International Journal of Core Engineering & Management
Volume-7, Issue-12, 2024 ISSN No: 2348-9510

PREDICTIVE MACHINE LEARNING ALGORITHM TOWARDS OPTIMISED CLOUD COMPUTING APPLICATIONS

Siva Parise
Senior Software Automation engineer
Cherry Hill, New Jersey, USA
reachmrparise@gmail.com

Abstract

Cloud computing is a significant technical advancement in the field of information technology. Organising and distributing a large amount of data and resources via the Internet is among the finest methods available. Cloud computing, in technical terms, is the ability to access IT infrastructure over a computer network without requiring any installation on your own computer. Machine learning is being used more often by cloud service providers to lessen the requirement for human interaction in identifying and fixing security flaws. With its scalability, flexibility, and cost-effectiveness, cloud computing has emerged as a key component of contemporary IT infrastructure. However, maintaining cloud resource security and optimization is still quite difficult. This paper explores the application of ML algorithms in enhancing cloud security and optimization, with a focus on resource management, energy-efficient task scheduling, and fine-grained access control. The report examines current research in various fields, such as the creation of adaptive access control systems, resource optimization strategies for cloud computing, and energy-efficient job scheduling algorithms. This study highlights the importance of machine learning in enhancing cloud performance and security by comparing different techniques and identifying major strengths and problems. The incorporation of sophisticated ML algorithms for fault tolerance, dynamic resource allocation, and more effective cloud system management is highlighted in the paper's discussion of possible future research areas. The results provide light on how cloud computing is developing and give helpful recommendations for future developments in cloud security and optimization.

Keywords: Cloud Computing, Machine Learning, Cloud Security, Resource Optimisation, Energy-Efficient Scheduling.

I. INTRODUCTION

Today, cloud computing is an essential part of the contemporary information technology landscape that has imbued virtually all sectors of the global economy and society with new possibilities for the handling of data. It enables users to require computing resources over the web with scalability, flexibility, and affordability. Cloud computing today owes its growth to human factors, hardware improvements, computer networks, and related software elements that provide foundational support for delivering services as desired [1]. Given modern high-performance microprocessors, enormous storage systems, and fast networks, cloud computing allows organisations to delegate their IT services, decrease expenses, and enhance outcomes [2].

The basic concept of cloud computing is founded on a unique structure of protocols and standards, which make it possible for users and services to interface properly and harmonise their resources. Such protocols do make the physical computing infrastructure less tangible, creating multiple virtual layers that are manageable and scalable based on the need[3]. As cloud platforms expand, standardised interconnect protocols are crucial to the effective functioning of the cloud infrastructure [4]. All of them are useful in terms of managing tasks and issues like distribution of loads, direction of data, or even guaranteeing the adequacy of services, so cloud providers can provide efficient and secure services to their users.

Optimising resources to maximise performance and reduce costs is one of the main issues in cloud computing. Optimisation algorithms always come in handy in this particular area to facilitate the proportional distribution of cloud exigencies such as computing capacity, storage space and data transfer limits. These algorithms seek to learn how resources are being utilised and make real-time decisions on how best to use them, minimising resource wastage and enhancing resource utilisation. [5] Optimisation is useful in improving data center energy efficiency, forecasting when demand is at its highest, and handling multitenant environments that are crucial for cloud platform sustainably [6]. With cloud computing services growing in importance, the performance of the services needed to run these systems require sophisticated optimisation algorithms while still considering the impact on the environment [7].

Cloud computing has changed industries by offering essentially available and unlimited computing resources that may be used in various ways, from hosting websites to conducting artificial intelligence algorithms. This decentralisation means that certain key processes can be the focus of businesses, and cloud solutions provide continuous services such as storage and computing to various enterprises [8]. The use of new generation applications such as cloud-based services is leading to increased uptake of cloud technology. It noted that there is growing dependency on cloud computing where ML is employed to support functions like resource orchestration [9], security and performance. The ML models detect patterns in the input data and make pronunciations of the trends and, more importantly, the productivity of available resources and their utilisation in real time [10][11]. Also, ML improves Cloud security as it identifies threats and exceptions to the system [12]. The adoption of predictive ML algorithms in cloud computing is expected to enhance service resource utilisation efficiency, security, as well as workload scalability, which will lead to the sustainability and profitability of the cloud platforms.

II. OVERVIEW OF CLOUD COMPUTING

Cloud computing has been a game-changing technology in the information technology sector throughout the last ten years. It signifies a fundamental change in how end users receive and utilise IT services. Storage, processing power, and apps are just a few of the many computing resources made accessible over the Internet in cloud computing, which operates on a pay-as-you-go model [13]. Through virtualised computing environments, end-users have access to computer resources that are separated from the physical infrastructure.

International Journal of Core Engineering & Management
Volume-7, Issue-12, 2024 ISSN No: 2348-9510

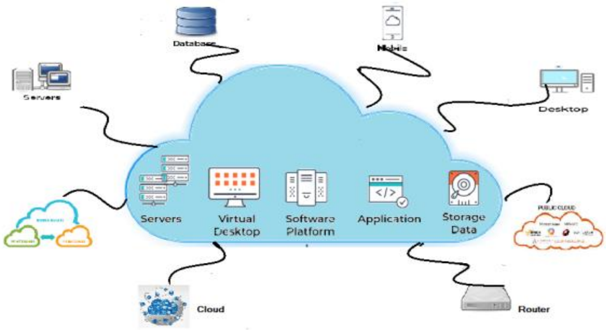

Fig. 1. Cloud Computing

Cloud computing is characterised by its scalability and the ability to be made available on demand. Without spending a fortune on costly physical infrastructure, end-users may consume and grow computer resources as needed [14]. Data stored in the cloud is subject to the care and management of the cloud provider. One of the main objectives of the cloud provider is, of course, data security [5]. As a wonderful bonus to adopting any innovation, information security is now offered. Undermining security risks the disastrous reality that might cause the technology's abandonment; however, keeping security would cause this technology to become very popular. Since the inception of grid computing, guidelines and regulations have been established to address any potential risk to data security.

