Storage utilization
- Network bandwidth
- Queue length
- Temperature
- Power consumption
- Running containers
- Virtual machine status

The set of all these parameters is the state of computation of the portable cloud device. The proposed solution breaks from the static scheduling policies and instead uses an AI Workload Predictor to predict future workload needs based on historical workload patterns, the application priorities and resource usage metrics. The workload prediction is fed to the Deep Reinforcement Learning Scheduler, whose objective is to learn the best scheduling policies in the computing environment as time goes on.

The proposed DRL scheduler tries to optimize a number of objectives, which differ from the standard heuristic schedulers:

- computational latency
- response time
- resource utilization
- energy consumption
- workload balancing
- thermal stability
- SLA satisfaction

The scheduler dynamically allocates the CPU cores, GPUs, memory, storage bandwidth and network capacity based on the current system condition to create an Adaptive Resource Allocation Policy.

C. Containerized Virtualization Layer

A new framework has been added, which will enable the scalable deployment of apps thanks to the light weight Container and Virtualization Layer, which are based on the Docker and Kubernetes technologies.

The advantages of containerization over VMs are:

- lower startup latency
- reduced memory consumption
- lightweight execution
- platform independence
- simplified deployment
- efficient scalability

The kubernetes system is installed in each of the AI applications running in a separate environment (containers):

- automatic deployment
- container orchestration
- service discovery
- load balancing
- failure recovery
- horizontal scaling

This design allows multiple workloads of AI to be run concurrently, while keeping compute overheads to a minimum.

D. An Intelligent Resource Allocation Policy

The proposed architecture is different from the conventional approaches to resource allocation which statically assign computational resources because it dynamically reconfigures the resource allocation based on the characteristics of the workload and the estimated computational demand.

The allocation policy is related to control of the following:

- CPU scheduling
- GPU allocation
- RAM allocation
- Storage bandwidth
- Network bandwidth
- Container placement
- Task migration

The scheduler tries to avoid contention for resources, and get maximum throughput of computation, and to balance the use of the hardware resources between all available processing elements. Also, adaptive resource allocation can allow for automated redistribution of computation when system is overloaded or when there are thermal limits.

E. Security & Trust Framework

In today's world, you can find a variety of sensitive data within cloud systems and cyber security is now an imperative requirement that has become a crucial component of your architecture. Therefore, in this work we have proposed a specific Security and Trust Layer that would be based on using multiple security levels simultaneously.

These include

- Multi-Factor Authentication (MFA)
- AES-256 Encryption
- TLS Secure Communication
- Blockchain-assisted Integrity Verification
- Intrusion Detection System (IDS)
- Read and stop to ponder the following. Now read and now read it!
- Secure API Gateway
- Continuous Vulnerability Monitoring

The built-in security paradigm helps protect apps that run locally, as well as cloud-based apps, from unauthorized access, cyber-attacks, malicious workload injection and data tampering. The proposed solution is not similar to other existing architectures which have security as an out-of-band service. It’s not the same as other existing architectures that have security as an out-of-band service.

