World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Review Article)

A review on AI techniques for cost optimization and forecasting in SAAS infrastructure

Pradeep Laxkar *

Department of Computer Science and Engineering, ITM(SLS) University, Vadodara, Gujarat.

World Journal of Advanced Engineering Technology and Sciences, 2025, 16(01), 375-384

Publication history: Received on 05 June 2025; revised on 15 July 2025; accepted on 17 July 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.16.1.1222

Abstract

The combination of Artificial Intelligence (AI) and Software as a Service (SaaS) platform has transformed the system of management of every available resource, allowing highly intelligent automation and increased scalability, as well as informed decision-making based on data. The current paper contains a detailed literature review of different machine learning (ML) techniques, including deep learning (DL), reinforcement learning (RL), etc., used to optimize resources deployed in SaaS environments. It discusses ways in which Artificial Intelligence (AI) can be used to increase cost-efficiency, boost service quality, enable predictive analytics and manage resources on the fly. What is also discussed in the study are the recent tendencies and strategies relying on AI to reinvent SaaS procedures and create user satisfaction. It also puts more focus on how AI can affect workload prediction, auto scaling of resources, and monitoring in real time, which are all part of operational excellence. It is with the help of this review that the paper seeks to give an insight into the changing importance of AI in SaaS applications and how its use can influence the future of cloud-based solutions. The results confirm the increasing significance of the AI-based solutions to the management of complicated SaaS infrastructures within an efficient environment. The next step in developing AI is further improvement of adaptability and responsiveness in the control of SaaS resources.

Keywords: Artificial Intelligence; SAAS; Resource Management; Machine Learning; Deep Learning; Reinforcement Learning; Predictive Analytics; Intelligent Automation; Cloud Computing

1. Introduction

In a modern business, it is not unusual to find software in every nook and corner of an enterprise, whether it is to keep its international shipments on track or to run a huge inventory or to train its employees or to maintain good relations with its customers. Conventionally, organizations have either had software housed on internal systems or on proprietary networks [1]. Nonetheless, within the past years, this model has changed in a big way. The Software as a Service (SaaS) paradigm is also becoming popular, where applications are provided over the Internet [2]. SaaS does not require local installations, it is easily maintained and supported and enables access to modular applications, managed services, shared resources and web-based capabilities.

As cloud infrastructure becomes the backbone of modern enterprises, managing its associated costs has emerged as a pressing challenge [3]. Organizations often face issues such as overprovisioning, inefficient resource utilization, and unexpected billing spikes. According to Flexera's 2023 State of the Cloud Report [4], based on insights from over 750 global IT decision-makers, enterprises estimate that approximately 32% of their cloud spend is wasted. The current state-of-the-art AI techniques are applied to cost optimization and forecasting within SaaS environments. It explores the taxonomy of methods, evaluates their effectiveness across various use cases, and highlights emerging trends and challenges in integrating AI for financial sustainability in cloud-native architectures.

* Corresponding author: Pradeep Laxkar

Copyright © 2025 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0.

Artificial Intelligence (AI) introduces a data-driven approach to address these inefficiencies and AI techniques have emerged as powerful tools for intelligent cost optimization and accurate forecasting in SaaS infrastructures, the study highlights that organizations adopting machine learning (ML) algorithms for cloud resource allocation achieved cost reductions of 23 - 47%, with development and testing environments where resource usage is highly variable, seeing average savings of 31.5%.The revolutionary effect of AI does not boil down to optimization per se. It is transforming business process in every industry such as healthcare, financial and marketing [5]. The emergence of SaaS-based AI platforms has made it easier for businesses of all technical skill levels to access sophisticated AI usage, and they can do so without having to build and maintain their own SaaS infrastructure [6]. These platforms offer on-demand scalable access to such tools as data processing, predictive analytics, automation.

The main AI methods like ML, Deep Learning (DL) and Reinforcement Learning (RL) play a critical role in cost reduction enhancement in cloud-based SaaS environments [7]. Such techniques are used to drive the following, When compute resources are needed, adapt dynamically to meet the demands, Real-time adaptations dynamically to ensure is efficient, Division of workloads so as to maximize performance, Minimizing the loss of power through analysis of history to make informed recommendations of pricing models according to the usage pattern [8]. Organizations may enhance resource efficiency, guarantee scalable, high-performing operations, and efficiently control cloud expenditures by incorporating AI into SaaS infrastructure.

1.1. Structure of the paper

This paper is structured as follows: Section II discusses resource management in SaaS, including AI integration and SaaS architecture. Section III focuses on cost control techniques using AI, such as auto-scaling and intelligent scheduling. Section IV highlights key challenges and emerging trends in AI-driven SaaS optimization. In Section V, AI Tools and Platforms for SaaS Cost Management. Section VI provides a review of relevant literature. Section VII summarizes the main conclusions and suggests areas for further investigation.