A. Characteristics of Cloud Computing

Cloud computing research areas have been widely explored due to various services and functionalities provisioned by them. Most of the researchers prefer cloud computing for completing their tasks compared to any other services [15]. A pictorial representation of characteristics of cloud computing services is given in Figure 2. The essential characteristics of cloud that distinguishes it from other types of services are given below [1]:

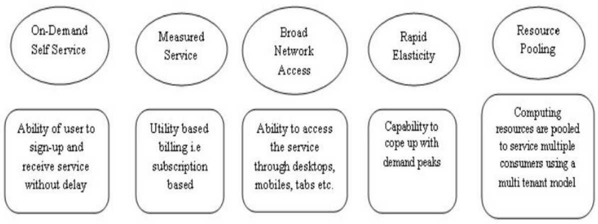

Fig. 2. Characteristics of cloud computing

- **On-demand self-service:** Cloud resources can be registered and utilised without human intervention with suppliers of cloud administration and the resources are virtual machines, storage, processing powers etc. [16].
- **Measured service:** Storage use, hours of CPU, usage of bandwidth and so on are measured through resource utilisation and these measurements are applied to the clouds but every cloud provider's uses a different abstraction level that can be alternate to an administration [17].
- **Broad network access:** The aforesaid resources can be obtained through the system, which utilises heterogeneous gadgets such as laptops or mobile phones.
- **Resource pooling:** Resources are pooled by cloud administration resources which are then shared with the clients. This is termed as multi tenure, where for instance, there can more than

International Journal of Core Engineering & Management
Volume-7, Issue-12, 2024 ISSN No: 2348-9510

virtual machines to a physical server with specific clients.

- **Rapid elasticity**: More resources can be gained rapidly from cloud through scaling out and scaling back and these resources are discharged when they are not needed

These characteristics of cloud computing services distinguish the cloud service providers from others.

III. CLOUD COMPUTING ARCHITECTURE AND SERVICES

Cloud computing architecture mainly focuses on configuring the system components which includes cloud resources, services, hardware, middleware and software, cloud consumers, cloud storage, and networks [18]. It is primarily involved in arranging all these components concerning the usage of the cloud consumers and the end-users. Cloud computing architecture is a modern concept derived based on the possibility of storing a huge amount of data and applications [19][20]. Also, it is related to providing these stored data and applications based on the demands of the customers along with flawless access to the hardware and software technologies without offering any significant investment for the own software, hardware, or infrastructure [21]. Figure 3 depicts the architecture of cloud computing and the components involved in the cloud computing architecture [22].

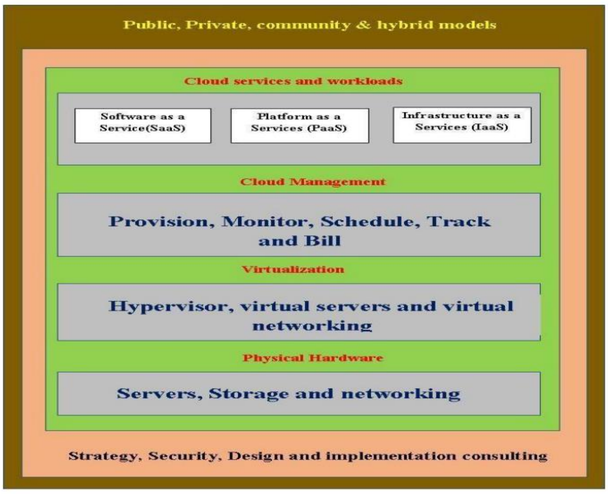

Fig. 3. Cloud Computing Architecture[23]

Figure 3 illustrates the architecture of cloud computing, highlighting the key layers and components involved. The cloud management layer focuses on the administration of resources, covering provisioning, monitoring, scheduling, tracking, and billing[24][25]. Virtualization plays a crucial role in efficiently allocating resources by creating virtual servers and networks, while the physical hardware layer underpins the infrastructure with servers, storage, and networking equipment. Additionally, the diagram emphasizes the importance of strategy, security, design, and implementation consulting to ensure effective cloud computing implementations. Overall, it provides a comprehensive view of the interaction between different layers, from deployment models to physical infrastructure.

Bulk processing solutions for the back office that operate on the web are only one of many applications that leverage cloud architectures.

- **Document Processing Pipelines:** Transfer documents from Word to PDF and make images and pages searchable.
- **Image Processing Pipelines**: Make video files encrypted in MPEG or AVI format, build an index that web may utilise, and search records using data mining [26].
- **Batch Processing System**: Used in industries like banking and insurance to analyse logs and generate daily/weekly reports.
- **Night Builds:** Automatically build source code at night and perform testing on multiple configurations.
- **Auto-Scaling Websites**: Scale up during the day and are redundant at night.
- **Seasonal/Instant Websites**: Created for specific events or operate seasonally (e.g., holidays or tax season).

A. Different Service Models of Cloud Computing

There are different types of service models for cloud end users to use on-demand services according to their requirements. As the cloud computing environment grows daily, the service models also grow. Figure 4 shows the service models of cloud.

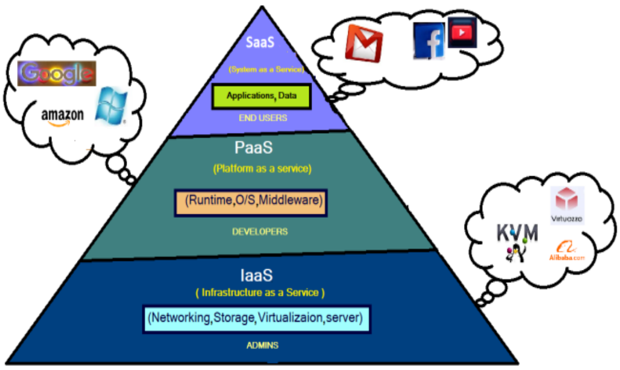

Fig. 4. Basic Service Models of Cloud Computing

1. *Infrastructure-as-a-service (IaaS)*

The cloud provides its customers with access to a shared pool of computing, storage, networking, and other IT resources and infrastructure. In the IaaS cloud, virtualisation is often used to merge or divide physical resources according to the fluctuating resource requirements of cloud customers.

2. *Platform as a service (PaaS)*

PaaS development environments facilitate the whole software lifecycle, allowing users to create cloud services and applications. Here they provide a development platform that differs from SaaS in that it hosts cloud applications in various stages of development, not just completed ones [27]. Applications that fall under the umbrella of PaaS include Google App Engine, Microsoft Azure, Java, and developer tools.