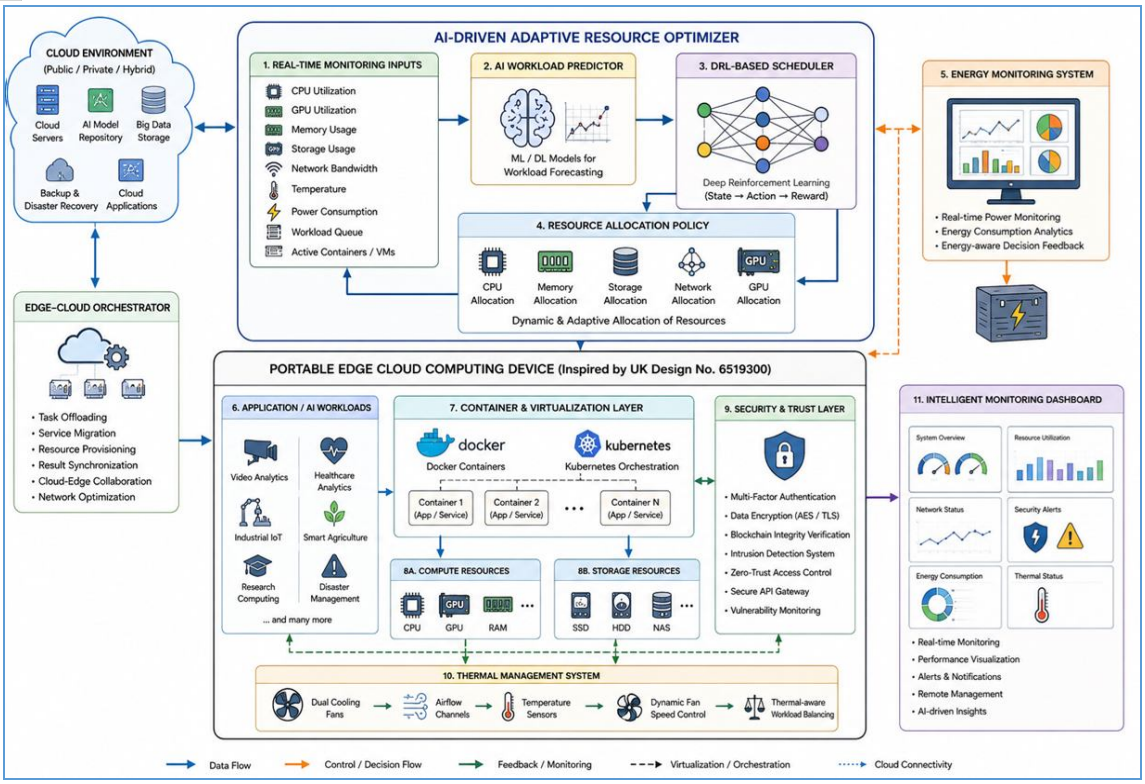


Fig 2. Detailed architecture of proposed intelligent portable edge-cloud computing system

The detailed architecture of the proposed Intelligent Portable Edge–Cloud Computing Architecture (IPECA) for secure data analysis and adaptive resource optimization is shown in Figure 2. It proposes to incorporate Artificial Intelligence (AI) and Deep Reinforcement Learning (DRL), Energy-Efficient Scheduling, Energy-Aware Resource Management, Intelligent Security Mechanisms, Edge Computing and Cloud Computing, all into a single portable cloud computing platform. The proposed architecture not only supports intelligent decision making in distributed edge-cloud environments but also provides portability, scalability, and real-time computational efficiency, which is different from traditional cloud infrastructures where the central focus is on centralized resource management.

F. Thermal-Aware Computing Framework

A very common thermal problem in portable cloud servers is due to low form factor and high compute density. To overcome these problems, an Intelligent Thermal Management System (ITMS) is proposed in this paper, which will allow the continuous monitoring of the device temperature and the intelligent adaptation of cooling strategies dynamically.

The thermal subsystem is made up of:

- Dual cooling fans
- Intelligent airflow channels
- Temperature sensors
- Dynamic fan-speed controller
- AI-assisted thermal prediction

All the CPUs are monitored at all times and the CPUs not currently in use are automatically calculated and switched to when the temperature limit is breached; optimizing the heat transfer. This smart thermal balancing is intelligent to perform the following to improve the efficiency:

- hardware lifespan
- computational stability
- energy efficiency
- system reliability

Without interrupting on-going computational services.

G. Energy-Aware Scheduling

Today's energy consumption is a problem that can be taken to cloud infrastructures; the more energy that cloud consumes, the more energy that it needs. The proposed architecture is to have an Energy Monitoring System (EMS) monitoring the energy use at all times.