2. Fundamentals of SAAS infrastructure-based AI

The combination of Software as a Service (SaaS) with AI is transforming the development, implementation, and management of digital services. SaaS is a cloud-based software delivery method in which applications are centrally hosted and made available to users from any location that has internet, and is usually requested as a subscription [9]. The multi-tenant architecture of SaaS provides tremendous scalability, flexibility, and cost advantages, which work well for businesses large and small. But as the SaaS platform grows in complexity and users, there are considerable challenges to operational efficiency and cost management [10]. AI, or more closely ML and predictive analytics, has many powered tools to leverage to address these challenges. AI can sift through massive amounts of usage data to uncover hidden patterns and assist in making decisions in real time about resource allocation, usage behaviour forecasts, and existing cost control strategies. Implementing AI methods in SaaS infrastructure allows organizations to automate many standard functions like user management, scaling, and performance tuning, in turn providing more reliable services to their customers while also more efficiently incorporating cost considerations. Organizations must first thoroughly comprehend the fundamental ideas of both SaaS and AI in order to properly utilize their combined potential in the implementation of intelligent and data-driven cloud solutions.

2.1. Overview of SaaS Architecture

Service-oriented architecture and SaaS architecture are comparable. A distribution model known as Software as a Service (SaaS) enables software to be accessed over a network. said that virtualization provides consumers with an abstracted, simulated computing platform rather than the multi-tenancy physical features. The two main architectural components of SaaS software are a single instance of an application program servicing multiple clients and web service communication over the HTTP protocol [11]. A well-specified SaaS architecture needs to be adaptable. In addition to maximizing concurrency and optimizing application resources, it is scalable and multi-tenant. The program's appearance and feel may be altered by users. Multi-tenancy is a critical feature that is necessary for a SaaS business to succeed [12]. In a multitenant architecture, several tenants share a single instance of common code and data. In addition to shared hardware resources, applications, and database instances, multitenant software must have a high degree of look-and-feel and workflow configurability. Additionally, some academics view multi-instance as a type of multi-tenancy, in which providers use shared hardware to host distinct instances for every client.

2.2. Using AI in Predicting Analysis

Predictive analysis using AI entails using sophisticated algorithms and ML techniques to analyze historical data and predict future events or outcomes. AI may be used in predictive analysis in a number of ways, depending on the specific application or subject [13]. These are some applications of predictive analysis in AI:

- **Fraud Detection:** In addition to analyzing enormous volumes of data, including past trends, user behavior, and transaction histories, the AI can also instantly spot potential fraud or questionable conduct. Businesses or organizations can save money by using ML algorithms, which can learn from past fraud cases and spot trends or anomalies that may indicate fraudulent behavior.
- **Credit Risk Analysis:** In order to assess a loan application's credit risk and likelihood of default, AI may evaluate various data, such as job status, income, credit history, and other characteristics. To find patterns, evaluate credit risk, and help financial organizations make informed lending decisions, these machine learning models may be trained on historical data.
- **Marketing Campaign Optimization:** AI can evaluate vast volumes of client data, such as demographic information, past purchases, internet presence, and social media activity, to optimize marketing strategies. Because They may more effectively target their marketing efforts by using ML algorithms to categories their clientele, ascertain their preferences, and forecast their responses to various marketing campaigns.
- **Supply Chain Management:** AI may analyze a wide range of elements, including demand, production capacity, inventory, and transportation costs, to optimize supply chain operations. In order to assist organizations in simplifying their supply chain process and saving costs, Transportation route optimization, inventory level optimization, and demand pattern forecasting are all possible using machine learning systems.
- **Human Resources:** AI may be used to analyses employee data, including performance ratings, feedback, and engagement surveys, in order to forecast employee attrition, identify prospective talent, and improve workforce planning. A company may take proactive steps to retain the best employees by using machine learning algorithms to monitor employee behavioral trends and predict employee turnover.

2.3. Relevance of AI in SaaS Operations

The centrality of AI in changing SaaS business is that it brings about automation, intelligence, and flexibility into key operations of the business. AI methods help to analyze in real-time using the huge volume of data that SaaS products create through interaction with users, logs, etc.[14]. This enables SaaS providers to optimize their resource usage, predict demand and tailor user experiences with the least possible human interaction. Operational situations benefit from the help of AI since it provides intelligent load balancing and auto-scaling functions that guarantee stable performance at busy usage periods. It also helps find anomalies, security threats, or inefficiencies that may result in poor performance or raise the costs. Moreover, with the involvement of AI, chatbots, recommendation systems, and predictive support tools enhance customer engagement and customer satisfaction [15]. AI incorporated into the SaaS environment can help companies become more proactive and, instead of dealing with problems, can manage them with a proactive approach that would occur before the problem even occurs and continuously adjust to continuously increase or decrease the workloads or expectations of the user. This means that, in addition to improving the performance of SaaS platforms in terms of efficiency, AI can be used to achieve long-term viability, cost-efficiency, and competitive advantage.