3. *Software as a service (SaaS)*

Programs published by cloud customers in a hosting environment are accessible to a range of clients with network access. By consolidating application users from different cloud providers into one logical environment, the SaaS cloud may optimise performance, availability, disaster recovery, maintenance, and security while also taking advantage of economies of scale [28]. Users do not have control over the cloud infrastructure, which often uses a multitenancy system design. Google,

International Journal of Core Engineering & Management
Volume-7, Issue-12, 2024 ISSN No: 2348-9510

Salesforce, and Google Docs are just a few examples.

B. Cloud Deployment models

The infrastructure of cloud is developed for its users and is owned according to cloud service provider. The high speed cloud infrastructure resides in a big land region. The usage of cloud services is based on a 'pay-per-use model' by Small and Medium Enterprises (SME) organizations or the general public. All resources under the cloud service deployment model are on-demand and dynamic in nature [29][30]. Figure 5 illustrates the cloud deployment models.

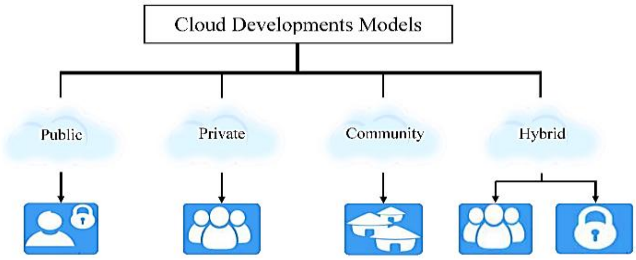

Fig. 5. Cloud deployment models [31]

1. *Private clouds*

A private cloud is an isolated instance of the cloud that is only accessible by the designated users of an enterprise. The virtualisation and provisioning of an organisation's servers, storage, and networking resources as a service is the essence of a private cloud [32]. A private cloud may be hosted either in-house by the company's data centre or by an external cloud service provider.

2. *Public clouds*

In public clouds, consumers do not typically control the underlying information technology infrastructure, but rather, the cloud service provider does. Microsoft Azure, Google Cloud, Alibaba Cloud, AWS, and IBM Cloud are among the most prominent public cloud providers [33].

3. *Hybrid clouds*

Hybrid clouds are a kind of IT infrastructure that consists of many separate but seemingly interconnected environments that are linked by various means like LANs, WANs, VPNs, and APIs [34]. A variety of needs may be applicable due to the complexity of hybrid cloud features.

4. *Community Cloud*

Community cloud providers essentially cater their services to those who have commonalities in terms of interests and issues. Concerns about security, privacy, etc., might be community-wide needs. Organisations or an outside entity (such as a network administrator) may oversee this cloud architecture [35].

IV. OPTIMIZATION CHALLENGES IN CLOUD COMPUTING

Algorithms for optimising cloud computing aim to improve cloud environments' efficiency, performance, and resource utilisation. These algorithms address challenges like cost minimization, task scheduling, resource allocation, energy consumption, and data management [36][37]. The purpose of any machine learning algorithm is to find the extreme value of some objective function, which is an optimisation issue. In machine learning, the first stage is to create models and logical goal functions. It is common practice to tackle optimisation problems using the defined objective function in conjunction with suitable numerical optimisation techniques.

There has been a lot of focus on task scheduling in cloud computing recently. Various scheduling techniques that function in a cloud computing setting have been suggested by many academics [38]. Despite the fact that the majority of suggested methods for scheduling activities are improvements to existing algorithms, the efficiency of user tasks and resource utilisation in a cloud environment are both impacted by task scheduling policies. Therefore, the primary challenge of cloud computing job scheduling is how to ensure optimum user task allocation [39].

A. Cloud optimization strategies

To maximise resource utilisation and cloud provider services, organisations may use a number of tactics. Among the techniques for cloud optimisation are:

- **Rightsizing:** The act of matching workloads to the amount and kind of cloud computing instances is known as rightsizing. By ensuring that a company buys the required number of cloud instances, rightsizing may assist in increasing cost effectiveness [40].
- **Auto scaling**: An essential component of cloud optimisation is automation. IT teams can no longer manually monitor and assign the appropriate resources for each task across many environments due to the growing use of hybrid and multi cloud systems [41].
- **Containerization:** Containerisation allows software to operate on any platform or cloud by combining the operating system, configuration files, libraries, and dependencies required to execute the code into a single "container." Compared with virtual machines (VMs), these containers are more effective and take less time to set up.
- **Data transfers**: Data is often transferred across clouds. Every transfer has a price, and making more transfers might result in higher cloud expenses [42]. Cloud management technologies may assist in minimising this expense by automating the process of effectively transferring data across clouds.
- **Cloud-native applications**: Applications that are cloud-native are composed of reusable parts called micro services [43][44]. Different components of an application's operation are controlled by tiny building pieces that make up a micro services architecture. Because micro services may be scaled and delivered independently of one another, they are adaptable and helpful for continuous delivery (CD) and continuous integration (CI) procedures.

B. Cloud Optimization Challenges

The cloud has revolutionised how businesses function and is now a crucial component of corporate operations. It makes it simpler for companies to scale and expand their operations by enabling them to store, access, and analyse data from any location. However, many organisations still struggle with cloud optimisation, which calls for meticulous administration, monitoring, and planning [45]. For organisations, cloud optimisation poses a number of challenges:

- **Cost Management:** The pay-as-you-go model of cloud services can quickly lead to unexpected costs if resources are not properly allocated. Monitoring tools and outsourcing to Managed Service Providers (MSPs) can help manage costs, although this may add extra expenses.
- **Complexity:** Managing cloud environments requires specialized skills. Over-optimization can lead to increased complexity, reducing performance and scalability. Balancing cost-saving efforts with performance, security, and scalability is key.
- **Security:** Cloud security is often not included by default and requires a multi-layered

approach. Many organizations face security challenges, with many experiencing incidents in the past year. Partnering with Managed Security Service Providers (MSSPs) can help mitigate risks without adding to internal IT staff costs[46].