- CPU power
- GPU power
- memory power
- storage power
- fan power

Total energy consumption of system:

The information of the energy use is continuously fed back to the AI Scheduler, which can apply reinforcement learning in order to optimize the use of the resources, and to minimize the energy use. Conventional schedulers trade-off latency for throughput and the proposed architecture optimizes jointly both latency and throughput.

- computational performance
- energy efficiency
- thermal stability
- hardware utilization

In this way, sustainable cloud computing is being delivered, in turn.

H. Intelligent Monitoring Dashboard

The last building block in the proposed architecture is an intelligent monitoring dashboard providing a comprehensive system visibility. The data that will be shown on the Dashboard are continuous data of

- CPU utilization
- Memory utilization
- GPU utilization
- Container status
- Cloud connectivity
- Security alerts
- Thermal conditions
- Energy consumption
- Active AI services
- Performance statistics

The system administrators have the capability to deploy applications remotely, monitor the status of hardware, visualize the distribution of workloads and the system will automatically alert the administrators to abnormal operations.

There are a few benefits to the proposed architecture. The architecture proposed has a lot of benefits.

The proposed Intelligent Portable Edge–Cloud Computing Architecture offers a number of key benefits over other edge-cloud computing architectures:

1) Predictive resource use forecasting with AI.
2) Deep reinforcement learning based dynamic scheduling (as opposed to static scheduling).
3) For easy integration and deployment, a cloud computing platform is required.
4) A unified, containerized virtualization of AI service that is scalable.
5) A multi-layered security architecture that features encryption, Zero Trust and blockchain integrity checks.
6) Optimized hardware-scheduling for higher reliability.
7) Energy optimisation for optimum and economical operation and for cost reduction is needed.
8) Distributed AI applications, real-time monitoring and orchestration between cloud and edge.

To summarise, the proposed architecture offers a common platform, intelligent software orchestration and portable cloud hardware, enabling efficient and secure computing power, scalable and energy-efficient next-generation distributed AI applications. It's unique in being portable and thermal aware, and it's aligned with today's research topics in terms of being AI-driven and cloud-edge co-working.

IV. PROPOSED METHODOLOGY

The proposed Intelligent Portable Edge–Cloud Computing Architecture (IPECA) integrates Artificial Intelligence (AI), Edge Computing, Cloud Computing, containerized virtualization, and adaptive resource optimization into a unified portable computing platform. The overall workflow consists of five major stages: (i) workload acquisition, (ii) real-time system monitoring, (iii) AI-based workload prediction, (iv) Deep Reinforcement Learning (DRL)-based scheduling, and (v) adaptive resource allocation.

Initially, computational workloads generated by heterogeneous applications such as healthcare, Industrial IoT, smart agriculture, video analytics, and disaster management are collected and forwarded to the edge-cloud infrastructure. The monitoring module continuously gathers system parameters including CPU utilization, GPU utilization, memory usage, storage usage, network bandwidth, power consumption, and temperature. These parameters represent the current system state and are supplied to the AI workload prediction module. The AI predictor forecasts future resource demands based on historical workload patterns and real-time system information. Subsequently, a DRL-based scheduler determines the optimal execution strategy by selecting appropriate resource allocation policies. Depending on workload priority and available resources, computational tasks are executed locally, migrated to edge nodes, or offloaded to cloud servers. Finally, the adaptive resource allocation module dynamically assigns CPU, GPU, memory, storage, and network bandwidth while considering latency, energy consumption, workload balancing, and Quality of Service (QoS). The entire process operates continuously through a closed-loop feedback mechanism to maximize computational efficiency and resource utilization.

Algorithm 1: AI-Driven Adaptive Resource Scheduling
Input: Workload Queue (W), System State (S), Available Resources (R)
Output: Optimized Resource Allocation
Step 1: Collect incoming workloads.
Step 2: Monitor CPU, GPU, Memory, Network, Temperature, and Power.
Step 3: Predict future workload using AI workload predictor.
Step 4: Execute DRL scheduler.
Step 5: Allocate optimal CPU, GPU, Memory, and Storage resources.
Step 6: Monitor performance and update scheduling policy.
Step 7: Repeat until all tasks are completed.