3. AI Techniques in Cost Optimization

The process of cost optimization using AI is one of the most vital contributors to higher efficiency and profitability within SaaS when operated. Using ML, predictive and reinforcement learning, AI will be able to find out and address unnecessary expenditures, automate allocation according to the expected workload patterns and automate decision-making, replacing the manual approach with strategies that absorb fewer costs. To give an example, intelligent resource allocation models can assist in cloud infrastructure management by preventing over-provisioning [16], whereas AI-powered anomaly detection can invite swift identification of, and attention to, sudden jumps in usage and billing, which can mitigate against over-spending or the development of risks to security. Moreover, AI helps to schedule intelligently, thus examining the historical data on the use to execute non-critical activities at off-peak, low-cost hours [17]. AI applications enable real-time pricing, as well, so SaaS providers can optimize subscription prices or their services according to utilization patterns or economic factors. Reinforcement learning, which analyses operational outcomes and aims to maintain optimal cost controls, can be used to constantly improve these tactics. By automating these cost-controlling processes, SaaS providers will save operating expenses while simultaneously enhancing the calibre and overall performance of their services.

3.1. Resource Allocation and Auto-scaling with AI

The predictive auto-scaling mechanisms, which are based on the earlier ML models the forecasting model and the performance model make up the third part of the AI-driven resource allocation and auto-scaling architecture. The time series forecast estimates the expected load level for the VNF in the following period for each scheduling period. The time series prediction forecasts the expected future load level for the VNF during each scheduling session [18]. The performance model then uses this expected demand level to forecast the resources required in the next scheduling period in order to prevent VNF overload. Let's look at an example where The CPU utilization percentage and the packet rate, measured in kips, are connected by the performance model. In order to keep CPU utilization from going above a specific threshold, we utilize the performance model to determine the required CPU allocation for this VNF after determining the maximum packet rate it can manage in the next time period. With this method, auto-scaling may be done both vertically and horizontally.

3.2. Applications of AI in anomaly detection in networks

There are many different applications of AI in network anomaly detection, and organizing methodical research is challenging. [19]. The chosen publications have been categorized into four classes, as seen in Figure 1, since the authors have used four machine learning methods for anomaly detection: unsupervised learning, supervised learning, hybrid learning, and reinforcement learning. It's crucial to use supervised learning (like SVM, Random Forest, and neural networks) for classification and unsupervised learning (like K-means, Isolation Forest, GANs, and autoencoders) for identifying strange patterns. AI techniques for network security anomaly detection. Reinforcement learning supports real-time threat response, while hybrid approaches combine methods for greater accuracy.

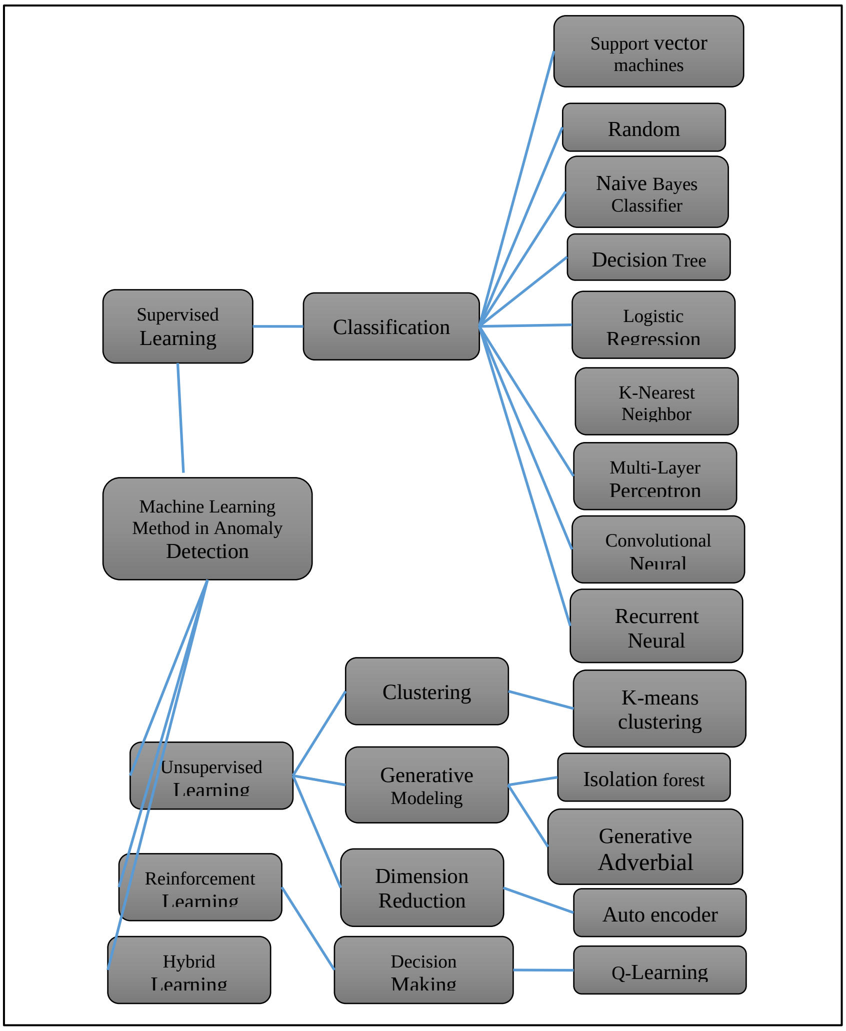

Figure 1 Classification of Approaches Related to the AI in Anomaly Detection in Networks

These techniques help detect cyber threats and intrusions, enhancing network reliability and security.

3.3. Load Prediction and Intelligent Scheduling

Load prediction and intelligent scheduling are critical for enhancing performance and minimizing operational costs in SaaS infrastructure. AI techniques, including time series analysis, regression models, and neural networks, may accurately forecast system load, including CPU utilization, memory demand, and traffic spikes, by utilizing historical use data, user behavior, and other contextual elements (such as the time of day or seasonal trends) [20]. Through such forecasts, smart scheduling algorithms can be used to optimize the distribution of resources, which allows them to automatically choose the most cost-effective and efficient time slots to use in ensuring that the non-urgent processes do not strain the resources during peak hours. This process can be even augmented by reinforcement learning and other optimization techniques, which keep adapting to new usage patterns and the performance of infrastructure. Collectively, these strategies can assist SaaS providers to eliminate the occurrence of resource bottlenecks, minimize response time, reduce power usage and total operational expenditures and also be able to provide a preference to the users during the service, especially in the multi-tenant case where resource interaction can decrease the quality of service.

4. Forecasting Techniques in Saas Using Artificial Intelligence (AI)

Forecasting is a very important role of SaaS operations, which allows the providers to be able to predict demand, use resources effectively, and make decisions in business. Methods of forecasting by means of AI have become central because they manage large amounts of information and identify complex patterns that statistical procedures could overlook. The methods employ past performance of the user in terms of user behavior, application usages, subscription patterns and system requirements to come up with precise and practical forecasts [21]. The most popular time series models for predicting future results of various key performance indicators (KPIs), including Monthly Recurring Revenue (MRR), user growth, and resource usage, are ARIMA, Facebook Prophet, and LSTM networks [22]. These models are effective in the dynamic SaaS and can be adjusted to seasonality, trends and sudden changes in user activity, which is why it is so effective in SaaS. Also, AI leads to scenario-driven forecasting in which various scenarios of the future are modeled depending on the observed changes in diverse inputs or market situations. This assists in capacity planning, budgeting and developing strategy. By leveraging AI forecasting techniques, SaaS providers gain deeper insights into their operations, improve customer satisfaction through better service delivery, and reduce unnecessary costs by avoiding over-provisioning or underutilization of resources.

4.1. Machine Learning Applications in Forecasting

One of the effective uses of ML-based models has been demand forecasting. Many scholarly research and consultancy reports have shown that ML models can perform better than conventional forecasting techniques. Additionally, big businesses like Walmart and Amazon have already developed ML algorithms for their demand projections [23]. Reducing obsolescence brought on by inaccurate projections and predicting the demand for new products are two examples of application cases where ML addresses urgent industrial issues [24]. Academic studies acknowledge the challenges of forecasting new items, yet only 58% of prediction accuracy has been seen in new products, for example. Because ML models are not constrained by the limitations of conventional forecasting techniques, ML applications typically produce superior results.

4.2. Revenue Prediction and Subscription Trends

Revenue estimation is a vital part of any SaaS business strategy that ultimately allows an organization to make coherent financial decisions, manage resources and plan for growth. AI techniques, especially ML and DL models, have shown good potential in accurate estimates of recurring revenue through analysis of historical subscription values, customer behavior, and usage patterns [25]. The most popular time series models for predicting future results of various key performance indicators (KPIs), including Monthly Recurring Revenue (MRR), user growth, and resource usage, are ARIMA, Facebook Prophet, and LSTM networks. These AI methods identify subscription trends (e.g., upgrades/downgrades, seasonality, the impact of promotions) that allow owners to better understand how to develop pricing strategies and tailor marketing campaigns. Ultimately, using predictive analytics, SaaS companies can decrease churn, increase customer lifetime value (CLV), and create financial predictability in a competitive market.

4.3. Capacity Planning and Usage Forecasting

Capacity planning and usage forecasting are critical to creating adequate infrastructure for your current and future workloads. AI-based models will allow you to provide accuracy in future compute, storage, and network demands by developing forecasts based on historical usage, user growth patterns, system workload variations, and other patterns detected through historical data [26]. Additionally, to forecast resource usage and predict peak periods, related ML algorithms could consist of regression models, recurrent neural networks (RNNs), and reinforcement learning algorithms (RLAs). Forecasting also helps SaaS providers to allocate adequate resources when scaling their service proactively, which minimizes service interruptions and optimizes operational capability. Also, AI is a great technology for scenario-based planning for any and all levels of SaaS plan and funding allocation, because it allows each organization to simulate additional growth paths and plan infrastructure investment accordingly.

5. AI Tools And Platforms for Saas Cost Management

AI tools and platforms for SaaS cost management include open-source and custom solutions aimed at enhancing visibility and optimizing cloud expenses. Open-source frameworks like Kub cost and Cloud Custodian offer vendor-neutral, customizable capabilities for real-time cost monitoring, policy enforcement, and budgeting across Kubernetes and multi-cloud environments [27]. Additionally, organizations often build custom AI pipelines tailored to their infrastructure, using machine learning for workload forecasting, anomaly detection, and dynamic resource optimization [28]. These approaches provide high scalability and flexibility, making them ideal for diverse and complex cloud architectures.

5.1. Commercial Tools

Commercial cloud service providers offer integrated AI-enhanced tools to assist organizations in managing and forecasting their cloud expenses [29]. These tools leverage machine learning models to detect cost anomalies, forecast spending trends, and provide recommendations.

- **AWS Cost Explorer with Machine Learning:** Offers visualizations and cost forecasting using ML-based algorithms. The tool allows users to filter usage data, explore cost trends, and predict monthly spend with models based on historical patterns.
- **Azure Advisor:** Integrates cost recommendations as part of its broader advisory service. It uses AI to provide actionable suggestions for cost savings, including underutilized resources, reserved instance purchasing, and scaling inefficiencies.
- **Google Cloud Recommender:** Utilizes ML to generate cost optimization insights, such as rightsizing virtual machines and identifying idle resources. Its explainable AI models help DevOps teams make informed financial decisions.

5.2. Challenges in Cost Management

5.2.1. There are several key challenges of cost management in SaaS based- AI as follows

- Lack of Real-Time Visibility Manual processes cannot continuously track dynamic usage patterns, service dependencies [30], and real-time pricing changes across cloud platforms.
- Human Error and Inefficiency Managing large volumes of billing and usage data manually is prone to errors, slow, and difficult to scale with growing infrastructure.
- Inaccurate Forecasting Traditional budgeting methods struggle to predict cost fluctuations caused by variable workloads, user demand, or traffic spikes.
- Delayed Decision-Making Slow data collection and analysis can result in delayed responses to cost overruns or inefficiencies, increasing overall cloud expenditure.
- Fragmented Cost Tracking Costs are often spread across different services, accounts, or regions, making manual consolidation and analysis complex and time-consuming.
- Poor Governance and Compliance Without automated policy enforcement, manual management makes it harder to detect unauthorized provisioning or cost anomalies, especially in multi-cloud environments.

6. Literature review

This literature review highlights recent advancements in integrating AI and ML into cloud systems, emphasizing improvements in cost-efficiency, security, workload forecasting, and business intelligence. It also explores strategic challenges, predictive analytics, and educational applications across diverse sectors.

Anand (2025) the potential of AI solutions for the advancement of reliability and cost-effectiveness of cloud systems. Self-controlled cooling systems enhance energy use, which is a huge saving for operations costs. Also, with the help of AI, one can predict possible security threats and prevent their occurrence at an early stage. There are benefits of implementing an intelligent cloud system in data centers that include cost-effectiveness, low emission of carbon emissions, and high service availability. Consequently, it is important to talk about the trend in the development of AI for automation in the near future and directions for further studies as well [31].

Katasani (2025) addresses critical challenges organizations face during cloud migrations, including resource allocation inefficiencies, data transfer costs, and operational complexities. Through an analysis of various predictive analytics approaches, machine learning techniques, and implementation strategies, the article demonstrates how organizations can achieve significant cost reductions and improved operational efficiency. The article encompasses privacy-preserving mechanisms, adaptive learning capabilities, and sophisticated monitoring systems, providing a holistic approach to cost management. This article also evaluates various implementation tools, including native cloud provider solutions, third-party platforms, and open-source frameworks, offering insights into their relative effectiveness [32].

Table 1 Summary of a study on AI Techniques for Cost Optimization and Forecasting in SaaS Infrastructure

Author	Study On	Approaches	Key Findings	Challenges	Future Directions
Anand et.al. (2025)	Role of AI in enhancing cloud system reliability and cost-effectiveness	Overview of AI frameworks for intelligent cloud systems	AI enables energy-efficient self-controlled cooling, early security threat detection, cost savings, and high service availability	Issues in practical implementation of intelligent frameworks	Trends in automation, need for more research into AI-driven cloud management
Katasani et.al. (2025)	Overcoming challenges in cloud migration using AI and analytics	Predictive analytics, ML techniques, and implementation strategies	Improved cost reduction and efficiency using adaptive learning and privacy-preserving models	Operational complexity, data transfer cost, and integration issues	Comparative evaluation of tools, enhanced adaptive and intelligent monitoring systems
Onabanjo A. et.al. (2024)	AI–cloud synergy to optimize infrastructure and innovation	Analytical study on AI's integration into cloud services	AI boosts scalability, security, and innovation in cloud transformation	Integration complexity, strategic alignment with enterprise goals	Strategic roadmaps for digital transformation with AI-cloud fusion
Mahbub et al. (2024)	DL-based cloud forecasting models' security flaws	Adversarial attacks using white-box techniques on DL models (RNN, LSTM, GRU, CNN, Attention)	DL models are highly susceptible to adversarial attacks, risking cloud workload prediction accuracy	Vulnerability of AI models to attacks in cloud environments	Advancing research in AI robustness, secure cloud forecasting models
Gupta et al. (2024)	Impact of AI and BDA on Business Intelligence in SaaS	Review of BI evolution via AI and Big Data Analytics (BDA)	AI enhances SaaS-based BI with better decision-making and sector-specific applications	Integration of BDA with legacy BI systems, scalability	BASOA model adoption, trend analysis of AI in enterprise BI
Naim et al. (2023)	ML and AI in E-Learning at King Khalid University	Qualitative study of LMS and BB platforms using ML/AI	Enhanced online learning experiences via AI-driven LMS	Lack of generalization to other institutions, limited evaluation metrics	Adaptable EL practices using AI across universities and future analytics-driven improvements

Onabanjo A. (2024) examines the mutually beneficial link between AI and cloud transformation, emphasizing how AI-powered solutions improve cloud infrastructure, streamline resource management, and spur industry innovation. Businesses are now able to be more agile and competitive due to the development of more secure, scalable, and efficient cloud solutions brought about by the integration of AI into cloud services. This paper explores the main advantages and difficulties of the AI-cloud synergy, including predictions for future developments and strategic ramifications for businesses starting their digital transformations [33].

Mahbub et al. (2024) four cutting-edge DL regression models: attention-based models, 1D Convolutional Neural Network (1D-CNN), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN). Reputable white-box adversarial attack generating methods from computer vision were used to build these hostile cloud workload samples. Our study is evaluated on three popular cloud benchmark datasets: Google trace, Bit brain trace, and Alibaba trace. The findings of our investigation unmistakably demonstrate how vulnerable DL-based cloud workload forecasting algorithms are to hostile assaults. The hazards to cloud data centres' cost-effectiveness and security are thus highlighted by the need for an in-depth investigation into the vulnerability of DL-based models for workload forecasting in cloud data centres [34].

Gupta et al. (2024) AI and data analytics are changing the face of BI. An aim of the study was to look at BI from every angle, particularly how it has changed with the addition of AI and sophisticated data analytics, and to see where these technologies are headed in the corporate world. The combined use of BI, AI, and BDA in SaaS products is the topic of this review. It explains big data analytics, enumerates its key components, and discusses its connection to business intelligence. Examined include successful sector-specific applications, AI advancements in BI, SaaS uptake in BI, and a proposal for a Big Data Analytics Service-Oriented Architecture (BASOA) [14].

Naim et al. (2023) the benefits of ML and AI in EL in general, and explain how King Khalid University's (KKU) EL Deanship is making the most of ML and AI to make EL effective and a good platform for L&T. The results of this descriptive research are derived on a qualitative analysis of the ML and AI applications at KKU's EL. Nonetheless, any institution may assess the effectiveness of their EL by using the same methods on its EL platform. KKU provides online learning methods through Learning Management Services (LMS) and online learning materials through Blackboard (BB) [35].

Table 1 presents a comparative analysis of recent studies highlighting the application of AI and ML techniques for cost optimization, forecasting, and system enhancement in SaaS infrastructure and cloud environments

7. Conclusion

This review highlights the transformative potential of integrating AI with SaaS platform, offering scalable solutions for resource optimization, cost efficiency, intelligent forecasting, and enhanced operational management. AI techniques like ML, DL, and reinforcement learning enable predictive analytics, automated scaling, anomaly detection, and intelligent scheduling—contributing significantly to the performance and profitability of SaaS infrastructures. AI plays an increasingly important role in tackling the problems of dynamic demand, consumer personalization, and real-time decision-making as SaaS use rises internationally. However, the quantity and quality of training data determine how successful AI models are, and real-time implementation can be resource-intensive in terms of system integration and processing cost.

Future research can explore the use of federated learning and edge AI to make SaaS platforms more efficient and privacy-preserving, especially in distributed environments. In complicated SaaS ecosystems, examining hybrid AI models that include machine learning and symbolic reasoning may help improve decision-making abilities. Additionally, more focus is needed on standardizing AI model governance within SaaS platforms to ensure ethical, transparent, and explainable AI-driven services. There is also potential in developing AI systems that can adapt autonomously to changes in user behavior and workload patterns without requiring human reconfiguration, making SaaS platforms even more intelligent and self-sustaining.

References

[1] R. Kumar, "Software as a Service (SAAS), Future of Information Technology & Business," Int. J. Adv. Res. Innov., vol. 5, no. 2, 2017.

[2] S. Garg, "AI/ML Driven Proactive Performance Monitoring, Resource Allocation and Effective Cost Management in SaaS Operations," Int. J. Core Eng. Manag., vol. 6, no. 6, pp. 263–273, 2019.

[3] A. R. Sarabu, "AI-Driven Strategies for Cloud Cost Optimization," Int. J. Sci. Technol., vol. 162, 2025, doi: 10.71097/IJSAT.v16. i2.4714.

[4] N. Patel, “Securing the Cloud: Strategies and Innovations in Network Security for Modern Computing Environments,” Int. Res. J. Eng. Technol., vol. 11, no. 04, 2024.

[5] P. Patel and S. Gajjar, “A Review: Comparative Analysis of Artificial Intelligence, Machine Learning (ML), and Data Science,” Int. J. Res. Publ. Rev., vol. 4, no. 10, 2023.

[6] V. Prajapati, “Cloud-Based Database Management: Architecture, Security, challenges and solutions,” J. Glob. Res. Electron. Commun., vol. 1, no. 1, 2025.

[7] M. Shahbaz and Deepshikha, “A Review of Cost-Effective Resource Management in Cloud Computing using AIBased Forecasting,” Int. Res. J. Eng. Technol., vol. 12, no. 4, 2025.

[8] S. Garg, “Predictive Analytics and Auto Remediation using Artificial Inteligence and Machine learning in Cloud Computing Operations,” Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci., vol. 7, no. 2, 2019.

[9] 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.

[10] F. Happy and J. Esite, “AI and SAAS Embedded System: Enhancing Content Creation Through Contextual Language AI and SAAS Embedded System: Enhancing Content Creation Through Contextual Language,” J. Sci. Technol., vol. Vol. 29, p. 2024, 2024, doi: 10.20428/jst. v29i2.2447.

[11] G. Laatikainen and A. Ojala, “SaaS Architecture and Pricing Models,” in Proceedings - 2014 IEEE International Conference on Services Computing, SCC 2014, 2015. doi: 10.1109/SCC.2014.84.

[12] N. Patel, “Secure Access Service Edge (Sase): Evaluating the Impact of Convereged Network Security Architectures in Cloud Computing,” J. Emerg. Technol. Innov. Res., vol. 11, no. 3, 2024.

[13] V. Božić, “AI and Predictive Analytics,” Gen. Hosp. Kopriv., no. May, 2023.

[14] S. Arora, S. R. Thota, and S. Gupta, “Artificial Intelligence-Driven Big Data Analytics for Business Intelligence in SaaS Products,” in 2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT), IEEE, Aug. 2024. doi: 10.1109/IC2SDT62152.2024.10696409.

[15] S. Bhekare, S. Bhoir, S. Deshmukh, O. Dhapte, and A. Kale, “Software as a Service AI Platform,” Int. J. Adv. Res. Sci. Commun. Technol., vol. 4, no. 4, 2024, doi: 10.48175/568.

[16] A. Goyal, “Optimising Cloud-Based CI / CD Pipelines: Techniques for Rapid Software Deployment,” vol. 11, no. 11, pp. 896–904, 2024.

[17] G. Olaoye, “The Impact of AI on Cloud Cost Optimization and Resource Management,” 2025.

[18] R. Moreno-Vozmediano, E. Huedo, R. S. Montero, and I. M. Llorente, “AI-Driven Resource Allocation and Auto-Scaling of VNFs in Edge-5G-IoT Ecosystems,” Electron., vol. 14, no. 9, pp. 1–32, 2025, doi: 10.3390/electronics14091808.

[19] M. Nikravan, M. Haghi Kashani, and S. Bazzaz Abkenar, “Anomaly Detection with Artificial Intelligence,” 2025, pp. 59–78. doi: 10.1201/9781003521020-4.

[20] I. S. Jahan, V. Snasel, and S. Misak, “Intelligent Systems for Power Load Forecasting: A Study Review,” Energies, vol. 13, no. 22, Nov. 2020, doi: 10.3390/en13226105.

[21] J. Fernando and S. Coelho, “Sales Forecasting for SaaS,” 2022.

[22] S. Singamsetty, “Dynamic Stock Price Prediction Leveraging LSTM, ARIMA, and Sparrow Search Algorithm,” Int. J. Comput. Math. IDEAS, vol. 16, no. 3, 2024.

[23] V. Prajapati, “Enhancing Supply Chain Resilience through Machine Learning- Based Predictive Analytics for Demand Forecasting,” 2025.

[24] C. Singh, “Demand forecasting with Machine Learning,” Thinkdeeply, 2021.

[25] R. M. Yewande, “International Journal of Research Publication and Reviews AI Enhanced Revenue Modeling and Financial Foresight for Risk- Responsive Growth in Minority-Led Enterprises,” Int. J. Res. Publ. Rev., vol. 6, no. 3, 2025.

[26] M. Palaniandi, N. Jayaprakash, and K. Subramanian, “Capacity Planning Analysis Model and Forecasting of Storage Usage for Storage Area Networks,” 2010.

[27] A. and P. Khare, “Cloud Security Challenges: Implementing Best Practices for Secure SaaS Application Development,” Int. J. Curr. Eng. Technol., vol. 11, no. 06, 2021, doi: https://doi.org/10.14741/ijcet/v.11.6.11.

[28] V. Thangaraju, “Enhancing Web Application Performance and Security Using AI-Driven Anomaly Detection and Optimization Techniques,” Int. Res. J. Innov. Eng. Technol., vol. 9, no. 3, p. 8, 2025.

[29] V. Thangaraju, “Security Considerations in Multi-Cloud Environments with Seamless Integration: A Review of Best Practices and Emerging Threats,” Trans. Eng. Comput. Sci., vol. 12, no. 2, p. 6, 2024.

[30] V. S. Thokala, “Scalable Cloud Deployment and Automation for E- Commerce Platforms Using AWS, Heroku, and Ruby on Rails,” pp. 349–362, 2023, doi: 10.48175/IJARSCT-13555A.

[31] K. Anand, “AI-Driven Optimization of Cloud Data Centers: Enhancing Reliability and Cost Efficiency,” in 2025 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), 2025, pp. 1–6. doi: 10.1109/ICDSAAI65575.2025.11011873.

[32] D. P. Katasani, “Optimizing Cloud Data Migration Costs Through Predictive Analytics: A Framework and Implementation Guide,” Int. J. Comput. Eng. Technol., vol. 16, no. 1, Jan. 2025, doi: 10.34218/IJCET_16_01_028.

[33] E. Onabanjo A., “Digital Transformation: The impact of AI on Cloud Transformation,” J. Artif. Intell. Gen. Sci. ISSN3006-4023, vol. 5, no. 1, pp. 174–183, Jun. 2024, doi: 10.60087/jaigs. v5i1.188.

[34] N. I. Mahbub, M. D. Hossain, S. Akhter, M. I. Hossain, K. Jeong, and E.-N. Huh, “Robustness of Workload Forecasting Models in Cloud Data Centers: A White-Box Adversarial Attack Perspective,” in IEEE Access, 2024. doi: 10.1109/ACCESS.2024.3385863.

[35] A. Naim, S. M. Alshawaf, P. Kumar Malik, and R. Singh, “Effective E-Learning Practices by Machine Learning and Artificial Intelligence,” in 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), 2023. doi: 10.1109/AISC56616.2023.10085391.