- **Performance:** As workloads grow, cloud performance can degrade without optimisation. Automation and effective monitoring are crucial to maintain high performance and avoid issues, as visibility directly impacts business value[47].
- **Governance:** Cloud governance is often challenging due to complexity and lack of visibility. Organisations must implement robust governance policies to ensure compliance with regulatory requirements and best practices.

C. Optimization Algorithm in Cloud Computing

Several resource optimisation methods are used in cloud computing to optimise the resources in order to provide cloud consumers the best services possible. In this overview, the following algorithms were examined: Genetic Algorithm (GA), Ant Colony Optimisation Algorithm (ACO), Particle Swarm Optimisation Algorithm (PSO), and Bacterial Foraging Optimisation (BFO) [48]:

1. *Genetic Algorithm (GA)*

This algorithm is a simultaneous, effective, worldwide search method. The resource optimisation issue is a good fit for GA. The genetic algorithm has an improved feature called multiple-point searching. The search space is shrunk when resource optimisation is done using GA approaches.

2. *Ant Colony Optimization (ACO) Algorithm*

Optimal route finding using this randomised probabilistic approach is analogous to how ants locate food. In the cloud, various resources are made available on several virtual computers by a cloud service provider. They think of virtual computers as nodes, ants as agents, and resources as food. While circling the network, Ant distributes tasks to available cloud resources. [49].

3. *Particle Swarm Optimization (PSO) Algorithm*

A subset of swarm intelligence, it is a computational approach. To find the best solution, this method uses an iterative approach that is self-adaptive. The population of potential solutions is called "particles" in PSO. PSO makes use of particle repositioning inside the search space. Based on the particle's location and velocity, basic mathematical equations dictate the motion.

4. *Bacterial Foraging Optimization (BFO) Algorithm*

The bacterial foraging algorithm is based on the hyper-heuristics method [50]. In this algorithm, the partial result is denoted by a bacterium and the motion of the bacterium is heuristics. The optimisation in bacterial foraging algorithm lies in processes like chemotaxis, swarming, reproduction, elimination and dispersal.

V. MACHINE LEARNING TECHNIQUES FOR CLOUD OPTIMIZATION

Machine learning (ML) techniques play a pivotal role in optimising cloud computing environments by improving resource allocation, energy efficiency, performance, and cost-effectiveness. ML algorithms such as regression, reinforcement learning, and clustering predict resource demands based on historical usage patterns, enabling dynamic scaling and load balancing to prevent under-provisioning or over-provisioning [51]. Additionally, ML models optimise energy consumption by forecasting power usage and adjusting workloads to reduce costs and environmental impact [52]. Auto-scaling, powered by predictive analytics, allows cloud providers to proactively adjust resources to handle traffic spikes, enhancing efficiency. ML-driven cost optimisation models analyse usage patterns to recommend the most cost-effective resource allocation and pricing strategies [53][54]. Furthermore, ML facilitates real-time performance monitoring and anomaly detection, identifying potential system failures or security threats early and enabling proactive responses [55]. Ultimately, the integration of machine learning in cloud optimisation not only enhances performance and reliability but also ensures more efficient, scalable, and sustainable cloud computing operations [56].

A. Benefits of ML in cloud computing

Cloud computing is expanding quickly, and more customers are following suit to take advantage of its advantages [57]. Cloud service providers are under the heat to meet this increased demand while also ensuring their customers' utmost safety. Under these conditions, they would benefit from automation or rapidly developing technologies that reduce the need for human intervention, allowing for more rapid response at lower cost.

- **Automation:** Detecting and blocking harmful network traffic automatically via the use of ML algorithms that analyse the traffic and find patterns that suggest a possible security risk.
- **Scalability:** By analysing data from many sources, including logs, network traffic, and system metrics, and detecting patterns that suggest when more resources are required, Machine Learning algorithms may automatically scale cloud resources depending on consumption patterns [58].
- **Adaptability:** Algorithms trained using Machine Learning may learn to spot new and emerging dangers by adjusting to trends in the data.
- **Proactivity:** Taking precautions to lessen the likelihood of a successful attack by using a supervised learning system to forecast possible security breaches [59][60].
- **Efficiency:** A more efficient and cost-effective cloud environment is possible with the use of ML algorithms applied to cloud infrastructure and resources like storage accounts, virtual machines, and network bandwidth.
- **Accurate and unbiased:** ML algorithms are able to sift through mountains of data in search of patterns that people would miss, resulting in more accurate security evaluations [61].
- **Personalization:** Profiles of people, either individually or collectively, may be constructed using ML algorithms to record their unique patterns of behaviour [62].

B. Applications of predictive ml in cloud computing

One of the most important ways that predictive machine learning in the cloud is improving decision-making and reshaping whole sectors is via the following:

- **Demand Forecasting:** Predicting resource usage to optimise cloud infrastructure and reduce operational costs.
- **Security:** With cybersecurity at the top of every business's agenda, it should come as no surprise that predictive analytics and Machine Learning play a key part in security. Predictive analytics is a common tool for security organisations to boost performance and service quality [63].
- **Security Enhancement:** Detecting potential cyber threats, anomalies, and unauthorised access in real time [64].
- **Customer Analytics**: Anticipating customer behaviour, such as churn or purchasing patterns, for personalised experiences.

- **IT Operations**: Enabling predictive maintenance by identifying potential hardware failures or system downtimes [65].
- **Healthcare:** Supporting cloud-based diagnostic tools by predicting patient outcomes using large datasets [66][67][68].
- **Financial Services:** Predicting market trends, credit risks, and fraudulent activities with high accuracy.
- **Energy Optimization**: Forecasting energy consumption patterns in data centres for sustainable operations.

VI. LITERATURE REVIEW

In this section, look at publications that address the subject of cloud security employing ML techniques. Then, examine how the associated publications compare to ours.

In this study, Khullar and Hossain (2022) a novel method for green cloud operation that schedules tasks with minimal energy use. To determine the sequence of tasks and allocate them to the processing nodes, the suggested method makes use of Dynamic Voltage and Frequency Scaling and DVFS enabled Efficient Energy Workflow Task Scheduling. The technique facilitates the use of CPU idle slots without going against overall process limitations. The deadline is calculated based on the task's heterogeneous end time, and slack time is maximised by combining inefficient processors. Instead of using virtual machines, containerisation is utilised with Kubernetes as a package, which provides benefits over Docker Swarm [69].