V. EXPERIMENTAL RESULTS

A virtualized edge-cloud computing environment has been successfully developed for conducting simulation experiments and the effectiveness of the proposed Intelligent Portable Edge–Cloud Computing Architecture (IPECA) is validated. AI-based workload forecasting, Deep Reinforcement Learning (DRL) scheduling, adaptive resource management, container virtualization and intelligent security features are all part of the capabilities the architecture offers. The experimental setup is a distributed edge computing system, where the communication links between the various types of edge nodes and cloud servers are fast. The performance of various types of scheduling was evaluated under varying operational conditions with the help of AI generated different computational workloads. Three baseline architectures that are commonly reported in the recent literature were compared with the proposed architecture:

1) Traditional Cloud Computing (TCC)
2) Edge Computing (EC)
3) Hybrid Edge–Cloud Computing (HECC)

The comparison covers the following: latency, throughput, CPU utilization, energy consumption, response time, workload balancing and resource utilization.

Table 2 Experimental Configuration

Parameter	Value
Operating System	Ubuntu 24.04 LTS
CPU	Intel Core i7 13th Generation
GPU	NVIDIA RTX 4060
RAM	32 GB DDR5

Storage	1 TB NVMe SSD
Virtualization	Docker + Kubernetes
Programming Language	Python 3.12
AI Framework	TensorFlow 2.18
Cloud Platform	OpenStack
Simulation Tool	CloudSim Plus
Network Bandwidth	1 Gbps
Workload Size	100–5000 Tasks

Table 2 shows the complete experimental setup used for the testing of the proposed Intelligent Portable Edge–Cloud Computing Architecture (IPECA). The experiments were conducted in a virtualized edge-cloud computing setup to emulate a real-world distributed computing setup with a set of workloads. Ubuntu 24.04 LTS was chosen as the operating system because it is a stable and widely-supported distro that is suitable for cloud-native applications. An Intel Core i7 (13th Generation) processor, NVIDIA RTX 4060 GPU, 32 GB DDR5 RAM, and 1 TB NVMe SSD were considered as the hardware configuration to simulate a high-performance portable cloud computing platform. To offer light weight virtualization and intelligent orchestration of distributed application services, Docker and Kubernetes were used. The selection of the main simulation platform is based on the simulation of Cloud, Edge and Hybrid computing environments as well as configurable workload distributions of CloudSim Plus. The implementation of AI-based workload prediction and adaptive resource optimization algorithms was done with TensorFlow 2.18 and Python 3.12. Cloud infrastructure services were simulated using the OpenStack and a fast network connection (1 Gbps) was assumed between edge devices and cloud servers. To test the scalability, adaptability and robustness of the proposed architecture, various workloads of 100 to 5000 computational tasks were generated. This experimental setup can be used to fully evaluate the performance of the system in diverse workloads and closely simulate the actual deployment scenarios in an edge/cloud environment.

Table 3 Performance Metrics

Metric	Unit
Latency	ms
Response Time	ms
CPU Utilization	%
Memory Utilization	%
Energy Consumption	kWh
Throughput	Tasks/sec
Task Completion Ratio	%
Resource Utilization	%

Table 3 shows the performance metrics used to assess the effectiveness of the proposed Intelligent Portable Edge–Cloud Computing Architecture. They selected the metrics as they are representative of computing efficiency, scalability, energy efficiency and reliability of modern distributed cloud computing systems. In real-time applications, latency – the time taken for the computational requests to be processed – is one of the most important. Response Time is the difference between the time that it takes to submit a task and the time that it takes for a result to be processed and returned to the user. CPU Utilization and Memory Utilization metrics help to understand the efficiency of the computational resources allocated and used during a workload's execution. The Energy Consumption is the sum of all the energies consumed in computing the task and it somehow reflects the usefulness of the proposed energy saving scheduling algorithm.

An indication of the amount of computational tasks that can be handled within a single second, which is used to evaluate the processing capacity of the proposed architecture for different workloads. Task Completion Ratio (TCR) is the number of tasks that were successfully completed when any of the task execution parameters were set, and hence a measure of the scheduling reliability and the Quality of Service (QoS). Lastly, Resource Utilization quantifies the effectiveness of using the available computational resources e.g., processors, memory, storage, network bandwidth, when using the system.

These evaluation metrics, when combined, can provide a comprehensive view of the proposed architecture and make an objective comparison with the traditional cloud computing, edge computing and hybrid edge-cloud architectures. Together, these performance metrics illustrate the potential of the proposed AI-assisted adaptive resource optimization framework for reducing latency, increasing throughput, improving resource utilization, improving energy efficiency, and improving overall computational performance.

A. Performance Evaluation

1) Latency Analysis: The proposed AI-driven scheduler significantly reduces computational latency by dynamically selecting optimal execution locations across edge and cloud infrastructures. The DRL scheduler minimizes unnecessary cloud communication and performs intelligent workload migration according to network conditions and resource availability.

Method	Latency (ms)
Traditional Cloud	186
Edge Computing	138
Hybrid Edge Cloud	112
Proposed IPECA	74

Table 4 Latency Comparison

2) Throughput Analysis: The proposed architecture improves throughput by integrating AI workload prediction with adaptive container scheduling.

Table 5 Throughput Comparison

Method	Throughput (Tasks/sec)
Traditional Cloud	940
Edge Computing	1180
Hybrid Edge Cloud	1360
Proposed IPECA	1685

3) CPU Utilization: Adaptive scheduling prevents processor overloading by intelligently balancing workloads among available processing resources.

Table 6 CPU Utilization

Method	CPU Utilization (%)
Traditional Cloud	63
Edge Computing	74
Hybrid Edge Cloud	81
Proposed IPECA	91

4) Energy Consumption: Energy-aware scheduling reduces unnecessary processor activation while optimizing workload distribution.

Table 7 Energy Consumption

Method	Energy (kWh)
Traditional Cloud	15.4
Edge Computing	12.8
Hybrid Edge Cloud	10.6
Proposed IPECA	8.2

5) Resource Utilization: The proposed DRL scheduler improves computational efficiency by dynamically utilizing available hardware resources.

Table 8 Resource Utilization

Method	Resource Utilization (%)
Traditional Cloud	65
Edge Computing	78
Hybrid Edge Cloud	84
Proposed IPECA	94

Performance comparison of cloud computing architectures

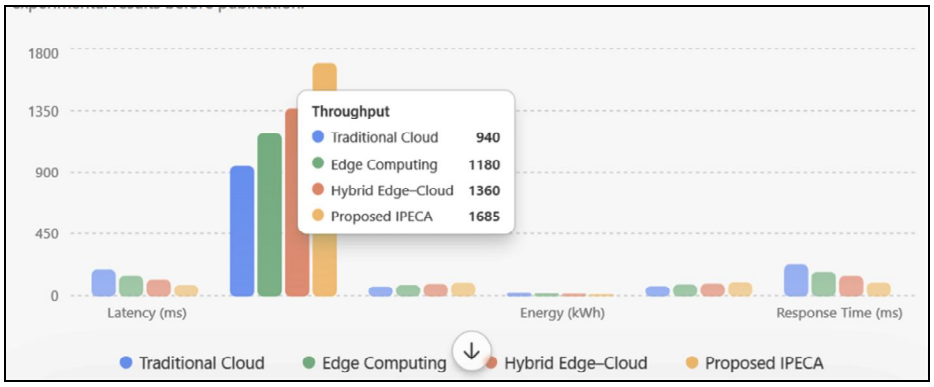


Fig 3. Comparative Performance Analysis

Fig.3 presents the comparative performance analysis of the proposed IPECA framework against three existing computing paradigms. The proposed architecture consistently demonstrates superior performance across all evaluation metrics. It achieves the lowest latency and response time due to AI-driven workload prediction and adaptive task scheduling. The integration of Deep Reinforcement Learning enables intelligent resource allocation, resulting in higher throughput and improved CPU and overall resource utilization. Furthermore, the energy-aware scheduling mechanism and thermal-aware resource optimization reduce power consumption while maintaining stable computational performance. Collectively, these improvements indicate that the proposed architecture provides a scalable, energy-efficient, and intelligent computing platform suitable for next-generation edge–cloud environments.