In this study, Gorantla et al. (2024) the optimisation algorithms developed for cloud computing management are targeted towards improving resource utilisation, failure handling, and load balancing. Furthermore, the intelligent use of techniques such as fault tolerance, capacity planning, and quality of service guarantees ensures that the users' experience within the cloud is of optimal quality. Parametric results from research on Cloud Computing management algorithm performance showed that, when correctly implemented, the algorithms can provide significant improvements to the management and performance of high-traffic networks [70].

In this study, Mengistu, Che and Lu (2019) three heuristic-based techniques to address the goals and limitations unique to Volunteer Cloud Computing. The VM placement issue in Volunteer Cloud Computing is a bounded 0-1 multi-dimensional knapsack problem. Test findings on a genuine Volunteer Cloud Computing testbed provide empirical confirmation of these algorithms' competitive performance [71].

This study, Patel and Halabi (2021) the use of cloud computing is growing. There is a new way to cache data in these apps using Redis, an in-memory data store, to improve QoS; many organisations have opted to move their workload to the cloud because of its scalability, resilience, and cost-reduction properties. An investigation reveals that caching improves performance by speeding up data retrieval by a factor of four, as measured by performance parameters including response time, throughput, latency, and number of hits [72].

This study, Shiftehfar, Mechitov and Agha (2014) continues to work on creating a versatile, granular access control mechanism for contemporary cloud-based apps. Programmers can't create

International Journal of Core Engineering & Management
Volume-7, Issue-12, 2024 ISSN No: 2348-9510

the application logic for modern cloud apps before the organisations that hold the data and resources set the necessary authorisation requirements. This simplifies the creation of cloud applications and makes them more adaptable to any changes in policy in the future, but it also shows how important it is to have an authorisation system that can react effectively [73].

In this study, Tandri, Nuha and Utomo (2023) so that a mobile app named Hening may have an API built on top of cloud computing. Accessing and using the power of Cloud Computing is made possible via the Hening application's API, which is a programming interface. This API may reduce the load on mobile devices and increase the efficiency of resource consumption by performing many operations and moving the weight of data processing and storage to Cloud Computing using the Express.js framework. The Hening API is put into action by use of PaaS technology [74].

Table I provides a comparison of related studies on cloud optimisation and security, focusing on their objectives, challenges, strengths, and potential future directions. It highlights the different approaches taken by each study, ranging from energy-efficient task scheduling to fine-grained access control and resource management. The Table II also emphasises areas for future research, particularly the integration of machine learning techniques in cloud environments.

TABLE I. COMPARISON OF RELATED STUDIES ON CLOUD OPTIMIZATION ALGORITHMS AND SECURITY

Study	Focus	Objectives	Challenges	Strengths	Future Study
Khullar & Hossain, 2022	Energy-efficient task scheduling in green cloud operations	Develop an energy-efficient algorithm for task scheduling using DVFS and Kubernetes	Optimising slack time and idle CPU usage without violating constraints	Efficient task scheduling; use of Kubernetes for better scalability and management	Further exploration into hybrid cloud architectures for improved energy savings.
Gorantla et al., 2024	Cloud computing resource management optimisation	Improve resource utilisation, fault tolerance, and load balancing in cloud computing.	Managing high-traffic networks; balancing cost vs. performance	Optimises cloud performance; focuses on high-traffic networks	Development of more advanced fault-tolerance algorithms incorporating AI and ML.
Mengistu, Che & Lu, 2019	Virtual Machine Placement in Volunteer Cloud Computing	Use heuristic algorithms to solve the Virtual Machine placement problem	Computational complexity in multi-dimensional optimisation	Effective in real-world test-bed environments; competitive performance results	Investigation into the use of ML for more dynamic placement strategies.
Patel & Halabi, 2021	Cloud computing adoption and data caching	Enhance Quality of Service using Redis caching for cloud applications	Ensuring low latency while scaling data caching solutions	Significant performance improvement in response time and throughput	Expanding caching techniques with AI-driven caching optimisations.
Shiftehfar, Mechitov & Agha, 2014	Fine-grained access control for cloud-based applications	Develop adaptive access control systems for modern	Balancing flexibility with the complexity of dynamic policies	Offers flexibility for policy changes and simplifies application	Exploring machine learning methods to enhance adaptive access

		cloud environments		development	control.
Tandri, Nuha & Utomo, 2023	Cloud Computing-based API development for mobile apps	Develop API to offload data processing and storage to the cloud	Managing mobile device load while ensuring efficient data transfer	Reduces mobile device load and improves cloud resource utilisation	Investigating the integration of AI for better API management and resource optimisation.

VII. CONCLUSION AND FUTURE SCOPE

"Cloud" is a relatively new term for a set of technologies that have been around for quite some time. It takes a lot of money, time, and expertise to go into machine learning and AI. IaaS must be a game-changing innovation in the cloud. Service offerings come from three main providers: AWS, Azure, and GCP. The most impressive thing is how easy and convenient it is. With an emphasis on energy economy, resource management, and security, this research explored the use of ML algorithms to optimise cloud computing systems. The reviewed studies illustrate the diverse ways ML techniques are being employed to enhance cloud performance, such as improving task scheduling, fault tolerance, and load balancing. While these techniques have demonstrated significant benefits, including better resource utilisation and faster data retrieval, challenges remain in terms of scalability and computational complexity. The integration of ML into cloud environments holds immense potential for future research, particularly in areas such as dynamic resource allocation, AI-based caching optimisations, and the evolution of flexible access control systems. As cloud computing continues to evolve, the adoption of ML will be key to unlocking smarter, more efficient, and secure cloud infrastructures.

REFERENCES

1. C. Gong, J. Liu, Q. Zhang, H. Chen, and Z. Gong, "The characteristics of cloud computing," Proc. Int. Conf. Parallel Process. Work., pp. 275–279, 2010, doi: 10.1109/ICPPW.2010.45.
2. R. Arora, S. Gera, and M. Saxena, "Impact of Cloud Computing Services and Application in Healthcare Sector and to provide improved quality patient care," IEEE Int. Conf. Cloud Comput. Emerg. Mark. (CCEM), NJ, USA, 2021, pp. 45–47, 2021.
3. A. Lopez Garcia et al., "A cloud-based framework for machine learning workloads and applications," IEEE Access, vol. 8, pp. 18681–18692, 2020, doi: 10.1109/ACCESS.2020.2964386.
4. A. P. Pandian, X. Fernando, and S. M. S. Islam, "Computer Networks, Big Data and IoT," Proc. ICCBI 2020, vol. 66, no. June, 2020, doi: 10.1007/978-981-16-0965-7.
5. X. S. Yang, "Optimization and Metaheuristic Algorithms in Engineering," Metaheuristics Water, Geotech. Transp. Eng., no. January 1970, pp. 1–23, 2013, doi: 10.1016/B978-0-12-398296-4.00001-5.
6. R. Arora, S. Gera, and M. Saxena, "Mitigating Security Risks on Privacy of Sensitive Data used in Cloud-based ERP Applications," in 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), 2021, pp. 458–463.
7. A. P. A. Singh, "Streamlining Purchase Requisitions and Orders: A Guide to Effective Goods Receipt Management," J. Emerg. Technol. Innov. Res., vol. 8, no. 5, pp. g179–g184, 2021.
8. S. K. R. Anumandla, V. K. Yarlagadda, S. C. R. Vennapusa, and K. R. V Kothapalli, "Unveiling the Influence of Artificial Intelligence on Resource Management and Sustainable Development:

A Comprehensive Investigation," Technol. \& Manag. Rev., vol. 5, no. 1, pp. 45–65, 2020.
9. R. Goyal, "An Effective Machine Learning Based Regression Techniques For Prediction Of Health Insurance Cost," Int. J. Core Eng. Manag., vol. 7, no. 11, pp. 49–60, 2024.
10. P. Nawrocki and M. Smendowski, "Optimization of the Use of Cloud Computing Resources Using Exploratory Data Analysis and Machine Learning," J. Artif. Intell. Soft Comput. Res., vol. 14, no. 4, pp. 287–308, 2024, doi: 10.2478/jaiscr-2024-0016.
11. J. Thomas, K. V. Vedi, and S. Gupta, "Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics," J. Emerg. Technol. Innov. Res., vol. 8, no. 9, 2021.
12. Mani Gopalsamy, "An Optimal Artificial Intelligence (AI) technique for cybersecurity threat detection in IoT Networks," Int. J. Sci. Res. Arch., vol. 7, no. 2, pp. 661–671, Dec. 2022, doi: 10.30574/ijsra.2022.7.2.0235.
13. R. K. Arora, Mohd.Muqeem, and M. Saxena, "Developing a Comprehensive Security Framework for Detecting and Mitigating IoT device Attack." Sep. 2024. doi: 10.21203/rs.3.rs-5165811/v1.
14. K. Patel, "A Review on Software Quality Assurance (QA): Emerging Trends and Technologies," Int. J. Tech. Innov. Mod. Eng. Sci., vol. 10, no. 10, pp. 9–14., 2024.
15. V. K. Y. Nicholas Richardson, Rajani Pydipalli, Sai Sirisha Maddula, Sunil Kumar Reddy Anumandla, "Role-Based Access Control in SAS Programming: Enhancing Security and Authorization," Int. J. Reciprocal Symmetry Theor. Phys., vol. 6, no. 1, pp. 31–42, 2019.
16. D. Puthal, B. P. S. Sahoo, S. Mishra, and S. Swain, "Cloud computing features, issues, and challenges: A big picture," in Proceedings - 1st International Conference on Computational Intelligence and Networks, CINE 2015, 2015. doi: 10.1109/CINE.2015.31.
17. B. P. Kishore Mullangi, Niravkumar Dhameliya, Sunil Kumar Reddy Anumandla, Vamsi Krishna Yarlagadda, Dipakkumar Kanubhai Sachani, Sai Charan Reddy Vennapusa, Sai Sirisha Maddula, "AI-Augmented Decision-Making in Management Using Quantum Networks," Asian Bus. Rev., vol. 13, no. 2, pp. 73–86, 2023.
18. B. Patel, V. K. Yarlagadda, N. Dhameliya, K. Mullangi, and S. C. R. Vennapusa, "Advancements in 5G Technology: Enhancing Connectivity and Performance in Communication Engineering," Eng. Int., vol. 10, no. 2, pp. 117–130, 2022, doi: 10.18034/ei.v10i2.715.
19. L. Wang, R. Ranjan, J. Chen, and B. Benatallah, Cloud computing: Methodology, systems, and applications. 2017. doi: 10.1201/b11149.
20. A. P. A. S. and N. Gameti, "Digital Twins in Manufacturing: A Survey of Current Practices and Future Trends," Int. J. Sci. Res. Arch., vol. 13, no. 1, pp. 1240–1250, 2024.
21. V. K. Yarlagadda and R. Pydipalli, "Secure Programming with SAS: Mitigating Risks and Protecting Data Integrity," Eng. Int., vol. 6, no. 2, pp. 211–222, Dec. 2018, doi: 10.18034/ei.v6i2.709.
22. R. Arora, A. Kumar, A. Soni, and A. Tiwari, "AI-Driven Self-Healing Cloud Systems: Enhancing Reliability and Reducing Downtime through Event-Driven Automation," 2024, doi: 10.20944/preprints202408.1860.v1.
23. N. R, "Enhanced data integrity and authenticity of dynamically stored data in cloud," 2019.
24. A. P. A. Singh and N. Gameti, "Leveraging Digital Twins for Predictive Maintenance: Techniques, Challenges, and Application," IJSART, vol. 10, no. 09, pp. 118–128, 2024.
25. K. Patel, "AN ANALYSIS OF QUALITY ASSURANCE FOR BUSINESS INTELLIGENCE PROCESS IN EDUCATION SECTOR," IJNRD - Int. J. Nov. Res. Dev., vol. 9, no. 9, pp. a884–

a896, 2024.
26. Muthuvel Raj Suyambu and Pawan Kumar Vishwakarma, "Improving Efficiency of Electric Vehicles: An Energy Management Approach Utilizing Fuzzy Logic," Int. J. Adv. Res. Sci. Commun. Technol., vol. 3, no. 2, pp. 737–748, Feb. 2023, doi: 10.48175/IJARSCT-9749V.
27. Sahil Arora and Apoorva Tewari, "Fortifying Critical Infrastructures: Secure Data Management with Edge Computing," Int. J. Adv. Res. Sci. Commun. Technol., vol. 3, no. 2, pp. 946–955, Aug. 2023, doi: 10.48175/IJARSCT-12743E.
28. M. Gopalsamy, "Predictive Cyber Attack Detection in Cloud Environments with Machine Learning from the CICIDS 2018 Dataset," IJSART, vol. 10, no. 10, 2024.
29. H. Hourani and M. Abdallah, "Cloud Computing: Legal and Security Issues," in 2018 8th International Conference on Computer Science and Information Technology, CSIT 2018, 2018. doi: 10.1109/CSIT.2018.8486161.
30. Y. ChakradharaRao Ch, Mogasala Leelarani & Ramesh Kumar, "Cloud: Computing Services And Deployment Models," J. Eng. Comput. Sci., vol. 2, no. 12, pp. 3389–3392, 2013.
31. U. A. Butt et al., "A review of machine learning algorithms for cloud computing security," Electronics (Switzerland). 2020. doi: 10.3390/electronics9091379.
32. A. P. A. Singh, "STRATEGIC APPROACHES TO MATERIALS DATA COLLECTION AND INVENTORY MANAGEMENT," Int. J. Bus. Quant. Econ. Appl. Manag. Res., vol. 7, no. 5, 2022.
33. Sahil Arora and Apoorva Tewari, "Zero trust architecture in IAM with AI integration," Int. J. Sci. Res. Arch., vol. 8, no. 2, pp. 737–745, Apr. 2023, doi: 10.30574/ijsra.2023.8.2.0163.
34. P. K. Vishwakarma and M. R. Suyambu, "An Analysis of Engineering, Procurement And Construction (EPC) -Contracts Based on Renewable Energy," IJSART, vol. 10, no. 10, pp. 26–35, 2024.
35. D. N. V. K. Shivaji P. Mirashe, "Cloud computing can simplify HIT infrastructure management.," Healthc. Financ. Manage., vol. 65, no. 8, pp. 52–5, 2011.
36. A. P. A. S. and NeepakumariGameti, "Asset Master Data Management: Ensuring Accuracy and Consistency in Industrial Operations," Int. J. Nov. Res. Dev., vol. 9, no. 9, pp. a861-c868, 2024.
37. J. Thomas, "Optimizing Nurse Scheduling: A Supply Chain Approach for Healthcare Institutions," pp. 2251–2259, 2024.
38. M. Gopalsamy, "Advanced Cybersecurity in Cloud Via Employing AI Techniques for Effective Intrusion Detection," Int. J. Res. Anal. Rev., vol. 8, no. 01, pp. 187–193, 2021.
39. C. Y. Liu, C. M. Zou, and P. Wu, "A task scheduling algorithm based on genetic algorithm and ant colony optimization in cloud computing," Proc. - 13th Int. Symp. Distrib. Comput. Appl. to Business, Eng. Sci. DCABES 2014, pp. 68–72, 2014, doi: 10.1109/DCABES.2014.18.
40. K. Ullah et al., "Ancillary services from wind and solar energy in modern power grids: A comprehensive review and simulation study," J. Renew. Sustain. Energy, vol. 16, no. 3, 2024, doi: 10.1063/5.0206835.
41. P. K. V. and M. R. Suyambu, "A Study on Energy Management Systems (EMS) in Smart Grids Industry," Int. J. Res. Anal. Rev., vol. 10, no. 02, pp. 558–563, 2023.
42. S. Bauskar, "Advanced Encryption Techniques For Enhancing Data Security In Cloud Computing Environment," Int. Res. J. Mod. Eng. Technol. Sci., vol. 05, no. 10, pp. 3328–3339, 2023, doi: : https://www.doi.org/10.56726/IRJMETS45283.
43. S. A. and A. Tewari, "AI-Driven Resilience: Enhancing Critical Infrastructure with Edge Computing," Int. J. Curr. Eng. Technol., vol. 12, no. 02, pp. 151–157, 2022, doi: https://doi.org/10.14741/ijcet/v.12.2.9.

44. A. Soni, R. Arora, and A. Kumar, “Enhancing Security in Cloud Native Applications: A Blockchain-Based Approach for Container Image Integrity.” Sep. 2024. doi: 10.21203/rs.3.rs-5063708/v1.
45. R. Bishukarma, “Scalable Zero-Trust Architectures for Enhancing Security in Multi-Cloud SaaS Platforms,” Int. J. Adv. Res. Sci. Commun. Technol., vol. 3, no. 3, pp. 1308–1319, 2023, doi: 10.48175/IJARSCT-14000S.
46. S. Arora and P. Khare, “The Role of Machine Learning in Personalizing User Experiences in SaaS Products,” J. Emerg. Technol. Innov. Res., vol. 11, pp. c809–c821, 2024.
47. Ramesh Bishukarma, “Privacy-preserving based encryption techniques for securing data in cloud computing environments,” Int. J. Sci. Res. Arch., vol. 9, no. 2, pp. 1014–1025, Aug. 2023, doi: 10.30574/ijsra.2023.9.2.0441.
48. R. Yamini and M. G. Alex, “Comparison of Resource Optimization Algorithms in Cloud Computing,” vol. 116, no. 21, pp. 847–855, 2017.
49. V. Kumar, V. V. Kumar, N. Mishra, F. T. S. Chan, and B. Gnanasekar, “Warranty failure analysis in service supply Chain a multi-agent framework,” in SCMIS 2010 - Proceedings of 2010 8th International Conference on Supply Chain Management and Information Systems: Logistics Systems and Engineering, 2010.
50. V. V Kumar, M. Tripathi, S. K. Tyagi, S. K. Shukla, and M. K. Tiwari, “An integrated real time optimization approach (IRTO) for physical programming based redundancy allocation problem,” Proc. 3rd Int. Conf. Reliab. Saf. …, no. August, 2007.
51. R. Goyal, “EXPLORING THE PERFORMANCE OF MACHINE LEARNING MODELS FOR CLASSIFICATION AND IDENTIFICATION OF FRAUDULENT INSURANCE CLAIMS,” Int. J. Core Eng. Manag., vol. 7, no. 10, 2024.
52. H. Sinha, “Predicting Employee Performance in Business Environments Using Effective Machine Learning Models,” Int. J. Nov. Res. Dev., vol. 9, no. 9, pp. 875–881, 2024.
53. H. Sinha, “Benchmarking Predictive Performance Of Machine Learning Approaches For Accurate Prediction Of Boston House Prices : An In-Depth Analysis,” ternational J. Res. Anal. Rev., vol. 11, no. 3, 2024.
54. S. shrivastava Khare, Pranav, “Transforming KYC with AI: A Comprehensive Review of Artificial Intelligence-Based Identity Verification,” J. Emerg. Technol. Innov. Res., vol. 10, no. 12, pp. 525–531, 2023.
55. S. Arora and P. Khare, “THE IMPACT OF MACHINE LEARNING AND AI ON ENHANCING RISK-BASED IDENTITY VERIFICATION PROCESSES,” Int. Res. J. Mod. Eng. Technol. Sci., vol. 06, no. 05, pp. 8246–8255, 2024.
56. P. Khare and Sahil Arora, “Predicting Customer Churn in SaaS Products using Machine Learning,” Heal. Inf. Sci. Syst., 2020, doi: 10.1007/s13755-019-0095-z.
57. A. and P. Khare, “Cloud Security Challenges : Implementing Best Practices for Secure SaaS Application Development,” Int. J. Curr. Eng. Technol., vol. 11, no. 6, pp. 669–676, 2021, doi: https://doi.org/10.14741/ijcet/v.11.6.11.
58. Muthuvel Raj Suyambu and Pawan Kumar Vishwakarma, “Improving grid reliability with grid-scale Battery Energy Storage Systems (BESS),” Int. J. Sci. Res. Arch., vol. 13, no. 1, pp. 776–789, Sep. 2024, doi: 10.30574/ijsra.2024.13.1.1694.
59. P. Khare and S. Srivastava, “AI-Powered Fraud Prevention: A Comprehensive Analysis of Machine Learning Applications in Online Transactions,” J. Emerg. Technol. Innov. Res., vol. 10, pp. f518–f525, 2023.
60. S. Bauskar, “A Review on Database Security Challenges in Cloud Computing Environment,”

Int. J. Comput. Eng. Technol., vol. 15, pp. 842–852, 2024, doi: 10.5281/zenodo.13922361.
61. R. Bishukarma, “Optimising Cloud Security in Multi-Cloud Environments : A Study of Best Practices,” TIJER – Int. Res. J., vol. 11, no. 11, pp. 590–598, 2024.
62. Pranav Khare and Shristi Srivastava, “Data-driven product marketing strategies: An in-depth analysis of machine learning applications,” Int. J. Sci. Res. Arch., vol. 10, no. 2, pp. 1185–1197, Dec. 2023, doi: 10.30574/ijsra.2023.10.2.0933.
63. V. V. Kumar, A. Sahoo, and F. W. Liou, “Cyber-enabled product lifecycle management: A multi-agent framework,” in Procedia Manufacturing, 2019. doi: 10.1016/j.promfg.2020.01.247.
64. S. Bauskar, “Navigating Database Security in Cloud Computing: Challenges and Solutions,” Int. J. Comput. Appl., vol. 186, no. 51, pp. 26–31, Nov. 2024, doi: 10.5120/ijca2024924173.
65. A. Goyal, “Optimising Cloud-Based CI/CD Pipelines: Techniques for Rapid Software Deployment,” TIJER – Int. Res. J., vol. 11, no. 11, pp. a896–a904, 2024.
66. V. V. Kumar, A. Sahoo, S. K. Balasubramanian, and S. Gholston, “Mitigating healthcare supply chain challenges under disaster conditions: a holistic AI-based analysis of social media data,” Int. J. Prod. Res., 2024, doi: 10.1080/00207543.2024.2316884.
67. V. K. Yarlagadda, “Harnessing Biomedical Signals: A Modern Fusion of Hadoop Infrastructure, AI, and Fuzzy Logic in Healthcare,” Malaysian J. Med. Biol. Res., vol. 2, no. 2, pp. 85–92, 2021.
68. R. Arora, S. Kumar, N. Jain, and M. T. Nafis, “Revolutionizing Healthcare with Cloud Computing: Superior Patient Care and Enhanced Service Efficiency,” SSRN, 2022, doi: http://dx.doi.org/10.2139/ssrn.4957197.
69. R. Khullar and G. Hossain, “A New Algorithm for Energy Efficient Task Scheduling Towards Optimal Green Cloud Computing,” in 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA), 2022, pp. 1–6. doi: 10.1109/AICCSA56895.2022.10017609.
70. V. A. K. Gorantla, S. K. Sriramulugari, B. Gorantla, N. Yuvaraj, and K. Singh, “Optimizing Performance of Cloud Computing Management Algorithm for High-Traffic Networks,” in 2024 2nd International Conference on Disruptive Technologies (ICDT), 2024, pp. 482–487. doi: 10.1109/ICDT61202.2024.10489018.
71. T. Mengistu, D. Che, and S. Lu, “Multi-Objective Resource Mapping and Allocation for Volunteer Cloud Computing,” in 2019 IEEE 12th International Conference on Cloud Computing (CLOUD), 2019, pp. 344–348. doi: 10.1109/CLOUD.2019.00063.
72. J. Patel and T. Halabi, “Optimizing the Performance of Web Applications in Mobile Cloud Computing,” in 2021 IEEE 6th International Conference on Smart Cloud (SmartCloud), 2021, pp. 33–37. doi: 10.1109/SmartCloud52277.2021.00013.
73. R. Shiftehfar, K. Mechitov, and G. Agha, “Towards a Flexible Fine-Grained Access Control System for Modern Cloud Applications,” in 2014 IEEE 7th International Conference on Cloud Computing, 2014, pp. 966–967. doi: 10.1109/CLOUD.2014.144.
74. H. I. H. Tandri, H. H. Nuha, and R. G. Utomo, “Cloud Computing-Based API Design and Implementation for Hening Mobile Application,” in 2023 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), IEEE, Nov. 2023, pp. 341–346. doi: 10.1109/COMNETSAT59769.2023.10420654.