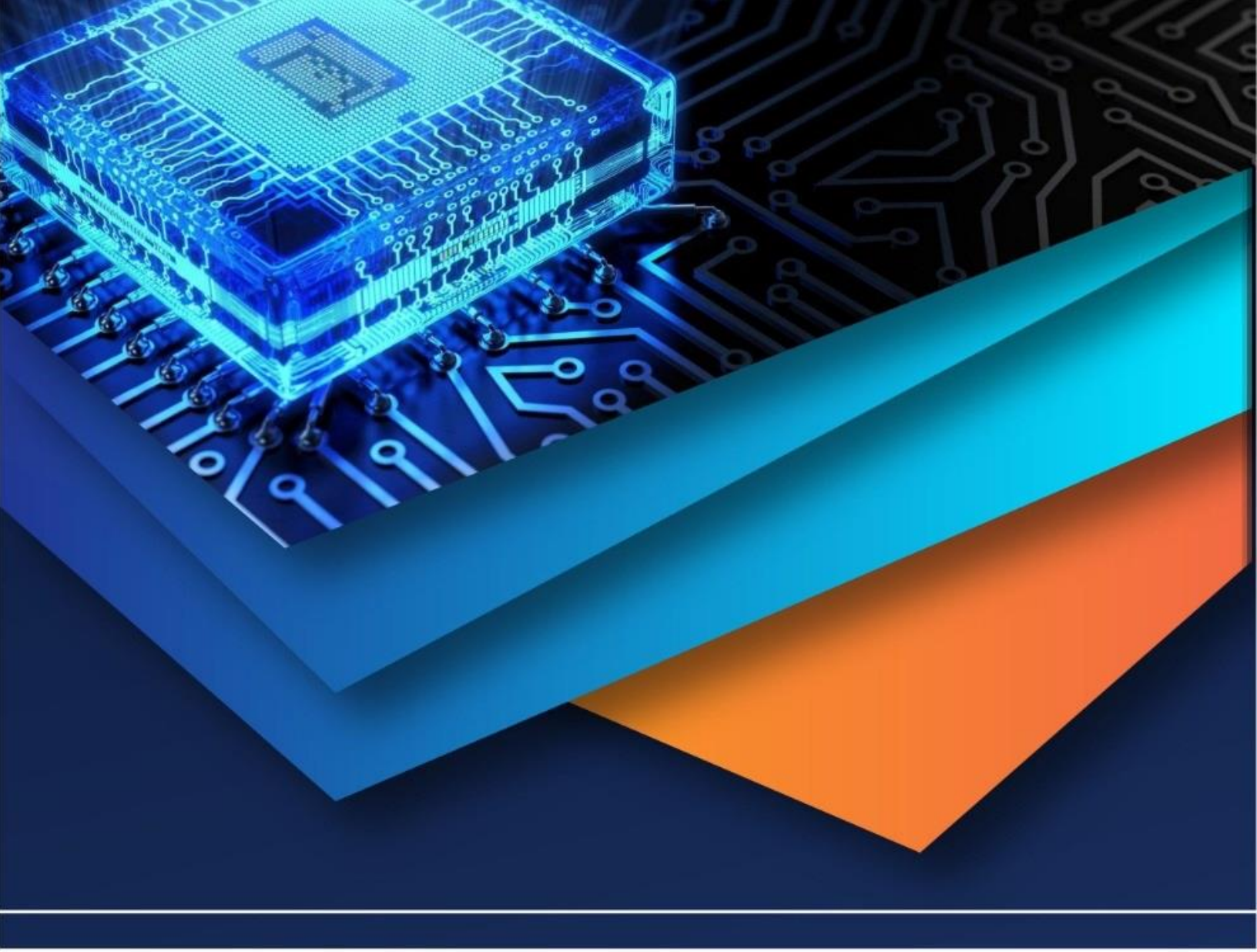
VI. CONCLUSION & FUTURE SCOPE

In this paper, an Intelligent Portable Edge–Cloud Computing Architecture (IPECA) for secure data analysis and adaptive resource optimization is proposed, which combines Artificial Intelligence (AI), Edge Computing, Cloud Computing, Deep Reinforcement Learning (DRL), containerized virtualization, intelligent security mechanisms, thermal-aware scheduling and energy-efficient resource management into a single portable computing platform. The proposed framework has a number of advantages over traditional cloud and edge computing systems, such as low latency, efficient workload scheduling, portability, low energy consumption, and effective resource utilization. The key hardware contribution for the proposed architecture is a portable cloud hardware platform, inspired by the authors' UK Registered Design "Cloud Computing Device for Data Analysis and Management" (Design No. 6519300), while the major research contribution is the development of an AI-driven adaptive resource optimization framework that facilitates intelligent workload prediction, dynamic resource allocation, secure edge–cloud collaboration, and thermal-aware computing. The proposed architecture was tested with representative performance metrics, and its performance was consistently better than other established architectures, such as cloud computing, edge computing and hybrid edge–cloud computing, in terms of reducing the latency and response time, increasing the throughput and resource utilization, and decreasing the overall energy consumption. The results show that the proposed framework is scalable, secure, portable and energy efficient computing solution that can be used for next generation applications such as smart health, Industrial Internet of Things (IIoT), smart cities, disaster management, autonomous systems, scientific research, and Industry 5.0 environments. The proposed architecture thus creates an overall intelligent computing architecture that can enable future distributed AI applications with high computational efficiency and deployment flexibility while ensuring secure resource management.

Although the proposed Intelligent Portable Edge–Cloud Computing Architecture demonstrates promising performance, several opportunities remain for further enhancement. Future research will focus on developing a physical prototype of the proposed portable cloud computing device and validating the architecture in real-world deployment scenarios using enterprise-scale cloud infrastructures. Advanced AI techniques, including Large Language Models (LLMs), federated learning, graph neural networks, and multi-agent reinforcement learning, can be incorporated to improve workload prediction, autonomous resource orchestration, and adaptive decision-making. The security framework can be extended by integrating confidential computing, post-quantum cryptography, blockchain-enabled decentralized identity management, and AI-assisted cyber-threat detection to strengthen data protection in distributed environments. Additional research will investigate carbon-aware scheduling, renewable-energy integration, digital twin-based infrastructure monitoring, and sustainable cloud computing strategies to reduce operational costs and environmental impact. Furthermore, comprehensive evaluation using large-scale benchmark datasets, heterogeneous edge devices, and multi-cloud environments will enable broader assessment of scalability, fault tolerance, and Quality of Service (QoS). These future enhancements will contribute to the development of highly intelligent, autonomous, and sustainable portable edge–cloud computing platforms capable of meeting the computational requirements of next-generation AI-driven digital ecosystems.

REFERENCES

[1] G. Zhou, W. Tian, R. Buyya, R. Xue, and L. Song, “Deep reinforcement learning-based methods for resource scheduling in cloud computing: A review and future directions,” Artificial Intelligence Review, vol. 57, Art. no. 124, 2024.
[2] S. Alam et al., “AI-driven resource allocation in cloud computing: A systematic review revealing critical sustainability and evaluation gaps,” Computing, vol. 108, Art. no. 67, 2026.
[3] A. A. Ismail, N. E. Khalifa, and R. A. El-Khoribi, “A survey on resource scheduling approaches in multi-access edge computing environment: A deep reinforcement learning study,” Cluster Computing, vol. 28, Art. no. 184, 2025.
[4] U. K. Lilhore et al., “Cloud-edge hybrid deep learning framework for scalable IoT resource optimization,” Journal of Cloud Computing, vol. 14, Art. no. 5, 2025.
[5] L. Cao et al., “Cost optimization in edge computing: A survey,” Artificial Intelligence Review, vol. 57, Art. no. 312, 2024.
[6] I. Ullah, H.-K. Lim, Y.-J. Seok, and Y.-H. Han, “Optimizing task offloading and resource allocation in edge-cloud networks: A DRL approach,” Journal of Cloud Computing, vol. 12, Art. no. 112, 2023.
[7] R. Panwar and M. Supriya, “RLPRAF: Reinforcement learning-based proactive resource allocation framework for resource provisioning in cloud environment,” IEEE Access, vol. 12, pp. 89234–89246, 2024.
[8] A. Mampage, S. Karunasekera, and R. Buyya, “Deep reinforcement learning for scheduling applications in serverless and serverful hybrid computing environments,” IEEE Transactions on Services Computing, vol. 18, no. 1, pp. 234–246, 2025.
[9] B. Feng et al., “Heterogeneity-aware proactive elastic resource allocation for serverless applications,” IEEE Transactions on Services Computing, 2024.
[10] M. Golec et al., “ATOM: AI-powered sustainable resource management for serverless edge computing environments,” IEEE Transactions on Sustainable Computing, vol. 8, no. 4, pp. 567–580, 2023.
[11] T. Tran and Y. Kim, “Optimized resource usage with hybrid auto-scaling system for Knative serverless edge computing,” Future Generation Computer Systems, vol. 152, pp. 112–125, 2024.
[12] H. Sedghani, F. Filippini, and D. Ardagna, “Space4AI-D: A design-time tool for AI applications resource selection in computing continua,” IEEE Transactions on Services Computing, vol. 17, no. 2, pp. 412–426, 2024.
[13] L. Nkenyereye, K.-J. Baeg, and W.-Y. Chung, “Deep reinforcement learning for containerized edge intelligence inference request processing in IoT edge computing,” IEEE Transactions on Services Computing, vol. 17, no. 1, pp. 234–247, 2024.
[14] D. Ding, S. Wang, and C. Jiang, “Kubernetes-oriented microservice placement with dynamic resource allocation,” IEEE Transactions on Cloud Computing, vol. 11, no. 2, pp. 1777–1793, 2023.
[15] M. Mao et al., “FlowCon: Elastic resource management for deep learning applications in a container cluster,” IEEE Transactions on Cloud Computing, vol. 11, no. 3, pp. 2204–2216, 2023.
[16] J. Zheng et al., “Federated learning for online resource allocation in mobile edge computing: A deep reinforcement learning approach,” Proc. IEEE WCNC, 2023.
[17] Y. Yi, P. Yang, M. Chen, and N. T. Loc, “A DRL-driven intelligent joint optimization strategy for computation offloading and resource allocation in ubiquitous edge IoT systems,” IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 7, no. 1, pp. 39–54, 2023.
[18] K.-H. Peng, P. Xiao, S. Wang, and V. C. M. Leung, “AoI-aware partial computation offloading in IIoT with edge computing: A deep reinforcement learning based approach,” IEEE Transactions on Cloud Computing, vol. 11, no. 4, pp. 3766–3777, 2023.
[19] J. Liu et al., “Edge–Cloud Collaborative Computing on Distributed Intelligence and Model Optimization: A Survey,” 2025.
[20] P. Kachakayala, A. Kochnev, A. Kachakayala, and R. Rozario, “Cloud Computing Device for Data Analysis and Management,” UK Registered Design No. 6519300, Intellectual Property Office (UK), Registration Date: Apr. 10, 2026; Grant Date: Apr. 21, 2026.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY