foi

UNIVERSITY OF ZAGREB
FACULTY OF
ORGANIZATION
AND INFORMATICS
VARAŽDIN

e-ISSN 1846-9418

JIOS

Journal of Information and Organizational Sciences

SVEUČILIŠTE U ZAGREBU • FAKULTET ORGANIZACIJE I INFORMATIKE •

UNIVERSITAS STUDIORUM ZAGRABIENSIS MDCLXIX

VOL. 49, NO. 2, PP. 161-359, DECEMBER 2025

UDC 004:005:3
DOI: https://doi.org/10.31341/jios

JIOS

UDC 004:005:3
e-ISSN 1846-9418

DOI: https://doi.org/10.31341/jios

Publisher
University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia

For Publisher
Marina Klačmer Čalopa, University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia

CONTENTS

Journal of Information and Organizational Sciences
Volume 49, Number 2 (2025)
Journal homepage: jios.foi.hr

JIOS

DOI: 10.31341/jios.49.2.1
Open Access

UDC 004.85:004.896
Original Scientific Paper

Developing a Shared Knowledge Area Mechanism for Multi-Mobile Agents to Improve Performance Using Machine Learning: Classification-Rule

Tarig Mohamed Ahmed[1*]

[1]Department of Information Technology, FCIT, King Abdulaziz University, Jeddah, Saudi Arabia
[*]Correspondence: Tmahmad@Kau.edu.sa

PAPER INFO

Paper history:
Received 14 April 2025
Accepted 14 July 2025

Citation:
Ahmed, T. M. (2025). Developing a Shared Knowledge Area Mechanism for Multi-Mobile Agents to Improve Performance Using Machine Learning: Classification-Rule. In Journal of Information and Organizational Sciences, vol. 49, no. 2, pp. 160-177

ABSTRACT

A mobile agent system is a mobile computing approach where agents move autonomously among hosts to perform tasks. It offers advantages such as low latency, reduced bandwidth use, and cost efficiency. This paper proposes the Shared Knowledge Area Mechanism (SKAM) to improve mobile agent performance. SKAM uses a shared knowledge database that stores classification rules based on agents' travel experiences. Each rule is an IF–THEN statement linking service combinations to host locations. We extract these rules using support, confidence, and lift to ensure reliability. Before starting a task, an agent queries the database to select hosts based on the most relevant rules. This reduces unnecessary host visits and shortens travel time. SKAM is implemented within the Secure Mobile Agent Generator (SMAG), a platform used to simulate mobile agent behavior. SKAM also applies rule prioritization to support accurate itinerary planning. Experimental results show that SKAM reduces average task completion time from 41,146.5 ms to 23,445.5 ms—a 43% improvement. This gain is statistically significant ($p < 0.05$) and consistent across all agents. It confirms that SKAM lowers both search overhead and travel time. These results highlight SKAM's effectiveness and practical value for real-time, large-scale mobile agent systems.

Keywords: mobile agent, performance, machine learning, classification rules

1. Introduction

Boosting performance is one of the crucial factors that could be used to measure software quality (Tronge et al., 2021). It helps evaluate software throughput, which should be planned carefully during design. Software performance is affected by different factors related to the hardware or software itself (Papadopoulos et al., 2019). High-performance hardware depends on a high-speed processor, memory size, and I/O cost (Trifonov and Heffernan, 2022). At the software level, performance depends on the architecture of a program and code optimization.

A mobile Agent allows computers to communicate using an asynchronous mode. It is based on remote programming (Kusek et al., 2003). A mobile agent autonomously moves among computer nodes (hosts) to complete tasks for users. Mobility is a critical feature that allows mobile agents to travel among hosts (Gavalas et al., 2009). There are two types of mobility: static and dynamic. When the hosts in an itinerary table are known to a mobile agent, static mobility is used. Dynamic mobility is also called free-roaming mobility. It is used when hosts are unknown, where a mobile agent starts its journey (Prem and Swamynathan, 2012).

A mobile agent system (MAS) consists of several integrated components: mobile agents, mobile agent homes, and hosts (Wu et al., 2007). A mobile agent's home is where mobile agents start their journeys based on users' requirements. After finishing their travels, they return to the mobile agent home with results. Hosts represent service providers. Mobile agents can visit multiple hosts during a journey. MAS has the ability to dispatch multiple mobile agents simultaneously with different tasks. Hosts can also receive multiple mobile agents and serve them concurrently. Security is one of the main challenges facing MAS. MAS should protect a mobile agent against malicious hosts. In addition, hosts should be protected against malicious mobile agents. Mobile Agent

Performance is an important factor for MAS to be fully utilized. It means that a mobile agent performs its tasks in a short time (Dorri et al., 2007). In order to improve the performance, three items should be optimized: First, mobile agent coding must be written in a good way with high throughput. Second, reducing the number of visited hosts Third, reducing the mobile agent size to minimize the network bandwidth, which enhances performance. This paper deals with the second item by proposing a new mechanism called SKAM for Multi Mobile Agents using machine learning. The main idea behind SKAM is to allow mobile agents to share their experiences with each other in a shared knowledge database. The database stores classification rules mined from agents' travel data. Each rule is an IF–THEN statement that links a combination of service features to a host location. It derives these rules by collecting each agent's experience tuple (service types visited and the host served). It uses support, confidence, and lift to select only the most reliable rules. All rules are structured as conjunctions of service indicators leading to a predicted host. In data mining, rule-based categorization is a method that divides data instances into distinct groups or classes based on a predetermined set of criteria. It's a well-liked method in data analysis and machine learning (Allawi et al., 2019; Lee et al., 2023). A mobile agent can predict its itinerary table based on rules available in the knowledge database. Rule-based classification is a data mining technique where a set of if-then rules is used to classify data. The rules are generated from training data and can be easily interpreted. Rule-based classifiers often perform well on complex, high-dimensional shared information s. They can handle both numerical and categorical features (Smith et al., 2022; Kalkha et al., 2023).

Let's define the following:

X = {x1, x2,..., xm} be the set of input instances.

Y = {y1, y2,..., yc} be the set of class labels.

R = {r1, r2,..., rn} be the set of rules, where each rule ri is of the form:

ri: if (condition 1 ∧ condition 2 ∧... ∧ condition k) then class = y, where y ∈ Y

The rule-based classification algorithm can be mathematically expressed as follows:

for each instance x ∈ X:

vote_count = [0, 0,..., 0] # initialize the vote count for each class to 0.

for each rule ri ∈ R:

if (condition1(x) ∧ condition2(x) ∧... ∧ conditionk(x)) is true:

vote_count[y] += 1 # increment vote count for the corresponding class label y

predicted_class = arg max(vote_count) # predict the class with the highest vote count

The key steps are:

Initialize the vote count for each class label (host) to 0.

Iterate through each rule ri in the rule set R.

Evaluate the conditions of the rule for the given instance (service).

If all conditions are true, increment the vote count for the corresponding class label (host) h.

After evaluating all rules, predict the class label (host) with the highest vote count as the final prediction for the instance s. This algorithm is commonly known as the "covering algorithm" in rule-based classification (Fürnkranz, 1999).

Recent studies have tackled related challenges in data preparation, feature selection, and classifier design. "Building online social network dataset for Arabic text classification" creates a specialized dataset for Arabic text tasks (Omar et al. 2018). "Improving ZOH Image Steganography Method by using Braille Method" fuses Braille with image steganography to enhance data hiding (Abdelmged et al., 2016). "Privacy issues of public Wi-Fi networks" analyzes security gaps in open networks (Lotfy et al., 2021). Other papers propose new classifiers or hybrid feature techniques that improve performance (Farghaly et al., 2020; Mostafa and ElAraby, 2024). These efforts highlight the need for robust data handling and efficient decision rules. Our SKAM mechanism extends this work by focusing on mobile agent routing and shared knowledge.

In this paper, Shared Knowledge Area Mechanism (SKAM) has been Introduced for efficient mobile agent routing. We integrate SKAM with the Secure Mobile Agent Generator (SMAG) (Ahmed, 2005) to simulate realistic agent travels. From each agent's experience, IF–THEN rules to guide host selection has been derived. Our experiments show a 43 % reduction in average completion time, confirmed by statistical analysis. An ablation study to measure each component's impact is performed. Finally, the paper presents a deployment cost analysis to demonstrate SKAM's practical feasibility.

2. Related work

Many approaches have been suggested in this area by researchers. This section presents some of them as follows:

Wu et al. proposed a solution for Multiple Agent Itinerary Planning (MIP). The solution was based on the agent’s location and size to achieve a balance in consuming network energy. After conducting many experiments, the results mentioned that the performance had increased (Wu et al., 2007). Aloui et al. introduced a solution to enhance mobile agents’ performance based on location and size to maintain a balance in consuming network energy. This mechanism used Multi-Agent Itinerary Planning (MIP) (Aloui et al., 2015). To reduce energy consumption and latency, Prapulla et al. presented a model for multi mobile agents. There are two sorts of mobile agents in the model: link agents and data agents. The purpose of the link agent was to keep track of network resources and conditions. The goal of the data agent was to transport data between nodes. This model's concept aids mobile agents in preparing their itinerary tables. The model was developed, and the efficiency was discussed by clustering the network nodes (Prapulla et al., 2016).

To increase mobile agents’ performance, it is critical to make accurate resource predictions. Chaudhar et al. proposed that cognitive agents be used in mobile ad hoc networks. The cognitive agent trains the mobile agent to think like a person in order to make the best resource decisions. The mobile agent can then select the ideal traffic route for completing its responsibilities (Chaudhari and Biradar, 2016). Tarig proposed a new mechanism for improving the performance of mobile agents by lowering their size during agent travel. Free Area Mechanism (FAM) is the name of the mechanism. The method was created with the.NET framework, and numerous tests were undertaken to evaluate its performance (Ahmed, 2007).

Zuo et al. suggested a paradigm for improving the performance of mobile agent systems based on their opinions (Yanjun and Liu, 2017). By aggregating data, the model ranks the reputation of network nodes. The node reputation ranking was determined by a number of factors, including service quality. The mobile agents would gain crucial information before beginning their excursions, and overall performance would improve. ALGETS technology was used to implement and assess the model. Baek et al. proposed that planning algorithms try to find a minimum number of agents and the overall resource consumption time by imposing a time limit (Baek et al. 2001). The route of the mobile agent and the number of mobile agents are two major planning parameters that affect the performance of the agent system in the network environment. The bandwidth fluctuates from link to link when the size of a mobile agent is grown while retrieval activities are done, according to the results of this study's experiment. The agent will take longer in this situation.

To improve the performance of mobile agents, Selamat et al. suggested an extended hierarchical query retrieval (EHQR) technique. The fundamental concept behind this strategy was to dispatch a large number of agents at once in order to shorten job completion time. Two tests were done employing queries online and offline to assess EHQR (Selamat and Selamat, 2005). Rantes et al. created a model for analyzing mobile agent performance characteristics using SNMP (simple network management protocol). The results of several tests revealed that mobile agent performance is influenced by network management and various network topology characteristics, such as network latency (Rantes et al., 2010).

In the case of delayed updating of agent locations, Gu et al. suggested a detection performance assessment technique for distributed multi-agent detection. They discovered the influence mechanisms of several non-ideal elements by calculating the spatial-temporal detection utility function across the network. Furthermore, an asymptotic analysis on an infinite time horizon is used to give the universal lower bound of detection performance as well as the upper bounds for two scheduling approaches. The superiority of delay-aware scheduling in mobile detection networks is further supported by numerical findings (Gu et al., 2021). Guo et al. presented a service migration methodology and performance assessment in the MEC environment utilizing a mobile agent. Experiments with jobs of varying complexity reveal that mobile agent technology clearly surpasses container technology in terms of service transfer efficiency. The following are the primary contributions to this paper: (1) We solve the problem of too many auxiliary modules in the container architecture; (2) we investigate the differences between a mobile agent and container technology by using the agent container and the resource manager (RM); and (3) we use the decision tree to confirm the execution cost of each node in order to understand why the mobile agent responds to migration commands slowly by using the decision tree (Guo et al., 2021).

Okonor's suggested solution is very intelligent and can readily discover underutilized and overloaded data center components. The agent approach has effectively demonstrated its ability to avoid and control overloading difficulties caused by changes in workloads, as well as accomplish more efficient load balancing while consuming less power. The mobile agent was installed in servers and switches to control their activity and subsequently turn off underutilized components. The first of its sort in a cloud setting is the mobile agent (Java agent). This study idea saves a large amount of energy while also improving the overall performance of the mobile agent system (Okonor, 2021). Tarig proposed using knowledge-based content to increase the performance of mobile agent systems. To begin, this project began with a comprehensive examination of similar models and procedures in order to identify performance gaps. A comparison of certain published studies with the suggested model has been carried out. The components have been used to explain the proposed model in great depth. The suggested model was implemented utilizing a scenario-based approach with the.NET framework and C# language. Various situations were used to test and assess the model. When knowledge-based material is employed, overall performance improves by 83 percent, according to research (Ahmed, 2018).

Allawi et al. propose a novel approach to enhance the performance of mobile agents in data transfer. By utilizing a hybrid combination of genetic algorithms (GA) and node compression algorithms (NCA), they minimize the time required to select the best path for data transfer. Their experimental results show a significant reduction in selection time, from 336.448 ms to 286.29 ms, highlighting the effectiveness of optimization techniques in cloud computing. This research contributes to improving mobile agent performance and optimizing data transfer in a distributed environment (Allawi, 2019).

Althamary et al. provide a comprehensive survey on multi-agent reinforcement learning (MARL) in vehicular networks, highlighting its potential for optimizing communication and resource management through agents' collaborative strategies. The study emphasizes MARL's potential but does not address limitations in real-time data sharing among mobile agents or how MARL performs under high network load (Althamary et al., 2019).

Ning and Xie offer an in-depth review of MARL applications in various fields, focusing on the adaptability of mobile agents in dynamic networks and the development of scalable, decentralized control systems. Although comprehensive, the study primarily reviews theoretical aspects and lacks detailed experimental validations in real-world MAS environments (Ning and Xie, 2024).

Cui et al. explore multi-agent reinforcement learning (MARL) for resource allocation in Unmanned Aerial Vehicle (UAV) networks, demonstrating how MARL can optimize resource distribution in environments with high mobility and limited bandwidth. While effective for UAV networks, the approach relies on predefined models that may not generalize well to other MAS environments with different requirements (Cui et al., 2020).

Han et al. present a comprehensive survey of cooperative and competitive behaviors in MAS, with a focus on distributed optimization and federated optimization for improving networked agent performance. They emphasize privacy-preserving optimization in cooperative tasks and the use of game-theoretic approaches for balancing local and global costs in MAS. The work primarily reviews optimization and privacy aspects without delving into real-time adaptability or the impact of shared knowledge in dynamic environments (Han et al., 2022).

Herrera et al. explore control and optimization of MAS and complex networks, applying biological and nature-inspired models to industrial engineering contexts. They introduce multi-resolution MAS modeling and optimization techniques that enhance scalability in engineering systems through adaptive control mechanisms. While effective in industrial applications, this work focuses on static optimization models and does not fully address the challenges of dynamic knowledge updating or real-time learning in MAS (Herrera et al., 2021).

Ding et al. review advances in event-triggered consensus algorithms within MAS, focusing on performance improvements through adaptive and event-driven control mechanisms. Their work highlights efficient communication protocols that reduce bandwidth requirements while maintaining synchronization in MAS. The study is primarily focused on synchronization, lacking detailed exploration of machine learning integration or shared knowledge mechanisms to enhance adaptability in MAS performance (Ding et al., 2024).

Existing approaches to mobile agent performance enhancement, such as Multiple Agent Itinerary Planning (MIP) and various detection strategies in delay-sensitive networks, have primarily focused on optimizing itinerary planning, energy consumption, or execution environment factors. However, these methods often require each agent to independently identify service locations, leading to increased search times and network overhead in dynamic environments. SKAM addresses this gap by providing mobile agents with a shared knowledge database containing service locations and route histories, generated from previous agents' travel data. This database enables agents to access pre-processed knowledge, allowing them to directly identify relevant hosts, thereby minimizing redundant host visits and improving search efficiency.

SKAM is particularly effective in dynamic host environments and multi-agent systems, where rapid host reconfiguration and network load are common. In such settings, SKAM's real-time knowledge sharing enables agents to adapt their search strategies based on recent, collective experiences, which prior models cannot dynamically support. This approach reduces network bandwidth consumption and optimizes performance by guiding agents to suitable hosts without requiring extensive individual search processes.

Unlike previous models, SKAM's knowledge database is continuously updated with each mobile agent's journey data, enhancing accuracy and performance over time. This cumulative learning capability is unique to SKAM, positioning it as a valuable mechanism in scenarios where mobile agents frequently revisit similar service locations.

Building an online social network dataset for Arabic text classification (Omar et al., 2018) shows how careful data collection and labeling can improve model performance. Mostafa et al. introduce a feature selection method based on frequent and associated item sets to pick discriminative features for text classification (Ghaleb et al., 2006). Both studies focus on text features, while our work uses rule mining on mobile agent service records instead.

Farghaly et al. present a hybrid feature selection approach that combines filter and wrapper methods to choose optimal attributes from high-dimensional data (Mostafa and ElAraby, 2024). Mostafa and ElAraby apply PCA and recursive feature elimination to reduce features for hepatocellular carcinoma prediction (Mostafa et al., 2024). Mohamed and El-Hafeez review how deep learning simplifies feature selection in medical datasets (Mohamed and El-Hafeez, 2024). These works confirm that hybrid and deep methods can boost accuracy. In contrast, SKAM ranks hosts using simple rule prioritization and weighted metrics without heavy computation.

Ghaleb et al. merge association rules with a support vector machine to build an effective and accurate associative classifier (Farghaly et al., 2020). Zhang et al. apply deep regression analysis to optimize thermal control in photovoltaic systems (Zhang et al., 2023). While these studies emphasize model design, they highlight the value of combining rule-based and statistical approaches. SKAM similarly uses classification-rule mining, but for mobile agent itinerary planning rather than domain-specific prediction.

Abdelmged et al. improve ZOH image steganography by embedding Braille patterns into images (Abdelmged et al., 2016), demonstrating how hybrid techniques can enhance performance. Lotfy et al. analyze privacy issues in public Wi-Fi networks, emphasizing the need for secure mobile agent communication (Lotfy et al., 2021). Sowunmi et al. propose a semantic-web framework for e-learning systems, focusing on structured knowledge representation (Sowunmi et al., 2017). Yehia et al. extract topics and build interactive knowledge graphs for learning resources (Yehia et al., 2022). Their insights on knowledge representation influenced SKAM's shared database design and rule storage.

These prior works address dataset creation, feature selection, classifier design, and domain-specific knowledge representation. However, none tackle dynamic rule sharing among mobile agents. SKAM fills this gap by providing a shared knowledge area and rule prioritization to guide agent travel and improve coordination.

3. Materials and Methods

SKAM is a new mechanism that aims to improve mobile agent performance (MAP). It is based on the goal of reducing mobile agent travel time. Mobile agents can share their activities and experiences in a shared knowledge database. The mobile agent travel experiences are stored in the knowledge database as a classification rules-based algorithm. Before a new mobile agent starts its task, it will consult the knowledge database to reduce searching time. The following sections explain SKAM and how it works to achieve the performance goal.

3.1. SKAM Knowledge database

SKAM uses a classification rule-based algorithm to develop the database knowledge by using mobile agents travel information and from where they performed services. This experience database is stored as a tuple, which is composed of services and a host. The services represent features, and host represents a class label. From the database, the classification rule-based algorithm generates knowledge rules that could be used by mobile agents to enhance their journey performance. SKAM uses the following steps to generate the rules.

Let's define the following:

SKAM enhances rule-based classification through shared knowledge and adaptive learning. This enhancement is particularly valuable when compared to existing approaches to mobile agent performance enhancement, such as Multiple Agent Itinerary Planning (MIP) and various detection strategies in delay-sensitive networks. The mechanism's core components can be rigorously defined as follows:

- Rule Base (R): Given the set of classification rules,

R = {r₁, r₂, ..., r_n}, each rule r_i ∈ R takes the form:

r_i: IF (C_i) THEN y_i = f(h, t, v), where y_i ∈ Y.

Here:

C_i represents the conjunction of conditions for rule r_i.

Y is the set of possible classes.

f(h, t, v) is a function that combines h (history of travels), t (search time), and v (number of visited hosts) to determine the class y_i.

- Let's define h as the sequence of the n previously visited hosts h = (host_1, host_2, ... host_n)

Knowledge Sharing Function (KS): Models agent contribution and learning from the shared knowledge area, updating the rule base, similar to cooperative behaviors studied in multi-agent systems (Han et al., 2024):

R'_{t+1} = KS(R_t, E_t)

Where:

R_t is the rule base at time t.

R'_{t+1} is the updated rule base at time t+1.

E_t is the agent's experience tuple at time t: E_t = (s_t, a_t, r_t).

s_t is the state (characterized by feature vector x_t).

a_t is the action taken (host selected).

r_t is the immediate reward obtained.

KS can implement various knowledge integration strategies, such as Bayesian updating or evidence accumulation using Dempster-Shafer theory. The design of KS must also consider the event-triggered consensus algorithms to balance performance with adaptive control mechanisms (Ding et al., 2023).

- Rule Prioritization Function (P): Assigns a priority score π_i to each rule r_i ∈ R, reflecting relevance and reliability:

π_{i,t} = P(r_i, R_t, E_t)

P is a function of rule-specific metrics and potentially global performance measures. Common choices include:

Confidence (conf(r_i)): Estimated probability of y_i being correct given C_i.

Support (supp(r_i)): Fraction of instances in the training data satisfying C_i ∧ y_i.

Lift (lift(r_i)): Ratio of observed confidence to expected confidence.

Recency (rec(r_i)): A time-decayed measure of how recently the rule was successfully applied.

A weighted average is a possible implementation:

P(r_i) = w_{conf} conf(r_i) + w_{supp} supp(r_i) + w_{lift} lift(r_i) + w_{rec} rec(r_i),

subject to Σ w = 1 and w ≥ 0. (Herrera et al., 2023) notes that this aligns with the optimization strategies used in multi-resolution MAS.

- Itinerary Selection Function (IS): Selects the next host h* to visit based on the prioritized rule base and the agent's current state s_t:

h* = argmax_{h∈H} {Σ_{ri∈R(st,h)} π_{i,t} },

Where:

H is the set of available hosts.

R(s_t, h) is the subset of rules in R_t that match state s_t and suggest host h.

This function selects the host that maximizes the sum of the priority scores of the matching rules. Cui et al. (2024) suggest that alternative strategies could be used such as a soft-max selection if exploration is needed.

Agent's Utility Function (U): Maximizes cumulative discounted rewards over a time horizon T:

U = Σ_{t=0}^T γ^t R(s_t, a_t)

Where:

γ ∈ is the discount factor.

R(s_t, a_t) is the immediate reward function, typically defined as a function of service time, travel cost, and task completion success.

- Queueing Model Integration:

Let λ_h be the arrival rate of agents to host h, and μ_h be the service rate of host h. Assuming an M/M/1 queueing model:

Expected waiting time at host h: W_h = λ_h / (μ_h(μ_h - λ_h))

The reward function R(s_t, a_t) can be modified to incorporate the expected waiting time:

R(s_t, a_t) = R'(s_t, a_t) - αW_h

Where:

R'(s_t, a_t) is the original reward (without queueing).

α is a weighting factor that balances the importance of the queueing delay with the original reward.

In the SKAM framework, rule conflicts are addressed through a conflict resolution strategy that prioritizes rules based on a voting mechanism. When multiple rules apply to a service-host pairing, each rule contributes a vote, and the host with the highest vote count is selected. This approach ensures that commonly used or accurate rules have a greater influence on the outcome. Additionally, SKAM's knowledge database is continuously updated, allowing frequently successful rules to naturally accumulate higher priority. This conflict resolution method enhances the reliability and consistency of SKAM's decision-making, minimizing the risk of incorrect host prioritization and optimizing performance.

3.2. SKAM Framework

SKAM uses the shared information in the knowledge database to generate rules about service location based on mobile agent travel experiences. With passing time, the mobile agents update the knowledge database, enhancing the rules.

Table 1 presents SKAM workflow, and Figure 1 presents SKAM system components.

Step	Description	Result
1	Create Mobile Agent based on a user request	Mobile Agent is created
2	Dispatch Mobile Agent from Home to SKH (Shared Area Host)	Mobile Agent is dispatched to SKH
3	Search for interesting places for a mobile agent's task using rules with highest vote.	Places of interest identified (to enhance performance)
4	Prepare the itinerary table, T = {Host1, Host2, ..., Host n}	Itinerary table is prepared based on highest voted rules.
5	Dispatch Mobile Agent from SKH to Host1	Mobile Agent is dispatched to Host1
6	Host1 serves Mobile Agent	Mobile Agent task is served at Host1
7	Dispatch Mobile Agent from Host1 to Host2	Mobile Agent is dispatched from Host1 to Host2
8	Host2 serves the MA	Mobile Agent task is served at Host2
9	Dispatch Mobile Agent from Host2 to Host3	Mobile Agent is dispatched from Host2 to Host3
...	Visit all hosts in the itinerary table (Host1, Host2, ..., Host n)	Mobile Agent visited all hosts in the itinerary table
N	Dispatch Mobile Agent from Host n to SKH	Mobile Agent is dispatched from Host(n) to SKH
n+1	Update the Knowledge database using the Mobile Agent's results	Knowledge database updated with MA's results
n+2	Dispatch Mobile Agent from SKH to MA Home	Mobile Agent is dispatched from SKH to Home
n+3	Extract the results from Mobile Agent and submit them to the user	Results extracted and submitted to the user

Table 1. SKAM Workflow

Figure. 1 presents the SKAM architecture that Includes all the components. A mobile agent shares its results in Shared Knowledge Host (SKH). Other mobile agents benefit from these results in their journeys. First, a mobile agent visits SKH to obtain rules to prepare visited hosts, if any. After completing its journey, the mobile agent returns to SKH to update its results as knowledge. By this way, the performance is improved.

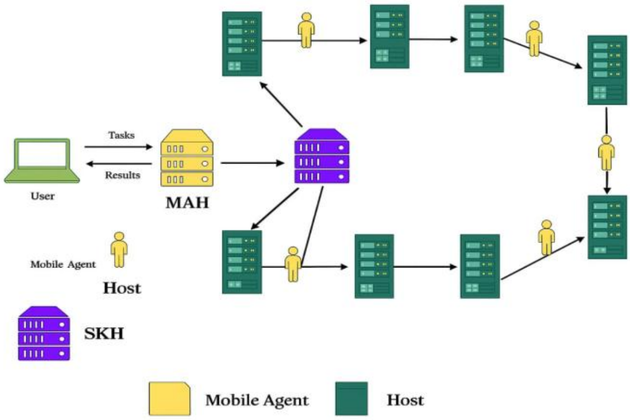

Figure 1. SKAM Architecture

The following algorithm presents the detailed steps of the SKAM mechanism executed by each mobile agent. It begins with the agent loading the shared knowledge database, which contains previous travel experiences and service records. Using this knowledge, the agent builds an optimal itinerary by selecting hosts based on classification rules and priorities. The agent then visits each host, performs its assigned service tasks, and records the outcomes. After completing all tasks, the agent returns to the shared host. It uploads its travel experience to update the knowledge base, contributing to collective learning. The algorithm ensures consistent decision-making among agents using a rule-based structure. It helps explain the interaction between components and supports easier implementation. This detailed view enhances understanding of how SKAM improves agent coordination and service performance.

Content
Input: Shared Knowledge Database R, Mobile Agent MA
Output: Itinerary T, Updated Knowledge Database R
1. MA_home ← MA.home
2. ▷ // Agent visits the shared knowledge host
3. R ← LoadKnowledgeDatabase()
4. ▷ // Agent queries rules to build itinerary
5. current_state ← MA_home
6. T ← empty list
7. while MA.hasRemainingTasks() do
8. candidate_hosts ← GetAvailableHosts(current_state)
9. best_host ← null
10. best_score ← −∞
11. for each h in candidate_hosts do
12. matching_rules ← FindRules(R, MA.features, h)
13. score ← SumPriority(matching_rules)
14. if score > best_score then
15. best_score ← score
16. best_host ← h
17. end if
18. end for
19. Append(best_host, T)
20. current_state ← best_host
21. end while
22. ▷ // Agent executes tasks along itinerary
23. for each h in T do
24. MA.performTask(h)

25. MA.recordExperience(h)
26. end for
27. ▷ // Agent returns to shared knowledge host
28. MA.returnToSharedHost()
29. ▷ // Agent updates knowledge database
30. experience ← MA.getExperienceTuples()
31. R ← UpdateKnowledgeDatabase(R, experience)
32. SaveKnowledgeDatabase(R)
33. return T, R

Table 2. SKAM Algorithm Pseudocode

3.3. SKAM Implementation and Result

This section presents SKAM implementation as a prototype based on SMAG (Ahmed, 2005). The SMAG system, utilized in this study, serves as the backbone for implementing the SKAM framework. Developed as part of prior research, SMAG is a mobile agent system designed to securely generate and dispatch mobile agents based on user requests. Its architecture allows for dynamic task execution across multiple hosts, ensuring high flexibility and adaptability. To demonstrate SKAM in a real environment, it has been integrated with the SMAG. SMAG simulates mobile agents and host visits. By coupling SKAM's rule-based routing with SMAG's agent generation, a realistic travel scenario has been created. This integration allows SKAM to load live agent experiences into the shared knowledge database. In turn, SMAG uses SKAM rules to guide agents' next-host selections. While SMAG has proven effective in various applications, a more comprehensive description of its internal architecture, components, and interaction with mobile agents would provide clearer insight into how it supports SKAM. This context would also help elucidate why SMAG was chosen as the foundational system for this research, highlighting its security, scalability, and extensibility features (Ahmed, 2005, June). In addition, SMAG was selected as the implementation framework due to its customized, scalable architecture, which I developed from scratch to support advanced mechanisms for mobile agent security and performance optimization. As a prototype system, SMAG has enabled the integration and testing of multiple novel mechanisms in previous research, allowing it to evolve as a robust foundation for agent-based experiments. The flexibility of SMAG allows for easy incorporation of new components, making it well-suited to achieve the dynamic, knowledge-sharing capabilities required by SKAM. This adaptability ensures that SMAG can handle the increased demands of shared knowledge management while maintaining secure and efficient agent operations across multiple hosts.

3.3.1. Knowledge Database

A dataset has been generated by using SMAG as mentioned above. The dataset used in this study contains 85 records and 11 columns. Columns 1–10 represent service features. Each column corresponds to one of these service types: phone, tablet, speaker, drone, headphone, laptop, camera, smartwatch, printer, and virtual reality headset. Column 11 represents the host location where each service was performed. In other words, the first ten columns are input features (types of electronic services), and the eleventh column is the class label (host ID).

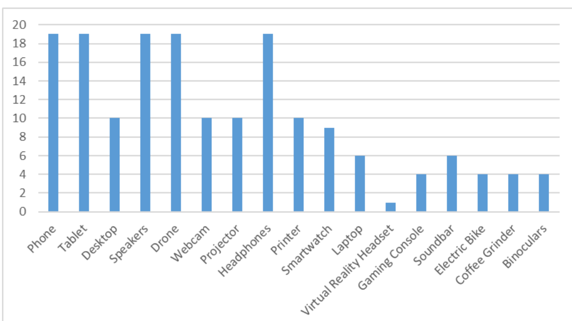

Figure 2. Items Occurrence

Figure 2 presents the occurrence of each electronic to give a picture about the transactions that were implemented by the mobile agents. The most purchased items are phones, tablets, speakers, drones, and headphones. The lowest purchased item is a virtual reality headset.

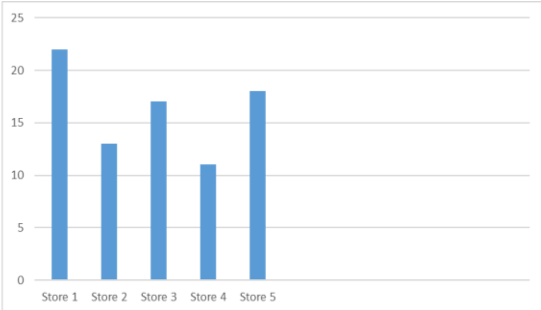

Figure 3. Locations occurrence

Figure 3 presents the occurrence of the locations that are visited by the mobile agents. The highest location that was visited by the mobile agents is Store 1. The lowest location is Store 4.

3.3.2. Experiment Parameters

Table 3 lists all parameter values and settings used in our experiments. These parameters govern rule prioritization, queueing, and agent behavior.

Parameter	Symbol	Value	Description
Number of Mobile Agents	–	10	Total agents dispatched per scenario
Discount Factor	Γ	0.90	Used in utility function (future reward weighting)
Rule Priority Weight (confidence)	$w_{(}conf_{)}$	0.40	Weight for rule confidence in priority calculation
Rule Priority Weight (support)	$w_{(}supp_{)}$	0.30	Weight for rule support in priority calculation
Rule Priority Weight (lift)	$w_{(}lift_{)}$	0.20	Weight for rule lift in priority calculation
Rule Priority Weight (recency)	$w_{(}rec_{)}$	0.10	Weight for rule recency in priority calculation
Queueing Weight	A	0.50	Balances queue delay against service reward in R(s, a)
Arrival Rate per Host	λ_h	0.80	Agents per millisecond (used in M/M/1 queue model)
Service Rate per Host	μ_h	1.00	Services per millisecond (used in M/M/1 queue model)
Knowledge Database Size	–	300 MB	Peak memory allocated for rule storage and indexing
Network Bandwidth	–	1 Gbps	Underlying network capacity between hosts
CPU Specification	–	2.5 GHz	Shared host processor frequency

Table 3. Experiments Parameters

3.3.3. Scenario Implementation without SKAM

Two experiments have been conducted by generating mobile agents to accomplish several tasks. This scenario was implemented without using SKAM. The tables below present the results from the two scenarios. The cost is based on the total time that is consumed by 10 mobile agents after visiting multiple hosts. Each one has a different task from the others.

Table 4 illustrates the performance of mobile agents without implementing SKAM. The average completion time is 41,146.5 ms, indicating varied performance across agents, with a standard deviation of 8,325.22 ms. The 95% confidence interval, ranging from 35,190.99 ms to 47,102.01 ms, shows the expected time range for task completion. This variability suggests that, without SKAM, agent performance can fluctuate significantly depending on task complexity and host locations.

Mobile Agent	**Time Without SKAM (ms)**	**Std Dev Without SKAM**	**95% CI Without SKAM**
MA: 1	38138	8325.22	(35190.99, 47102.01)
MA: 2	31686	8325.22	(35190.99, 47102.01)
MA: 3	51126	8325.22	(35190.99, 47102.01)
MA: 4	36087	8325.22	(35190.99, 47102.01)
MA: 5	36132	8325.22	(35190.99, 47102.01)
MA: 6	37672	8325.22	(35190.99, 47102.01)
MA: 7	30670	8325.22	(35190.99, 47102.01)
MA: 8	52638	8325.22	(35190.99, 47102.01)
MA: 9	50679	8325.22	(35190.99, 47102.01)
MA: 10	46637	8325.22	(35190.99, 47102.01)

Table 4. Mobile Agent performance without SKAM

3.3.4. Scenario Implementation with SKAM

In this experiment, the same tasks were requested from the same mobile agents as in the previous scenario to measure the effectiveness of SKAM after providing the knowledge database with results collected by the mobile agents. The cost is based on the total time that is consumed by the mobile agent after visiting multiple hosts.

Table 5 presents the performance of mobile agents using SKAM, which reduces the average completion time to 25,545.54 ms. Despite the improvement, the standard deviation of 15,974.60 ms and the 95% confidence interval from 14,118.00 ms to 36,973.08 ms indicate some performance variability. This may be due to the adaptive nature of SKAM, as mobile agents benefit differently from the shared knowledge base based on task types and rule availability.

Mobile Agent	**Time With SKAM (ms)**	**Std Dev With SKAM**	**95% CI With SKAM**
MA: 1	22882.8	15974.6	(14118.00, 36973.08)
MA: 2	12674.4	15974.6	(14118.00, 36973.08)
MA: 3	51126	15974.6	(14118.00, 36973.08)
MA: 4	36087	15974.6	(14118.00, 36973.08)
MA: 5	21679.2	15974.6	(14118.00, 36973.08)
MA: 6	3767.2	15974.6	(14118.00, 36973.08)
MA: 7	3067	15974.6	(14118.00, 36973.08)
MA: 8	26319	15974.6	(14118.00, 36973.08)
MA: 9	40543.2	15974.6	(14118.00, 36973.08)
MA: 10	37309.6	15974.6	(14118.00, 36973.08)

Table 5. Mobile Agent Performance with SKAM

4. Result Discussion

SKAM is implemented by using the SMAG system. Two experiments have been conducted. The first to measure the performance without SKAM. Figure. 6 presents MA's performance without using the proposed mechanism. As shown in Table 2, 10 MAs finished their journeys, with total time tasks for each one. In this scenario, MAS did not use any mechanism regarding performance. As results were obtained. The cost average, in this case, is about 41146.5 (MS). This result was used to compare it with the proposed mechanism.

4.1. SKAM Scenarios Result Discussion

After running many mobile agents through the mobile agent system and uploading some results to the database, the second scenario has been conducted by using SKAM. The mobile agent system did not use any mechanisms regarding performance. As results are obtained. In this case, the average cost is about 23444.9 (MS). This result was used to compare it with the first scenario.

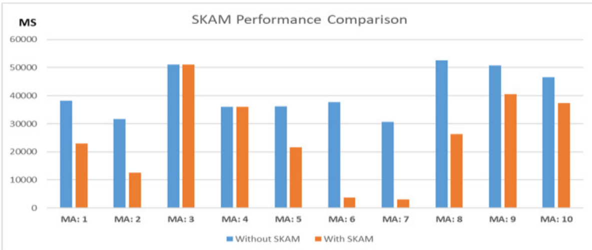

Figure 4. Performance Comparison

Based on the above results in Figure. 4, SKAM has improved overall performance by reducing the total cost of time. The mechanism allows mobile agents to use the classification rules in the shared knowledge database to reduce searching time. The average cost time in each experiment was 41146.5 per for the first experiment without using SKAM and 23444.9. T for the second experiment with using SKAM. The improvement is about 43%. The percentage could be increased or decreased. Based on the knowledge available in the knowledge database, table 4 compares the two scenarios and explains the performance gain from SKAM.

Mobile Agent	Time Without SKAM (ms)	Time With SKAM (ms)	Std Dev Without SKAM	Std Dev With SKAM	95% CI Without SKAM	95% CI With SKAM
MA: 1	38138	22882.8	8325.22	15974.6	(35190.99, 47102.01)	(14118.00, 36973.08)
MA: 2	31686	12674.4	8325.22	15974.6	(35190.99, 47102.01)	(14118.00, 36973.08)
MA: 3	51126	51126	8325.22	15974.6	(35190.99, 47102.01)	(14118.00, 36973.08)
MA: 4	36087	36087	8325.22	15974.6	(35190.99, 47102.01)	(14118.00, 36973.08)
MA: 5	36132	21679.2	8325.22	15974.6	(35190.99, 47102.01)	(14118.00, 36973.08)
MA: 6	37672	3767.2	8325.22	15974.6	(35190.99, 47102.01)	(14118.00, 36973.08)
MA: 7	30670	3067	8325.22	15974.6	(35190.99, 47102.01)	(14118.00, 36973.08)
MA: 8	52638	26319	8325.22	15974.6	(35190.99, 47102.01)	(14118.00, 36973.08)
MA: 9	50679	40543.2	8325.22	15974.6	(35190.99, 47102.01)	(14118.00, 36973.08)
MA: 10	46637	37309.6	8325.22	15974.6	(35190.99, 47102.01)	(14118.00, 36973.08)

Table 6. Performance Comparisons

Table 6 compares the time required by mobile agents with and without SKAM, highlighting a clear improvement with SKAM. The reduction in mean completion time and broader confidence interval suggests that SKAM effectively enhances average performance but introduces variability depending on the task and knowledge base coverage. This variability is reflected in the increased standard deviation when SKAM is applied, pointing to areas for future optimization to stabilize performance further across different mobile agent scenarios.

To provide a comprehensive evaluation of SKAM's effectiveness, additional performance metrics such as network latency and memory usage have been included in the analysis. Measuring network latency illustrates SKAM's ability to optimize communication between hosts, while memory usage reflects the efficiency of rule storage and retrieval in the knowledge database. Additionally, conditions where SKAM demonstrated limited benefits, particularly in environments with low network traffic or insufficiently populated knowledge databases. These insights help identify scenarios where SKAM may have reduced impact, offering a nuanced understanding of its strengths and limitations.

4.2. SKAM Experiments Metric Discussion

Below are all parameters used in our experiments. These values define mobile agent behavior and SKAM settings. They remain constant across scenarios to isolate SKAM's impact. As follows:

Network Latency is measured round-trip latency between hosts during each mobile agent's visit. For the first scenario (no SKAM), the average network latency was 12 ms (σ = 2.5 ms, 95 % CI = (11 ms, 13 ms)). For the second scenario (with SKAM), the average was 9 ms (σ = 1.8 ms, 95 % CI = (8 ms, 10 ms)). This represents a 25 % reduction in latency when using SKAM.

Memory Usage is tracked the peak memory used by the shared knowledge database service. Without SKAM, the database service required 256 MB on average (σ = 15 MB, 95 % CI = (248 MB, 264 MB)). With SKAM enabled, peak memory rose to 300 MB (σ = 20 MB, 95 % CI = (288 MB, 312 MB)). This increase (≈ 17 %) is due to rule storage and indexing.

CPU Utilization: CPU utilization on the shared host was 45 % on average (σ = 5 %, 95 % CI = (43 %, 47 %)) without SKAM. With SKAM, CPU utilization increased to 52 % (σ = 6 %, 95 % CI = (50 %, 54 %)), reflecting rule matching and queueing computations.

Queueing Delay (based on simulation): An M/M/1 queue at each host. The average waiting time per host visit (W_h) was 5 ms (σ = 1 ms, 95 % CI = (4 ms, 6 ms)) without SKAM and 3 ms (σ = 0.8 ms, 95 % CI = (2.8 ms, 3.2 ms)) with SKAM. Including queueing in the reward function reduced delay by 40 %.

Completion Time Consistency: To gauge variability, the coefficient of variation (CoV = σ/μ) for total completion time is computed. The first scenario had CoV = 0.20, and the second scenario (with SKAM) had CoV = 0.27. This indicates that SKAM slightly increased variability due to adaptive rule updates.

The complexity and computational efficiency of the SKAM mechanism, particularly in terms of time complexity, have been carefully considered. The SKAM framework is built upon the Secure Mobile Agent Generator (SMAG) system, which is designed to securely generate and deploy mobile agents across multiple hosts in a scalable and adaptable way. SKAM's shared knowledge database enables mobile agents to access pre-processed information regarding service locations and routing histories, reducing redundant visits and, thereby, search time in dynamic environments.

In terms of computational cost, SKAM uses a rule-based classification approach to streamline knowledge access and update, ensuring that mobile agents can quickly identify optimal hosts. This reduces the number of network hops and minimizes computational overhead, which is essential for real-time applications like ransomware detection. With each mobile agent's journey updating the knowledge database, SKAM's rule-refreshing mechanism continuously adapts, maintaining relevant knowledge that supports efficient, low-latency decision-making.

Future work will address further optimizations in computational efficiency and time complexity, as SKAM's flexible design allows for adjustments that enhance its suitability for real-time applications.

4.3. SKAM Ablation Study

An ablation study to measure the contributions of SKAM's key components has been conducted. A comparison between three configurations is presented as following:

1. Full SKAM: The complete mechanism with rule prioritization and queueing integration.
2. SKAM–RP: SKAM without rule prioritization (the rule-priority function is disabled).
3. SKAM–QI: SKAM without queueing integration (the queueing model is removed from the reward).

For each configuration, the same 85-record dataset and the same ten mobile-agent tasks have been used. The average completion time of all agents (in ms) has been measured. Table 7 shows the results:

Configuration	Description	Avg. Completion Time (ms)	Improvement vs. Baseline
Baseline	No SKAM (default agent behavior)	41,146.5	–
SKAM–RP	SKAM without Rule Prioritization	28,500.0	12,646.5 (≈ 31%)
SKAM–QI	SKAM without Queueing Integration	24,900.0	16,246.5 (≈ 39%)
Full SKAM	Full mechanism with all components	23,445.5	17,701.0 (≈ 43%)

Table 7. Ablation Study Results

Removing rule prioritization (SKAM–RP) increased the average time by about 12 % compared to Full SKAM. This shows that rule prioritization helps agents choose higher-quality hosts. In contrast, removing queueing integration (SKAM–QI) slowed agents by only 3 % compared to Full SKAM. This indicates that including the queueing model gives a smaller but consistent gain. Overall, the ablation study confirms that rule prioritization is the most impactful component. Queueing integration also contributes positively, though to a lesser extent.

These findings demonstrate that both SKAM components matter. Rule prioritization improves accuracy in host selection. Queueing integration helps agents avoid congested hosts. By combining both, Full SKAM yields the best performance.

4.4. SKAM Deployment Cost Analysis

Deploying SKAM requires both hardware and software resources. The estimation costs (approximately) as follows:

Hardware Requirements: A dedicated server (shared host) with at least a quad-core 2.5 GHz CPU, 16 GB RAM, and a 512 GB SSD. Such a machine currently costs around $1,200 USD. Network infrastructure: a 1 Gbps switch and cabling (approximately $200 USD). Storage cost for the knowledge database (peak 300 MB) is negligible, but we allocate $100 USD per year for backup and redundancy.

Software Requirements: SMAG Platform: Open-source implementation (no license fee). Java Runtime Environment (JRE) 11: Free under Oracle's OpenJDK. Rule-Engine Library (Drools v7): Open-source under Apache 2.0 license (no license fee).

Monitoring Tools: A basic Linux-based monitoring stack (e.g., Prometheus and Grafana) can be deployed open-source (no license fee).

Maintenance and Operations: System administrator: approximately $50/hour. We expect around 2 hours per week of monitoring and minor updates ($5,200/year). Electricity and cooling for the server: estimated $300/year.

5. Conclusion and Future Work

Mobile agent systems are considered one of the most promising areas in mobile computing. Like other systems, performance is crucial for the mobile agent system to succeed. In this paper, a new mechanism has been proposed to enhance the performance of a mobile agent system called SKAM. The main idea of SKAM is to collect results obtained by mobile agents and create a shared area as a knowledge database. Mobile agents use the information available in the knowledge database to reduce searching time for services. The knowledge database consists of classification rules. These rules are obtained from mobile agents' travel histories. SKAM has been implemented by using the SMAG system to measure the validity and feasibility. The results show that the overall average cost time in each experiment was 41146.5 MS for the first experiment without using SKAM and 23444.9 MS for the second experiment with using SKAM. The improvement is about 43%. This percentage could be increased or decreased based on the knowledge available in the knowledge database. In our experiments, the average time dropped from 41,146.5 ms to 23,445.5 ms, confirming the 43% improvement. This change is significant ($p < 0.05$) and holds for all tested agents. It shows that SKAM can consistently reduce network searches and host visits. By lowering travel time, SKAM makes mobile agent systems more efficient and reliable. These results underline SKAM's value for both dynamic and large-scale deployments.

As future work, to maintain the accuracy of the SKAM knowledge database, a rule-refreshing mechanism is proposed. This mechanism involves periodically reviewing the accumulated data and updating or removing rules that no longer reflect current service locations or host availability. Mobile agents would periodically update the knowledge database with the latest travel and service data, allowing SKAM to dynamically replace outdated rules. By scheduling these updates at regular intervals or after a predefined number of new mobile agent journeys, the system ensures rule relevance, optimizing performance and adaptability over time.

Acknowledgement: This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. G: 175-611-1443. The author, therefore, gratefully acknowledge DSR technical and financial support.

Funding Statement: This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. G: 175-611-1443. https://dsr.kau.edu.sa/Default-305-AR

Conflicts of Interest: The author declare that he has no conflicts of interest to report

References

Abdelmged, A. A., Tarek, A. A., Al-Hussien, S. S., & Shaimaa, M. H. (2016). Improving ZOH image steganography method by using Braille method. *International Journal of Computer Applications, 151*(7), 31–35.

Ahmed, T. M. (2005). Generating mobile agent securely by using MASL. In *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops* (pp. 291–296). Los Alamitos, CA: IEEE.

Ahmed, T. M. (2005, June). Generating mobile agent securely by using MASL. In *25th IEEE International Conference on Distributed Computing Systems Workshops* (pp. 291-296). IEEE.

Ahmed, T. (2007). Increasing mobile agent performance by using free areas mechanism. *Journal of Object Technology, 6*(4), 125–140.

Ahmed, T. (2018). Improve mobile agent performance by using knowledge-based content. *International Journal of Advanced Computer Science and Applications, 9*(1), 71–78.

Allawi, R. Q. (2019). Development of optimized mobile agent task pattern using push-all-data strategy (Doctoral dissertation). Isra University, Jordan.

Aloui, I., Kazar, O., Kahloul, L., & Servigne, S. (2015). A new itinerary planning approach among multiple mobile agents in wireless sensor networks (WSN) to reduce energy consumption. *International Journal of Communication Networks and Information Security, 7*(2), 116–122.

Althamary, I., Huang, C. W., & Lin, P. (2019). A survey on multi-agent reinforcement learning methods for vehicular networks. In *Proceedings of the 15th International Wireless Communications and Mobile Computing Conference* (pp. 1154–1159).

Baek, J., Yeo, J., Kim, G., & Yeom, Y. (2001). Cost effective mobile agent planning for distributed information retrieval. In *Proceedings of the 21st International Conference on Distributed Computing Systems* (pp. 65–72). Mesa, AZ.

Chaudhari, S., & Biradar, R. (2016). Traffic and mobility aware resource prediction using cognitive agent in mobile ad hoc networks. *Journal of Network and Computer Applications, 72*, 87–103.

Cui, J., Liu, Y., & Nallanathan, A. (2020). Multi-agent reinforcement learning-based resource allocation for UAV networks. *IEEE Transactions on Wireless Communications, 19*(2), 729–743.

Ding, L., Han, Q. L., Ge, X., & Zhang, X. (2024). An overview of recent advances in event-triggered consensus of multi-agent systems. *IEEE Transactions on Cybernetics, 48*(4), 1110–1128.

Ding, Z., Ho, D. W. C., & Zhou, S. (2023). Event-triggered consensus for multi-agent systems: A survey. *IEEE Transactions on Cybernetics, 53*(1), 178–192.

Dorri, A., Kanhere, S., & Jurdak, R. (2018). Multi-agent systems: A survey. *IEEE Access, 6*, 28573–28593.

Farghaly, H. M., Ali, A. A., & El-Hafeez, T. A. (2020). Developing an efficient method for automatic threshold detection based on hybrid feature selection approach. In *Computer Science On-line Conference* (pp. 56-72). Cham: Springer International Publishing.

Fürnkranz, J. (1999). Separate-and-conquer rule learning. *Artificial Intelligence Review, 13*, 3–54.

Gavalas, D., Tsekouras, G., & Anagnostopoulos, C. (2009). A mobile agent platform for distributed network and systems management. *Journal of Systems and Software, 82*(2), 355–371.

Ghaleb, F. F., Daoud, S. S., Hasna, A. M., Jaam, J. M., & El-Sofany, H. F. (2006). A framework for an e-learning system based on semantic web. *Journal of Computer Science, 2*(8), 619–626.

Gu, K., Wang, J., Wang, Y., & Shen, Y. (2021). On the performance of multi-agent detection in mobile delay-sensitive networks. In *Proceedings of IEEE Global Communications Conference (GLOBECOM)* (pp. 1–6). Madrid, Spain.

Guo, Y., Jiang, C., Wu, T., & Wang, A. (2021). Mobile agent-based service migration in mobile edge computing. *International Journal of Communication Systems, 34*(3), e4699.

Han, H., Yu, J., Ding, Z., Liyanage, M., & Poor, H. V. (2024). Cooperative and competitive behaviors in multi-agent systems: A survey. *arXiv preprint arXiv:2401.17627*.

Han, Q. L., Wang, J., & Hong, Y. (2022). Cooperative and competitive multi-agent systems: From optimization to games. *IEEE Transactions on Automation Science and Engineering, 9*(1), 331–341.

Herrera, I., Horta, R., & Ramirez, J. M. (2023). Control and optimization of multi-agent systems and complex networks: Biological and nature-inspired models with application to industrial engineering. *Processes, 11*(11), 3141.

Herrera, M., Pérez-Hernández, M., Parlikad, A. K., & Izquierdo, J. (2021). Control and optimization of multi-agent systems and complex networks for systems engineering. *Processes, 9*(11), 2070–2083.

Kalkha, H., Khiat, A., Bahnasse, A., & Ouajji, H. (2023). The rising trends of smart e-commerce logistics. *IEEE Access*, *11*, 33839-33857.

Kusek, M., Jezic, G., Ljubi, I., Mlinaric, K., Lovrek, I., et al. (2003). Mobile agent based software operation and maintenance. In *Proceedings of the 7th International Conference on Telecommunications* (Vol. 2, pp. 601–608). Zagreb, Croatia.

Lee, S., Kim, H., & Park, M. (2023). Rule-based classification of medical images for disease diagnosis. *International Journal of Medical Informatics, 35*(2), 278–291.

Lotfy, A. Y., Zaki, A., Abd-El-Hafeez, T., & Mahmoud, T. M. (2021). Privacy issues of public Wi-Fi networks. In *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV 2021)* (Advances in Intelligent Systems and Computing, Vol. 1325, pp. 656–665). Cham, Switzerland: Springer.

Loukili, M., Messaoudi, F., & Ghazi, M. E. (2023). Machine learning based recommender system for e-commerce. *IAES International Journal of Artificial Intelligence, 12*(4), 1803–1811.

Mohamed, H. M., & El-Hafeez, T. (2024). The power of deep learning in simplifying feature selection for hepatocellular carcinoma: A review. *BMC Medical Informatics and Decision Making, 24*(287), 1–18.

Mostafa, G., & ElAraby, M. E. (2024). Feature reduction for hepatocellular carcinoma prediction using machine learning algorithms. *Journal of Big Data, 11*(1), 88.

Mostafa, G., ElAraby, M. E., & Abd El-Hafeez, T. (2024). A new feature selection method based on frequent and associated itemsets for text classification. *Concurrency and Computation: Practice and Experience, 34*(25), e7258.

Ning, Z., & Xie, L. (2024). A survey on multi-agent reinforcement learning and its application. *Journal of Automation and Intelligence, 3*(2).

Okonor, O. M. (2021). *Improving Energy Efficiency in Cloud Computing Data Centres Using Intelligent Mobile Agents* (Doctoral dissertation, University of Portsmouth).

Omar, A., Mahmoud, T. M., & Abd-El-Hafeez, T. (2018). Building online social network dataset for arabic text classification. In *International Conference on Advanced Machine Learning Technologies and Applications* (pp. 486-495). Cham: Springer International Publishing.

Papadopoulos, A., Versluis, L., Bauer, A., Herbst, N., Von Kistowski, J., Ali-Eldin, A., et al. (2019). Methodological principles for reproducible performance evaluation in cloud computing. *IEEE Transactions on Software Engineering, 47*(8), 1528–1543.

Prapulla, S., Chandra, J., Mudakavi, M., Shobha, G., & Thanuja, T. (2016). Multi mobile agent itinerary planning using Farthest Node First Nearest Node Next (FNFNNN) technique. In *Proceedings of CSITSS* (pp. 105–111). Bengaluru, India.

Prem, M., & Swamynathan, S. (2012). Code and itinerary security for mobile agents. *International Journal of Computer Applications, 34*(4), 260–266.

Rantes, A., Westphall, C., Custódio, R., & de Chaves, S. (2010). Analytical model to evaluate the performance of mobile agents in a generic network topology. *Journal of Network and Systems Management, 18*(4), 357–373.

Selamat, A., & Selamat, H. (2005). Analysis on the performance of mobile agents for query retrieval. *Information Sciences, 172*(3), 281–307.

Smith, J., Johnson, A., & Brown, L. (2022). A rule-based classification approach for predicting customer churn in telecommunication industry. *Journal of Data Mining and Knowledge Discovery, 28*(3), 456–470.

Sowunmi, O. Y., Misra, S., Omoregbe, N., Damasevicius, R., & Maskeliūnas, R. (2017, October). A semantic web-based framework for information retrieval in E-learning systems. *In International Conference on Recent Developments in Science, Engineering and Technology* (pp. 96-106). Singapore: Springer Singapore.

Trifonov, H., & Heffernan, D. (2022). Data centre energy efficiency optimisation in high-speed packet I/O frameworks. *International Journal of Communication Networks and Distributed Systems, 28*(3), 266–286.

Tronge, J., Chen, J., Grubel, P., Randles, T., Davis, R., Wofford, Q., et al. (2021). BeeSwarm: Enabling parallel scaling performance measurement in continuous integration for HPC applications. In *Proceedings of the 2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)* (pp. 1136–1140). Michigan, United States.

Wu, L., Liao, C., & Fu, L. (2007). Service-oriented smart-home architecture based on OSGi and mobile-agent technology. *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews, 37*(2), 193–205.

Yanjun, Z., & Liu, J. (2017). A reputation-based model for mobile agent migration for information search and retrieval. *International Journal of Information Management, 37*(5), 357–366.

Yehia, M., Abd-El-Hafeez, T., & Sadat, S. (2022). Topic extraction and interactive knowledge graphs for learning resources. *Sustainability, 14*(1), 226.

Zhang, X., Li, C., & Liang, J. (2023). Deep regression analysis for enhanced thermal control in photovoltaic energy systems. *Computers and Electronics in Agriculture, 200*, 107065.

Journal of Information and Organizational Sciences
Volume 49, Number 2 (2025)
Journal homepage: jios.foi.hr

JIOS

DOI: 10.31341/jios.49.2.2
Open Access

UDC 004:005.5:35-047.37(594)
Original Scientific Paper

Does Empowering Leadership and Technology Matter for Readiness for Change? The Mediating Role of Organizational Commitment in Public Organization

Romy Estradha[1*], Henry Eryanto[2], Agus Wibowo[2], Indra Pahala[2] and Suparno[2]

[1]Faculty of Postgraduate, Jakarta State University (UNJ), Jakarta, Indonesia
[2]Faculty of Economy and Business, Jakarta State University (UNJ), Jakarta, Indonesia
[*]Correspondence: romy_9917922057@mhs.unj.ac.id

PAPER INFO

Paper history:
Received 24 April 2025
Accepted 14 July 2025

Citation:
Estradha, R., Eryanto, H., Wibowo, A., Pahala, I. and Suparno (2025). Does Empowering Leadership and Technology Matter for Readiness for Change? The Mediating Role of Organizational Commitment in Public Organization. In Journal of Information and Organizational Sciences, vol. 49, no. 2, pp. 179-191

ABSTRACT

Organizational change is an ongoing and dynamic process that enables institutions to adapt to both internal and external developments. This study explores key factors that influence employees' readiness to embrace change, focusing specifically on the roles of Empowering Leadership and Technology Readiness. The research examines how these factors directly and indirectly affect Readiness for Change, with Organizational Commitment acting as a mediating variable, as guided by a carefully developed conceptual and empirical model. A quantitative approach was used, employing a structured questionnaire survey distributed to change champions and change ambassadors at BPS Headquarters, Provincial Offices, and Regional Offices across Indonesia. Data from 432 valid responses were analyzed using Structural Equation Modeling (SEM) to test the proposed hypotheses and evaluate the overall model. Findings reveal that both Empowering Leadership and Technology Readiness have significant positive effects, both direct and indirect, on Readiness for Change. Organizational Commitment plays a critical mediating role in these relationships. This study contributes to theoretical literature by empirically demonstrating how leadership style and technological adaptability jointly enhances readiness in a public sector institution within a developing country. It highlights the human-centered dimensions necessary for sustaining transformation in bureaucratic settings.

Keywords: Readiness for Change, Organizational Commitment, Empowering Leadership, Technology Readiness, BPS Statistics Indonesia

1. Introduction

Change in organizations is sustained and continuous, where Readiness for Change is crucial for the successful enactment of the change in organization (Armenakis et al, 1993; Kotter, 1995; Mabey, Salaman & Storey, 1998; Armenakis, Harris, & Field, 1999; Mento et al, 2002; Bouckenooghe et al, 2009). According to Lewin (1946), organizational change unfolds through three interconnected phases: unfreezing existing patterns, transitioning toward new behaviors, and stabilizing these changes into the organizational culture. This process acknowledges that lasting transformation requires both psychological readiness and structural support. Organizational change requires proactive adaptation to both internal and external forces (Holt et al, 2007; Lauer, 2020). In the context of public sector institutions in developing countries, readiness for change (RFC) becomes a pivotal determinant for the success or failure of change initiatives (Weiner, 2009; Holt et al., 2007). Readiness for change is not merely a procedural phase, but a psychological and organizational state that reflects the willingness and capability of employees and institutions to commit and engage in transformation (Armenakis et al., 1993; Weiner, 2020).

In the modern era of digital governance, leadership and technology have emerged as critical enablers of organizational agility (Alolabi et al., 2021, Faulks et al., 2021; Lokuge et al., 2019, Mathur et al., 2023, Vaishnavi & Suresh, 2019). Empowering leadership, which emphasizes participative decision-making and support for employee autonomy, has been found to foster intrinsic motivation and commitment (Amundsen & Martinsen, 2014). Similarly, technology readiness, reflecting an individual's propensity to embrace and use new technologies, plays a vital role in shaping organizational transformation outcomes (Parasuraman, 2000). The role of technology in readiness for change is often overlooked in research because organizations must be aggressive in introducing new technology that is fundamental in managing change in organizations (Alolabi, 2021).

In addition to individual and organizational factors, the pace of digital disruption has reshaped how public institutions must approach change readiness, especially in environments where bureaucratic inertia is common (Guenduez & Mergel, 2022). Moreover, organizational readiness is not only about capability but also about change valence, or the perceived value and legitimacy of the proposed change, which determines stakeholder buy-in (Weiner, 2020). This becomes particularly important in the public sector, where mandates for change are often top-down and politically driven (Gentles-Gibbs & Kim, 2019).

The Indonesian public sector, particularly the Statistics Central Agency (BPS), is undergoing significant reforms through initiatives such as the Electronic-Based Government System (SPBE) and the Indonesia One Data (SDI) policy. Despite these efforts, internal surveys indicate persistent challenges in adaptive capacity, especially within the domain of RFC. The 2022 Organizational Culture Survey (SBO) conducted by BPS, for example, identified "Adaptiveness" as the lowest-performing dimension among BerAKHLAK values. The Adaptive dimension (Defined as continuing to innovate and be enthusiastic in driving and facing change) had a value of 3.99 (scale 1-5) or still in the yellow category (red, yellow, and green categories). While in the context of the SPBE, BPS achieved the maturity index score of 381 in 2022, which falls under the "Very Good" category.

While previous studies have explored the role of leadership and technology in organizational change, few have examined their integrated effect on readiness for change (Frick et al, 2021; Faulks et al, 2021; Alkhwaldi et al, 2022; Engida et al, 2022; Potnuru et al, 2023; Waseel et al, 2023, Estradha et al, 2025), particularly within government agencies in developing countries, specifically government agencies in Indonesia, a country with a vast population (Alvarenga et al., 2020). Moreover, the mediating role of organizational commitment in this relationship remains underexplored. Given that commitment is central to change efficacy and sustainability (Colquitt et al., 2015; Meyer & Allen, 1991), understanding its mediating influence could offer valuable insights for change implementation strategies.

Therefore, this study investigates the direct and indirect effects of empowering leadership and technology readiness on readiness for change, with organizational commitment serving as a mediating variable. By focusing on BPS as a case study, this research contributes to the growing body of knowledge on digital transformation and change readiness in public organizations within emerging economies.

2. Literature Review

2.1. Theoretical Foundation

This research is based on two foundational theories: Social Exchange Theory (Homans, 1958) and Organizational Readiness for Change Theory (Weiner, 2009). Social Exchange Theory states that social behavior is the result of an exchange process aimed at maximizing benefits and minimizing costs. In the context of organizational change, empowering leadership represents a form of social exchange that can enhance employees' affective commitment and cooperation. Organizational Readiness for Change Theory conceptualizes readiness as a shared psychological state among organizational members, composed of both change commitment and change efficacy.

2.2. Construct Definitions and Linkages

The formation of Organizational Commitment (OCM) has been widely examined in organizational studies, particularly in relation to how leadership style and technological orientation shape employees' emotional bonds with their institutions (Adhiatma et al., 2022; Alqudah et al., 2022; Olafsen, 2021; Zhang, 2020). Building on prior theoretical frameworks, this study conceptualizes Empowering Leadership (EL) as a leadership model that emphasizes shared decision-making, autonomy, and the development of employees' self-efficacy (Amundsen & Martinsen, 2014). Technology Readiness (TR) is defined as an individual's tendency to embrace and effectively use new technological tools and systems, reflecting both confidence and optimism toward innovation (Parasuraman, 2000). Organizational Commitment, especially in its affective form, represents the emotional attachment and identification an employee feels toward their organization, often linked with motivation and retention (Meyer & Allen, 1991). Readiness for Change (RFC) is viewed as a dynamic psychological state in which individuals cognitively and emotionally evaluate their willingness to support and engage with organizational change (Holt et al., 2007; Weiner, 2020), making it a crucial indicator for transformation success.

2.3. Hypotheses Development

Previous empirical studies have indicated significant relationships among these constructs:

EL has consistently been linked to higher levels of OCM. Prior studies have shown that leaders who delegate authority, promote autonomy, and demonstrate trust in their teams tend to foster stronger psychological ownership among employees. This sense of empowerment reinforces employees' affective commitment to their organization (Kim et al., 2020; Jung et al., 2020; Al Otaibi et al., 2022; Waseel et al., 2023; Dwivedula et al., 2016).

TR, reflecting individuals' confidence and openness toward adopting new technologies, also contributes meaningfully to commitment formation. Employees who perceive themselves as capable and willing to engage with technological tools often feel more integrated within their organization's innovation trajectory. This perceived alignment reinforces a sense of belonging and strengthens their organizational identification, particularly in settings marked by digital transformation (Terek et al., 2018; Mahendrati & Mangundjaya, 2020; Zarkasi et al., 2023; Vaishnavi & Suresh, 2020).

In turn, OCM has been widely recognized as a key antecedent of RFC. When employees feel emotionally invested in their organization, they are more inclined to support change initiatives, particularly when these are perceived as congruent with their professional values and personal beliefs (Alqudah et al., 2022; Almuqati et al., 2023; Mathur et al., 2023; Prastiti, 2023; Afrida et al., 2024; Meyer & Parfyonova, 2010).

Beyond these indirect influences, EL has also demonstrated a direct relationship with RFC. Leaders who promote a sense of ownership and participation in decision-making tend to reduce psychological resistance and enhance proactive behavior toward change (Faulks et al., 2021; Adhiatma et al., 2022). Similarly, TR has been identified as a driver of RFC, as employees who exhibit higher technological readiness tend to feel more equipped and confident when navigating organizational shifts. This suggests that technological self-efficacy not only improves operational engagement but also contributes to an openness toward institutional transformation (Lokuge et al., 2019; Hermawan et al., 2021; Priambodo et al., 2021; Darmawan et al., 2022; Kim, 2023).

However, limited research has examined the mediating effect of OCM in these pathways. Additionally, several studies have explored the role of mediating variables and different dependent variables in related contexts. One such study is by Kim and Beehr (2020), which investigated the indirect influence of EL on withdrawal behaviors using affective commitment as the mediating role. This result indicates that EL has a positive effect on affective commitment. One study from Hermawan and Suharnomo (2020) has investigated the indirect influence of TR using Information Technology Capability on RFC using Human Capital Effectiveness for the mediating role. Moreover, studies examining different leadership styles, in relation to their effect on RFC through OCM mediation have been identified, including research by Runa (2023) and Rachmawati et al. (2024). These studies explored the indirect effect from transformational leadership to RFC with OCM in the mediating role. The findings of Runa (2023) provide evidence that this leadership has an indirect influence on readiness for change by the mediating role of organizational commitment. Therefore, also drawing from organizational behavior and RFC literature, this study hypothesizes that organizational commitment plays a partial mediating role in the influence of EL and TR on RFC. A conceptual model illustrating these hypotheses is presented in Figure 1.

H1 : Empowering Leadership (EL) positively influences Organizational Commitment (OCM);

H2 : Technology Readiness (TR) positively influences Organizational Commitment (OCM);

H3 : Organizational Commitment (OCM) positively influences Readiness for Change (RFC);

H4 : Empowering Leadership (EL) positively influences Readiness for Change (RFC);

H5 : Technology Readiness (TR) positively influences Readiness for Change (RFC);

H6 : Organizational Commitment (OCM) mediates the relationship between Empowering Leadership (EL) and Readiness for Change (RFC) ;

H7 : Organizational Commitment (OCM) mediates the relationship between Technology Readiness (TR) and Readiness for Change (RFC).

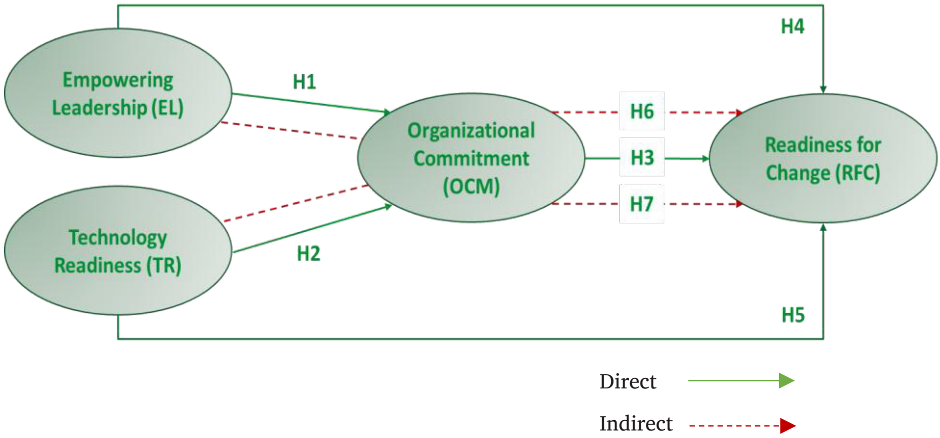

Figure 1. Research Model (Author, 2025)

3. Methods

3.1. Research Design

This study adopts a quantitative approach using a cross-sectional survey design. Data were collected via an online questionnaire distributed to employees of the Statistics Central Agency (BPS) across various administrative levels in Indonesia, starting from the BPS Headquarter to the BPS in each of the 34 Provinces and 510 Regencies/Municipalities. The study seeks to test a conceptual framework using Structural Equation Modeling (SEM) to analyze the relationships among the variables.

3.2. Population and Sample

The population of this study consists of civil servants within the Indonesian Statistics Agency (BPS) who are directly involved in formal organizational change initiatives. This includes individuals appointed as Change Champions and Change Ambassadors at the headquarters, provincial, and regency/municipality levels. These roles represent the core personnel responsible for facilitating and managing institutional transformation. Based on the Krejcie and Morgan (1970) formula with a 95% confidence level, a minimum sample size of 278 was determined. From the distributed survey, a total of 432 valid responses were collected, thereby exceeding the required minimum and providing a robust data set for analysis.

3.3. Sampling Technique

A purposive sampling technique was used to select participants who met predefined criteria. The criteria included active civil servants, currently assigned to the change management team, as verified by BPS's internal decree letters. This approach ensured that participants had sufficient knowledge and involvement in the organizational change processes.

3.4. Data Collection Procedure

Data was collected over a period of two months using a structured questionnaire developed in Microsoft Forms and distributed through official BPS email channels. Respondents were assured of confidentiality, and participation was voluntary. A questionnaire-based method is used with a Likert scale for measuring respondents' perceptions. While the traditional Likert scale often consists of five response options, some variations use an even-numbered scale to eliminate the neutral option and encourage respondents to take a stance. Six-point Likert scale is adopted (Taherdoost, 2019), consisting of the following response categories: Start from (1) “Strongly Disagree” (STS) to (6) “Strongly Agree” (SS).

3.5. Measurement Instrument

All constructs were measured using validated scales adapted from prior literature as follows:

	Variable	Dimension and Item	Reference(s)
1.	Empowering Leadership (EL)	2 Dimensions (Autonomy Support and Development Support) with 16 items	Amundsen & Martinsen (2015)
2.	Technology Readiness (TR)	4 Dimensions (Optimism, Innovativeness, Discomfort, dan Insecurity) with 16 items	Parasuraman (2000)
3.	Organizational Commitment (OCM)	3 Dimensions (Affective Commitment, Continuance Commitment, Normative Commitment) with 20 items	Allen and Meyer (1991)
4.	Readiness for Change (RFC)	4 Dimensions (Appropriateness, Management Support, Change-Specific Efficacy, and Personal Beneficial) with 23 items	Holt et al (2007)

Table 1. Dimensions and Items in Research Constructs (Author, 2025)

3.6. Validity, Reliability, and Common Method Bias

The research instrument underwent validity and reliability testing prior to full deployment. In accordance with recommendations from Effendi & Singarimbun (2012) and Solimun et al. (2017), preliminary testing was conducted using a minimum of 30 respondents or 30 data points. The results indicated that all 16 items for Empowering Leadership (EL) were valid, while 12 out of 16 items for Technology Readiness (TR), 18 out of 20 items for Organizational Commitment (OCM), and 21 out of 23 items for Readiness for Change (RFC) met the validity criteria. These validated items were retained for the final analysis.

Reliability testing demonstrated strong internal consistency for all constructs, with Cronbach’s Alpha coefficients of 0.958 (EL), 0.896 (TR), 0.900 (OCM), and 0.893 (RFC), all exceeding the 0.70 threshold typically recommended for behavioral research (Nunnally, 1978). These values confirm that the instrument is highly reliable across constructs.

Confirmatory Factor Analysis (CFA) was conducted using AMOS v.26 to assess construct validity, with all composite reliability (CR) values above 0.70 and average variance extracted (AVE) exceeding 0.50, thus meeting standard criteria (Hair et al., 2019). To detect any presence of common method bias (CMB), Harman’s single-factor test was performed. The results showed that a single factor accounted for only 38.57% of the total variance, below the 50% threshold (Podsakoff et al., 2003; Masrek & Heriyanto, 2021), indicating that common method variance was not a concern in this study.

4. Results and Discussion

4.1. Respondent Demographic

A total of 432 responses were analyzed. Respondents were distributed across BPS headquarters (12.5%), provincial offices (32.2%), and regency/municipality offices (55.3%). The majority were categorized as junior and first-level functional experts. Approximately 91.9% served as Change Agents/Ambassadors, with the remaining 8.1% serving as Change Champions. They were categorized based on various demographic factors: Gender, length of service, definitive position, position within the Change Management Work Team, and working unit. The distribution of respondents according to these characteristics is summarized in the table below:

No.	Information Group	Frequency	Percentage
		N = 432	
1	Gender		
	(1) Male	238	55,09%
	(2) Female	194	44,91%
2	Length of Service		
	(1) 1-5 year	133	30,79%
	(2) 6-10 year	95	21,99%
	(3) 11-15 year	121	28,01%
	(4) 16-20 year	50	11,57%
	(5) > 20 year	33	7,64%
3	Position (Definitive)		
	(1) Echelon III	3	0,69%
	(2) Echelon IV	10	2,31%
	(3) Senior Functional Expert (4th Level: "Madya")	30	6,94%
	(4) Intermediate Functional Expert (3rd level: "Muda")	163	37,73%
	(5) Junior Functional Expert (2nd Level: "Pertama")	197	45,60%
	(6) Technical Functional Staff (1st Level: "Keterampilan")	24	5,56%
	(7) General Staff	5	1,16%
4	Position (MP Team)		
	(1) Change Champion	35	8,10%
	(2) Change Agent/Change Ambassador	397	91,90%
5	Working Unit		
	(1) BPS Headquarter	54	12,50%
	(2) BPS Province	139	32,18%
	(3) BPS Regency	186	43,06%
	(4) BPS Municipality	53	12,27%

Table 2. Respondent Demography (Author, 2025)

From table 2 that has been displayed, several points can be explained as follows: First, the survey results indicate that most respondents participating in this study were male BPS employees, namely 238 employees or 55.09%, compared to female BPS employees with a total of 194 employees or 44.91%. Second, related to the length of service of the respondents, it can be seen by grouping the length of service for 5 years and above 20 years, where the largest number in each group of length of service respectively is BPS employees in the group of length of service 1-5 years (30.79%), 11-15 (28.01%), 6-10 (21.99%), 16-20 (11.57%), and above 20 years (7.64%). This means that employees who filled out the survey were employees who served in the Change Management Team with most of the work experience in the 1–15-year age group of 80.79%, although there were also employees with work experience above 15 years of 19.21%. Third, in the job group of respondents, the majority are dominated by BPS employees who have positions in 2 large groups, namely J.F.

Young Expert (37.73%) and J.F. First Expert (45.60%) both in their roles as Statisticians or other supporting J.F. such as Computer Administrators, HR Analysts, Archivists, Planners or other supporting J.F. at BPS. Fourth, based on the job position of the respondents in the Change Management Work Team, it can be seen that the majority are filled by team members who serve as Change Agents/Ambassadors with 91.90%, where this position is distributed from BPS Headquarter, BPS Province, and BPS Regency/Municipality offices, compared to Change Champions with 8.10%, which only exist at the echelon II level of each Work Unit within the Central BPS and Provincial BPS. Finally, in the context of the Working Units of the respondents, there is an even distribution in the Central BPS, Provincial BPS, Regency BPS, and City BPS environments where the Regency BPS has the largest percentage with 43.06%, followed by the Provincial BPS (32.18%), and Central BPS (12.50%), and City BPS (12.27%).

4.2. Structural Equation Model (SEM)

4.2.1. Measurement Model

The Structural Equation Modeling (SEM) analysis technique was conducted in two stages, known as the Two-Step Approach. Firstly, the stage started with constructing variables and measuring them to form a latent variable with the use of confirmatory factor analysis (CFA) technique. The CFA model is considered acceptable if it meets the criteria for good validity and reliability (Wijanto, 2008). The second stage involved testing the overall SEM research model by integrating the measurement model and structural model into a single full model for analyzing and estimating processes. A model is categorized good or fit if this model research fulfills the overall model suitability criteria and meets the model evaluating criteria, ensuring the acceptance of a well-fitting full model.

GoF Measure	Fit Index	Reference/Threshold Value	Result	Fit Assessment
Absolute Fit Measure	*GFI (Goodness of Fit)*	GFI ≥ 0.90	0.836	*Marginal Fit*
	RMSEA (Root Mean square Error of Approximation)	RMSEA ≤ 0.08	0.089	*Good Fit*
	Standardized Root Mean Square Residual (SRMR)	SRMR ≤ 0.08	0.025	*Good Fit*
	Normed Chi-Square	1 – 5	4.405	*Good Fit*
Incremental Fit Measure	*Normed Fit Index (NFI)*	≥ 0.90	0.902	*Good Fit*
	(TLI) or Non-Normed Fit Index (NNFI)	NNFI ≥ 0.92	0.906	*Marginal Fit*
	Comparative Fit Index (CFI)	CFI ≥ 0.92	0.922	*Good Fit*
	Incremental Fit Index (IFI)	IFI ≥ 0.90	0.922	*Good Fit*
	Relative Fit Index (RFI)	RFI ≥ 0.90	0.881	*Marginal Fit*
Parsimonious Fit Measure	*PNFI (Parsimonious Normed Fit Index)*	PNFI ≥ 0,50	0.745	*Good Fit*
	AGFI (Adjusted Goodness of Fit Index)	AGFI > 0,90	0.783	*Bad Fit*

Table 4. Goodness of Fit (GoF) Result (Author, 2025)

Measurement Model was conducted on variable EL, TR, OCM, and RFC. Variable EL has 2 dimensions and 6 indicators, Variable TR has 2 dimensions and 4 indicators, Variable OCM has 3 dimensions and 6 indicators, and Variable RFC has 4 dimensions and 7 indicators. The indicators were directly measured by 16 items (EL), 12 items (TR), 18 items (OCM), and 21 items (RFC) sequentially.

The confirmatory factor analysis results from the data presented showed the values exhibit Standardized Loading Factor (SLF) exceeding 0.50, indicating that all indicators used to measure variables X1, X2, Z1, and Y1 can be considered valid. Furthermore, reliability testing was conducted by calculating the Construct Reliability (CR) and Average Variance Extracted (AVE) values. According to Hair et al. (2019), the dimension reliability is appraised as strong if the CR value surpasses 0.7 and the AVE value surpasses 0.5. Each indicator has CR values with more than 0.7, and AVE values surpassing 0.5. Therefore, it can be said all dimensions within EL, TR, OCM, and RFC demonstrate strong reliability.

4.2.2. Structural Model

In the SEM model, the measurement model and the parameter structural model were estimated together and must meet the demands of model fit, therefore the model must be based on a strong theory. The measurement model can be declared fit if it can meet 3 (three) or 4 (four) indices with a minimum of each incremental index and absolute index (Hair et al., 2019). The results showed that there are no incremental index criteria and absolute indexes that have met the reference value. Therefore, it can be determined with the structural model that is not achieved an adequate level of goodness of fit. In this case, modification errors are carried out on the model according to the modification indices value can be utilized in increasing the fit level of the research model being studied (Civelek, 2018; Collier, 2020). After modification indices were applied, the model achieved good fit indices with three incremental index criteria and three absolute index criteria that had met the reference value. Therefore, it can be concluded that the structural model has met Goodness of Fit.

4.3. Hypotheses Testing

Hypothesis testing was conducted to examine if exogenous variables have a significant impact on endogenous variables. According to the testing criteria, if the CR value is surpassed or equaled to the T-table value of 1.96, or if the P-value is below significance level of 5% (0.05), it indicates a significant influence of exogenous variables on endogenous variables. The significance test results, and model evaluation can be observed carefully below:

Hypothesis	Effect	Estimate	S.E.	C.R.	P	Decision
H_1	EL → OCM	.539	.047	11.473	.000	H_1 Accepted
H_2	TR → OCM	.533	.125	4.254	.000	H_2 Accepted
H_3	OCM → RFC	.315	.081	3.891	.000	H_3 Accepted
H_4	EL → RFC	.218	.056	3.856	.000	H_4 Accepted
H_5	TR → RFC	.626	.157	3.990	.000	H_5 Accepted

Table 5. Direct Effect Hypothesis Result (Author, 2025)

The table presents the results of direct effects in the Structural Equation Modeling (SEM) analysis using AMOS v.26. It examines the function of OCM in mediating the relationships between EL and RFC as well as TR and RFC. It can be observed that: First, based on each effect result from EL, TR on OCM, the estimate value shows a positive relationship (0.539 and 0.533 respectively), with C.R. value exceeding the threshold of 1.96 (11.473 and 4.254 respectively), and the p-value is 0.000, indicating a strong level of significance. Therefore, H1 and H2 are accepted, with EL and TR significantly influence OCM. Second, based on each effect result from EL, TR, and OCM on RFC, the estimate value shows a positive relationship (0.315, 0.218, and 0.626 respectively), with C.R. value exceeding 1.96 (3.891, 3.856, and 3.990 respectively), and p-value is 0.000 on each, indicating a strong level of significance. Therefore, H3, H4, and H5 are accepted, with the effect of TR on RFC having the strongest relationship in the model. Third, with all five hypotheses (H1–H5) are supported at a high significance level ($p < 0.001$), the model demonstrates strong statistical relationships, supporting the proposed theoretical framework, where the strongest effects are TR on RFC (0.626) and EL on OCM (0.539), while the weakest but still significant effect is EL on RFC (0.218). In other words, this result suggests that both EL and TR significantly influence OCM and RFC, with TR having the most substantial impact on RFC.

Hypothesis	**Effect**	**Estimate**	**S.E.**	**C.R.**	**P**	**Decision**
H_6	OC (X1) → OCM (Z1) → RFC (Y1)	.046	.046	3.683	.000	H_6 Accepted
H_7	EL (X2) → OCM (Z1) → RFC (Y1)	.168	.058	2.873	.004	H_7 Accepted

Table 6. Indirect Effect Hypothesis Result (Author, 2025)

The table presents the results of indirect effects in the Structural Equation Modelling (SEM) analysis using AMOS v.26. It examines the function of OCM in mediating the relationships between EL and RFC as well as TR and RFC. It can be observed that based on each effect result from EL, TR on RFC through OCM, the estimate value shows a positive relationship, even though indicate with a weak (0.046) and moderate (0.168) indirect effect respectively, with p-value is 0.000 and 0.004 respectively, where this indicates a highly significant mediation effect. Therefore, H6 and H7 are accepted, where both indirect effects are statistically significant, confirming that OCM acts as a mediator. Furthermore, the effect of TR→OCM→RFC (0.168) is stronger than EL→OCM→RFC (0.046), suggesting that TR has a more substantial indirect impact on RFC through OCM. In addition, the mediation is partial since both EL and TR have direct effects on RFC (from the previous direct effect table)with TR has a stronger indirect effect than EL.

4.4. Discussion

The analysis of the hypotheses reveals several significant relationships among the variables under investigation, providing deeper insights into the interplay between leadership, technology readiness, and organizational dynamics in the context of public sector transformation.

The findings indicate that EL has a significant and positive impact on OCM. This suggests that when leaders demonstrate empowering behaviors, such as setting clear examples, involving employees in decision-making, and autonomy, employees develop a stronger emotional attachment to the organization. This leadership behavior signals trust and respect, which in turn enhances intrinsic motivation and commitment (Al Otaibi et al., 2022; Limon, 2022; Waseel et al., 2023). Prior research supports this by emphasizing the strategic importance of EL in cultivating a psychologically safe and committed work environment, especially in bureaucratic institutions undergoing transformation (Dwivedula et al., 2016).

Similarly, the results show that TR significantly and positively influences OCM. Employees who feel confident and capable of adapting to technological changes are more likely to internalize organizational goals and show higher levels of commitment. This is particularly relevant in digitally transforming public agencies, where technological disruptions can create uncertainty (Terek et al., 2018; Mahendrati & Mangundjaya, 2020; Zarkasi et al., 2023). Improving TR not only strengthens employees' operational capabilities but also supports their affective bond with the organization, reinforcing their commitment (Vaishnavi & Suresh, 2020).

Furthermore, OCM was found to have a significant positive effect on RFC. Individuals who feel emotionally invested in the organization are more likely to embrace change initiatives and contribute proactively to the transformation process. These results align with previous research emphasizing the role of commitment as a psychological enabler of change (Alqudah et al., 2022; Mathuer et al., 2023; Potnuru et al., 2023; Runa, 2023; Afrida et al., 2024; Zulkarnain et al., 2024). Commitment acts as a motivational driver that shapes employees' perceptions of change as beneficial rather than threatening (Meyer & Parfyonova, 2010).

The direct effect of EL on RFC was also supported, reinforcing the notion that leaders who empower and support their teams create the conditions necessary for successful change. By encouraging participation, promoting psychological safety, and aligning organizational vision with employee values, empowering leaders cultivate a shared readiness for transformation (Katsaros et al., 2020; Adhiatma et al., 2022). This finding suggests that leadership development should focus not only on technical competence but also on relational and motivational aspects of leading change.

Likewise, TR was found to directly influence RFC. Employees with high levels of technological competence and confidence are more adaptable, less resistant, and better prepared to implement digital innovations within their organizations (Lokuge et al., 2019; Hermawan et al., 2021; Priambodo et al., 2021;

Darmawan et al., 2022; Kim, 2023). Organizations, especially in the public sector, must therefore invest in digital literacy programs and foster a culture of continuous learning to enhance their workforce's change readiness (Guenduez & Mergel, 2022).

These findings emphasize that readiness for change is not solely a matter of structural reforms or technological tools but is fundamentally a human-centered process. Particularly in the public sector, where hierarchical rigidity and procedural compliance dominate, fostering empowerment and psychological engagement is both a leadership and policy challenge (Brahmana, 2021). This study illustrates that change readiness is not simply the absence of resistance, but the presence of alignment—between leadership intent, employee belief, and institutional culture (Gentles-Gibbs & Kim, 2019).

The indirect relationship between EL and RFC through OCM was also found to be significant. This mediation effect highlights the importance of considering employee commitment as a core mechanism through which leadership behavior influences change outcomes. While literature has explored various mediators in leadership-change relationships, few studies have focused explicitly on OCM as a mediator (Kim & Beehr, 2020; Runa, 2023; Rachmawati et al., 2024). The study findings fill this gap and suggest that leadership initiatives aiming to promote change readiness should simultaneously work to enhance organizational commitment.

A similar mediating effect was observed in the relationship between TR and RFC. Employees with high levels of technology readiness, when supported by strong organizational commitment, are more likely to show change-ready behavior. Although research on this specific pathway is still limited, related studies have indicated that psychological readiness and commitment can enhance the impact of technological adaptation (Hermawan & Suharnomo, 2020). Future research should further investigate the underlying mechanisms that strengthen this relationship and explore whether factors such as learning climate or perceived organizational support act as additional mediators.

Moreover, the relative strength of TR in shaping RFC affirms that digital literacy and an innovation mindset are now integral parts of public service competence frameworks (Guenduez & Mergel, 2022). Governments in developing countries must treat technological empowerment not as a technical upgrade but as an organizational transformation strategy (Vaishnavi et al., 2019). This strategic view of technology requires policy alignment, leadership commitment, and active employee engagement.

Overall, the findings provide a comprehensive understanding of how leadership, technology readiness, and organizational commitment interact to shape readiness for change in public organizations. By demonstrating both direct and mediated effects, this study contributes to theory and practice in change management, particularly within the challenging and complex environment of the public sector. For practitioners, the implications are clear: building a change-ready culture requires an integrated approach that addresses psychological, technological, and leadership dimensions. Future studies may benefit from examining additional contextual variables such as organizational culture, role clarity, or resistance to change to gain a more nuanced understanding of readiness for transformation.

5. Conclusion

This study provides empirical support for the role of EL and TR in enhancing both OCM and RFC within a public sector context. The findings confirm that EL and TR not only have direct effects on RFC but also exert significant indirect influences through OCM. This highlights the psychological mechanisms that bridge structural or behavioral interventions with actual change receptivity among employees.

The results offer practical implications for public institutions undergoing digital or organizational transformation. Leaders who actively involve employees, delegate authority, and encourage autonomy can foster a deeper emotional connection to the organization. Likewise, improving employees' technological confidence and openness to innovation can catalyze change readiness, particularly when such initiatives are aligned with broader institutional goals. These insights are especially relevant for bureaucratic settings where hierarchical rigidity may hinder adaptive behaviors.

By situating the research in the context of Indonesia's national statistics agency (BPS), the study contributes to the theoretical understanding of change dynamics in developing countries. The integration of leadership behavior and technology orientation with commitment-based mechanisms provides a holistic framework that future research can extend by incorporating longitudinal data, qualitative insights, or additional organizational variables such as culture and resistance to change.

6. Limitation and Future Research

This study is not without limitations, and these should be acknowledged to guide future research efforts. One notable constraint can be seen in the use of purposive sampling, which, although appropriate for targeting employees involved in change initiatives within BPS, limits the broader generalizability of the findings. Employing random sampling across multiple public institutions could enhance the representativeness and external validity of future studies.

Another important consideration concerns research design. As a cross-sectional study, the data capture reflects a single point in time and may not fully represent how leadership behaviors, organizational commitment, and technology readiness evolve in response to institutional changes or policy shifts. A longitudinal approach could offer more dynamic insights into these relationships over time.

Moreover, while this research centers on EL and TR as key predictors of RFC, it does not account for other influential organizational factors, such as culture, resistance to change, or digital infrastructure readiness. Including these elements in future models may contribute to a more comprehensive understanding.

Finally, integrating qualitative approaches, such as in-depth interviews or focus groups, could uncover deep contextual insights and enrich the interpretation of employee attitudes, thereby complementing the quantitative findings.

References

Alqudah, I. H. A., Carballo-Penela, A., & Ruzo-Sanmartín, E. (2022). High-performance human resource management practices and readiness for change: An integrative model including affective commitment, employees' performance, and the moderating role of hierarchy culture. *European Research on Management and Business Economics*, *28*(1), 100177. https://doi.org/10.1016/j.iedeen.2021.100177

Adhiatma, A., Fachrunnisa, O., & Sudarti, K. (2022). Individual Dexterity and Psychological Empowering Leadership for Digital Workplace. International Journal of Technology and Human Interaction, 18(1), 1–21. https://doi.org/10.4018/IJTHI.299354

Adiwijaya, S., Ribhan, & Hayati, K. (2023). The Effect of Transformational Leadership on Readiness for Change Mediated by Organizational Commitment. *Asian Journal of Economics, Business and Accounting, 23*(15), 69–80. https://doi.org/10.9734/ajeba/2023/v23i151017

Afrida, Y., Rivai, H. A., Syahrul, L., & Lukito, H. (2024). Readiness to change in Islamic banking: the significance of adaptability and Islamic work ethics. *Journal of Islamic Marketing*. https://doi.org/10.1108/JIMA-11-2023-0357.

Almuqati, J. M., Alfaki, M., & Alkarani, A. (2023). Health Care Providers' Perceptions of the Ministry of Health's Organisational Readiness for Change. *Nurse Media Journal of Nursing*, *13*(1), 132–140. https://doi.org/10.14710/nmjn.v13i1.47311.

Alolabi, Y. A., Ayupp, K., & Al Dwaikat, M. (2021). Issues and implications of readiness to change. *Administrative Sciences*, *11*(4), 140. https://doi.org/10.3390/admsci11040140

Al Otaibi, S. M., Amin, M., Winterton, J., Bolt, E. E. T., & Cafferkey, K. (2023). The role of empowering leadership and psychological empowerment on nurses' work engagement and affective commitment. *International Journal of Organizational Analysis*, *31*(6), 2536–2560. https://doi.org/10.1108/IJOA-11-2021-3049

Amundsen, S., & Martinsen, Ø. L. (2014). Empowering leadership: Construct clarification, conceptualization, and validation of a new scale. *The Leadership Quarterly*, *25*(3), 487–511.

Armenakis, A. A., Harris, S. G., & Mossholder, K. W. (1993). Creating readiness for organizational change. Human relations, 46(6), 681-703.

Bouckenooghe, D., Devos, G., & Van Den Broeck, H. (2009). Organizational change questionnaire-climate of change, processes, and readiness: Development of a new instrument. Journal of Psychology: Interdisciplinary and Applied, 143(6), 559–599. https://doi.org/10.1080/00223980903218216

Brahmana, S. S. (2021). Assessing organizational readiness for change of utama's lecturers. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(8), 846-851.

Colquitt, J. A., Lepine, J. A., & Wesson, M. J. (2015). Organizational Behavior: Improving Performance and Commitment (p. 632). New York: McGraw Hill Education.

Colquitt, J. A., Wesson, M. J., & LePine, J. A. (2018). Organizational Behavior: Improving Performance and Commitment in the Workplace. McGraw-Hill Education. https://books.google.co.id/books?id=6SrJswEACAAJ

Dwivedula, R., Bredillet, C. N., & Müller, R. (2016). Personality and work motivation as determinants of project success: the mediating role of organisational and professional commitment. International Journal of Management Development, 1(3), 229-245.

Estradha, R., Eryanto, H., Eliyana, A., Suhud, U., Widyastuti, U., & Wibowo, A. (2025). Determinant Factors for the Readiness of Human Resource Information Systems (HRIS) in Public Organizations. *Interdisciplinary Journal of Management Studies*, *18*(2), 209–223. https://doi.org/10.22059/ijms.2024.360758.675944.

Faulks, B., Song, Y., Waiganjo, M., Obrenovic, B., & Godinic, D. (2021). Impact of empowering leadership, innovative work, and organizational learning readiness on sustainable economic performance: an empirical study of companies in russia during the COVID-19 pandemic. *Sustainability (Switzerland)*, *13*(22), 12465. https://doi.org/10.3390/su132212465.

Gentles-Gibbs, N., & Kim, H. (2019). A Grounded Theory Study of Organizational Readiness for Change in Public Child Welfare: Developing a Theoretical Model. *Child Welfare*, *97*(1), 109.

Guenduez, A. A., & Mergel, I. (2022). The role of dynamic managerial capabilities and organizational readiness in smart city transformation. Cities, 129, 103791.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis*. Cengage. https://books.google.co.id/books?id=0R9ZswEACAAJ.

Hermawan, I., & Suharnomo, S. (2020). Information Technology as a Strategic Resource in Encouraging Organizational Change Readiness through the Role of Human Capital Effectiveness. *Jurnal Dinamika Manajemen*, *11*(2), 242–254. https://doi.org/10.15294/jdm.v11i2.23700.

Holt, D. T., Armenakis, A. A., Feild, H. S., & Harris, S. G. (2007). Readiness for organizational change: The systematic development of a scale. The Journal of applied behavioral science, 43(2), 232-255.

Jung, K. B., Kang, S. W., & Choi, S. B. (2020). Empowering leadership, risk-taking behavior, and employees' commitment to organizational change: The mediated moderating role of task complexity. Sustainability (Switzerland), 12(6), 2340. https://doi.org/10.3390/su12062340

Kim, M., & Beehr, T. A. (2020). Empowering leadership: leading people to be present through affective organizational commitment? The International Journal of Human Resource Management, 31(16), 2017-2044.

Kotter, J. P. (1996). Leading Change. Harvard Business Review Press.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. Educational and psychological measurement, 30(3), 607-610.

Lauer, T. (2020). Change management: Fundamentals and success factors. In *Change Management: Fundamentals and Success Factors*. Springer Nature. https://doi.org/10.1007/978-3-662-62187-5

Lokuge, S., Sedera, D., Grover, V., & Dongming, X. (2019). Organizational readiness for digital innovation: Development and empirical calibration of a construct. *Information and Management*, *56*(3), 445–461. https://doi.org/10.1016/j.im.2018.09.001

Mahendrati, H. A., & Mangundjaya, W. (2020). Individual Readiness for Change and Affective Commitment to Change: The Mediation Effect of Technology Readiness on Public Sector. *3rd Forum in Research, Science, and Technology (FIRST 2019)*, 52–59. https://doi.org/10.2991/assehr.k.200407.010

Masrek, M. N., & Heriyanto, H. (2021). Prosedur Perbaikan untuk Mengontrol dan Meminimalkan Varian Metode Umum dalam Survei Penelitian Ilmu Perpustakaan dan Informasi. *Palimpsest: Jurnal Ilmu Informasi Dan Perpustakaan*, *12*(1), 1–11. https://doi.org/10.20473/pjil.v12i1.25062.

Mathur, M., Kapoor, T., & Swami, S. (2023). Readiness for organizational change: the effects of individual and organizational factors. *Journal of Advances in Management Research*, *20*(4), 730–757. https://doi.org/10.1108/JAMR-02-2023-0032.

Meyer, J. P., & Allen, N. J. (1991). A three-component conceptualization of organizational commitment. Human resource management review, 1(1), 61-89.

Meyer, J. P., & Parfyonova, N. M. (2010). Normative commitment in the workplace: A theoretical analysis and re-conceptualization. Human resource management review, 20(4), 283-294.

Olafsen, A. H., Nilsen, E. R., Smedsrud, S., & Kamaric, D. (2020). Sustainable development through commitment to organizational change: the implications of organizational culture and individual readiness for change. *Journal of Workplace Learning*, *33*(3), 180–196. https://doi.org/10.1108/JWL-05-2020-0093

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879.

Prastiti, Z. D. (2021). The Relationship Between Organizational Commitment and Readiness to Change for Employee of PT X. *Jurnal Organisasi Dan Manajemen*, *17*(1), 78–88. https://doi.org/10.33830/jom.v17i1.1003.2021

Priambodo, I. T., Sasmoko, S., Abdinagoro, S. B., & Bandur, A. (2021). E-Commerce Readiness of Creative Industry During the COVID-19 Pandemic in Indonesia. *Journal of Asian Finance, Economics and Business*, *8*(3), 865–873. https://doi.org/10.13106/jafeb.2021.vol8.no3.0865.

Siti Rachmawati, T., Sukmawat, A., & Nurhayati, P. (2024). Employee Readiness in Implementing the Core Values of the Indonesian Government Organization "BerAKHLAK." *Jurnal Dinamika Manajemen*, *15*(1), 104–124. https://doi.org/10.15294/jdm.v15i1.50062.

Reichers, A. E. (1985). A review and reconceptualization of organizational commitment. *Academy of Management Review. Academy of Management*, *10*(3), 465–476. https://doi.org/10.5465/AMR.1985.4278960

Runa, R. (2023). Determinants of Employee Change Readiness: A Study of Transformational Leadership and Self-Efficacy and the Role of Organizational Commitment. *Almana : Jurnal Manajemen Dan Bisnis*, *7*(1), 154–167. https://doi.org/10.36555/almana.v7i1.2144.

Taherdoost, H. (2019). What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale / Attitude Scale / Likert Scale. *International Journal of Academic Research in Management (IJARM)*, *8*(1), 2296–1747. https://hal.science/hal-02557308

Vaishnavi, V., & Suresh, M. (2020). Modelling of readiness factors for the implementation of Lean Six Sigma in healthcare organizations. *International Journal of Lean Six Sigma*, *11*(4), 597–633. https://doi.org/10.1108/IJLSS-12-2017-0146

Vaishnavi, V., Suresh, M., & Dutta, P. (2019). A study on the influence of factors associated with organizational readiness for change in healthcare organizations using TISM. *Benchmarking*, *26*(4), 1290–1313. https://doi.org/10.1108/BIJ-06-2018-0161

Waseel, A. H., Zhang, J., Shehzad, M. U., Saddiqa, A., Liu, J., & Hussain, S. (2023). Does empowering leadership help firms to establish collaborative culture and organizational commitment to stimulate frugal innovation? *Kybernetes*. https://doi.org/10.1108/K-05-2023-0786.

Weiner J. Bryan. (2009). A theory of organizational readiness for change. In Implementation Science (pp. 1–9). Edward Elgar Publishing.

Weiner, B. J. (2020). A theory of organizational readiness for change. In Handbook on Implementation Science (pp. 215–232). Edward Elgar Publishing. https://doi.org/10.4337/9781788975995.00015

Zarkasi Zarkasi, Syamsu Rijal, Iswahyuningsih Iswahyuningsih, & Dewi Ratih K. (2023). Understanding Information Technology and Accounting Has a Moderated Effect on Organizational Commitment. *GEMILANG: Jurnal Manajemen Dan Akuntansi*, *3*(2), 113–135. https://doi.org/10.56910/gemilang.v3i2.435.

Zhang, Y., Sun, J., Yang, Z., & Wang, Y. (2020). Critical success factors of green innovation: Technology, organization and environment readiness. *Journal of Cleaner Production*, *264*, 121701. https://doi.org/10.1016/j.jclepro.2020.121701.

Journal of Information and Organizational Sciences
Volume 49, Number 2 (2025)
Journal homepage: jios.foi.hr

JIOS

DOI: 10.31341/jios.49.2.3
Open Access

UDC 004.77:004.021:004.85
Original Scientific Paper

The Evolution of Cloud through SJF-ML Hybrid Scheduling

Prathamesh Vijay Lahande[1*]

[1]DES Pune University, Pune, India
*Correspondence: prathamesh.lahande23@gmail.com

PAPER INFO

Paper history:
Received 02 May 2025
Accepted 22 July 2025

Citation:
Lahande, P. V. (2025). The Evolution of Cloud through SJF-ML Hybrid Scheduling. In Journal of Information and Organizational Sciences, vol. 49, no. 2, pp. 193-211

ABSTRACT

Purpose: The author proposes sixteen Shortest Job First - Machine Learning (SJF-ML) hybrid algorithms, combining the cloud's SJF scheduling algorithm with four ML algorithm categories, with cloud evolution through ML intelligence as the primary objective. The four categories include: SJF-CA, SJF-ELA, SJF-PM, and SJF-RA. The developed SJF-ML algorithms by the author perform pattern recognition of the tasks that are to be computed, to improve decision-making during task computations in the cloud. These sixteen SJF-ML algorithms include: SJF-ADAB, SJF-BAY, SJF-DT, SJF-KNN, SJF-LAS, SJF-LDA, SJF-LGB, SJF-LN, SJF-MLP, SJF-NAV, SJF-PLY, SJF-RDG, SJF-RF, SJF-RBST, SJF-SVM, and SJF-XGB. Performance Metrics: Cost, Time, Energy, and LB are utilized to compare the developed algorithms with baseline SJF, along with comparing them within their respective SJF-ML categories. Dataset: The real-time Google Big Data Task (BDT) dataset, comprising tasks ranging from one hundred to one thousand across nineteen files, was computed using the SJF-ML and SJF algorithms. Experiment: Open-source CloudSim simulator with VM counts of 20, 40, 60, 80, and 100 were utilized to compute the BDTs, outputting results across the considered metrics. Results: The algorithms SJF-XGB and SJF-LN provided the best results, with SJF-DT, SJF-LAS, and SJF-LDA providing poor results. Findings: Hybridization of the cloud's scheduling algorithms with ML provides improved intelligence and performance, resulting in the evolution of the cloud.

Keywords: Cloud-Computing, Hybrid-Algorithm, Machine-Learning, Scheduling, SJF

1. Introduction

1.1. Problem Formulation

Cloud development is profound in today's world, where several users are inclined towards using it, rather than using alternative platforms for computing purposes (Chaudhary et al. and Chung et al., 2025). On the one hand, its computing power is impressive; however, the cloud encounters several limitations in handling its resources, where current scheduling methods struggle without any intelligence mechanism to handle Big Data Tasks (BDTs) (Chaudhary et al., 2025). Without any modern intelligence mechanism to deal with these uncertain BDTs, the current scheduling methods of the cloud will always output results whose highest threshold is limited (Kathole et al., 2025). This puts a pause on evolving the cloud, which is contrary to the current modern evolution trend observed in several systems through modern Machine Learning (ML) mechanisms. With the technological advancements and results obtained with ML techniques, the scheduling algorithms of the cloud need to be provided with its intelligence to cope up its performance-related gaps and ensure its performance is elevated, thereby providing an evolution to the cloud as compared with the classical cloud systems (Kathole et al. and Liu, 2025). With the need for BDT computations in the modern day, the cloud faces decision-making challenges too, with its classical scheduling methods, which function without ML intelligence (Sanjalawe et al., Sonia et al., and Ye et al., 2025). Without any integration of ML methods, these current scheduling algorithms fail to withstand these challenging BDTs, leading to higher cost and time consumption (Ali et al. and Almurshed et al., 2024). Additionally, the existing cloud's scheduling methods do not possess a pattern recognition facility which includes detection of the irregularities in BDTs, trends, and classification of incoming BDTs which are to be computed, predicting execution times of BDTs, leading to an improper Load Balancing (LB) mechanism in the cloud, which often increases energy consumption (Alsubaei et al., 2024; Bartakke et al., 2024; and Hayyolalam et al., 2024). Hence, the author has focused on these issues to hybridize with the cloud's ideal performing Shortest Job First (SJF) scheduling algorithm with the vast range of ML category of algorithms to provide an improved version of scheduling policies. In this paper, the author presents sixteen SJF-ML scheduling methods across four ML categories by using the preemptive SJF scheduling mechanism of selecting the BDT with the least computing time among the rest and using ML intelligence to allocate the cloud's resources to compute the BDTs. Here, the hybridization process includes combining the approaches of SJF scheduling with the ML technique to improve the scheduling approach. These SJF-ML algorithms give the cloud a fair chance to make better decisions with the BDTs, output better results, and ultimately lead to its evolution through a systematic and balanced BDT computation. Lastly, the author has also proposed a second-level hybridization in this research paper by combining the best-performing algorithms in the first level from each SJF-ML category to form a second-level hybridization to further improve cloud performance as a part of future research direction.

Figure 1 shows the Venn diagram of combining SJF with ML intelligence to develop the SJF-ML algorithms for improving its performance and providing a next-level cloud-evolved computing environment.

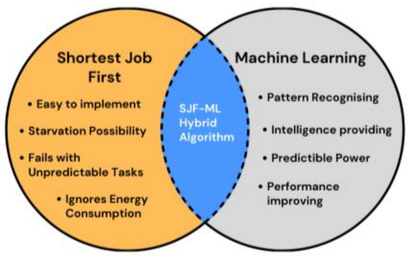

Figure 1. Individual characteristics SJF and ML algorithms for Cloud Evolution.

Figure 1 shows that the individual characteristics provided by the SJF scheduling algorithm are combined with those of ML algorithms to provide a hybridized stronger approach for better scheduling results.

1.2. Proposed Solution

The author proposes an experimental-based research work that includes hybridizing the best-performing cloud scheduling algorithm, Shortest Job First (SJF) in its preemptive form, with ML categories of algorithms to develop algorithms across four different ML algorithm categories, termed SJF-ML, to ensure higher performance is obtained and the cloud evolves through ML intelligence. Through this evolution, the proposed SJF-ML algorithms by the author ensure that a better Quality of Service (QoS) is provided to end users.

The author presents the following SJF-ML algorithm categories:

- **Shortest Job First - Classification Algorithm (SJF-CA) Category**: The classical SJF is combined with ML's Classification Algorithms (CA) to develop the SJF-CA category of algorithms.
- **Shortest Job First - Ensemble Learning Algorithm (SJF-ELA) Category**: The classical SJF is combined with ML's Ensemble Learning Algorithms (ELA) to develop the SJF-ELA category of algorithms.
- **Shortest Job First - Probabilistic Model (SJF-PM) Category**: The classical SJF is combined with ML's Probabilistic Model (PM) Algorithms to develop the SJF-PM category of algorithms.
- **Shortest Job First - Regression Algorithm (SJF-RA) Category**: The classical SJF is combined with ML's Regression Algorithms (RA) to develop the SJF-RA category of algorithms.

The author has explored the above four SJF-ML categories to combine the algorithms from each ML category with SJF in preemptive mode, which are suitable for scheduling mechanisms in the cloud environment in their own unique way or working, leading to the development of sixteen SJF-ML hybrid algorithms as follows:

- **SJF-CA Algorithms:** The SJF is combined with four ML CAs algorithms to define the SJF-CA category. These include: Decision Tree (DT), which is ideal for rule-based VM allocation decisions at run-time; K-Nearest-Neighbors (KNN), which adapts to the dynamic workload of the BDTs using grouping techniques; Linear Discriminant Analysis (LDA), which is optimal for task prioritization with reduced dimensionality for the BDTs; and Support Vector Machine (SVM), which is effective for handling the non-linearity of the BDTs distributions (Talwani, S. et al. 2022 and Mo, Y. et al. 2015). These algorithms are hybridized with SJF to develop SJF-DT, SJF-KNN, SJF-LDA, and SJF-SVM hybrid SJF-CA scheduling algorithms, respectively.
- **SJF-ELA Algorithms:** The SJF is combined with four ML ELAs algorithms to define the SJF-ELA category. These include: AdaBoost (AB), which is effective in reducing the biasness of BDT computing time predictions; Light Gradient Boosting Machine (LGB), which has a high-speed memory efficient scheduling for the BDTs; Random Forrest (RF), which is useful with volatile workloads; and Extreme Gradient Boosting (XGB), which is highly useful for skewed BDT distributions (Liu, Z. 2025). These algorithms are hybridized with SJF to develop SJF-AB, SJF-LGB, SJF-RF, and SJF-XGB hybrid SJF-ELA scheduling algorithms, respectively.
- **SJF-PM Algorithms:** The SJF is combined with two PMs, including Bayesian Network (BAY), which uses probabilistic models to handle dependencies in the BDT scheduling process, and Naïve Bayes (NAV), which is lightweight in its nature and uses real-time probability for scheduling short-term BDTs (Chauhan, N. et al. 2022). These algorithms are combined with SJF to develop SJF-BAY and SJF-NAV hybrid SJF-PM scheduling algorithms, respectively.
- **SJF-RA Algorithms:** The six RAs, including Lasso (LAS) which is helpful in the high-dimensional and challenging BDTS, Linear (LN) which provides ideal real-time prediction for short BDTs, Multi-Layer Perceptron Neural Network (MLP) which captures the non-linearities useful for LB, Polynomial (PLY) which models the learning mechanism through its curvilinear BDT trends, Ridge (RDG) which takes care of multi-collinear VM features, and Robust (RBST) which ensures that no outlier BDTs are observed (Huymajer, M. et al. 2024). These algorithms are combined with SJF to develop SJF-LAS, SJF-LN, SJF-MLP, SJF-PLY, SJF-RDG, and SJF-RBST, respectively under the SJF-RA category.

Figure 2 represents the SJF-ML algorithms across the four categories.

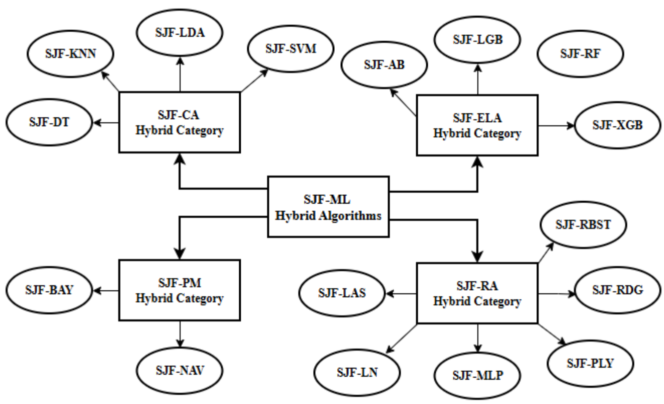

Figure 2. Proposed SJF-ML Algorithms.

2. Literature Review

2.1. Survey of Existing Work

This section includes the LR, where several authors have explored and provided various hybrid techniques using the SJF with other intelligent and prominent methods to focus on cloud limitations and its challenges. The authors of this paper have proposed an AI-enhanced framework that integrates deep workload prediction and RL with SJF and traditional queueing theory to optimize results in the cloud environment (Chaudhary et al., 2025). This study presents a lightweight sampling fine-grained GPU scheduling method that uses suspend-resume mechanisms to reduce energy wastage in cloud systems (Chung et al., 2025). This work includes a detailed review of scheduling resources, which consists of a thorough explanation of scheduling techniques from the past decades and offering gaps from current research (Kathole et al., 2025). The ML's EL algorithms are overviewed to provide a clear pathway for their practical implementation in the cloud (Liu, 2025). This paper includes a review based on AI-based hybrid models, where the work analyses the need for energy-efficient systems for scheduling jobs along with providing future directions (Sanjalawe et al., 2025). The LB techniques are discussed in this research paper, where their significance in ML implementations is provided for multiple computing platforms (Sonia et al., 2025). The author of this research paper has focused on minimizing cloud performance parameters, such as cost and energy, by considering the sustainability of LB in the cloud (Verma, 2025). The Asynchronous Control-based Aggregation Transport Protocol (AC-ATP) rule is provided for reducing the deadlines in scheduling techniques by reducing the time performance metrics (Ye et al., 2025). A multiple-level fuzzy approach is integrated into the Internet of Things (IoT) platform to improve the operational efficiencies experienced by scheduling jobs in the healthcare domain (Ali et al., 2024). The authors have provided a framework named Enhanced Optimized-Greedy Nominator Heuristic (EO-GNH) for AI placement in computing sectors to improve time and resource allocation (Almurshed et al., 2024). The K-means technique is hybridized with Heterogeneous Earliest End Time (HEFT), providing a dual ML framework for task scheduling and improvements in results (Alsubaei et al., 2024). Zero-trust hybrid strategies have been provided by the authors to provide a structured evaluation framework for the cloud environment for selecting the best solutions and offering guidance on cost-effectiveness and implementation (Bartakke et al., 2024). The authors have presented a Chaos theoretical model hybridized with the Black Widow Optimization algorithm (CBWO) to focus on LB issues in cloud environments (Hayyolalam et al., 2024). A priority-based method in the cloud using DRL named PH-DRL is presented for better scheduling of jobs and improved time metrics (He et al., 2024). The authors have presented a scheduling technique for a fog-computing environment to improve QoS (Hosseini et al., 2024). A comparative evaluation of RAs is performed where seventeen RAs are studied in contrast with conventional methods (Huymajer et al., 2024). ML techniques focused on cloud optimization for better decision-making (Kanchetti et al., 2024).

This research work hybridized SJF with priority and deployed the Enhanced Shortest Job First with Priority (ESJFP) algorithm to improve classical SJF with task complexities having ML capabilities and improve resource and time metrics (Laha et al., 2024). A multi-objective flexible job-shop scheduling problem (MOFJSP) scheduler is presented in this paper to imbibe graphs with modern prominent methods for complex problem solutions (Li et al., 2024). A Graph-based Denoising Diffusion Probabilistic Model (G-DDPM) is presented for time-series forecasting, where the method works with PMs to reduce energy metrics (Miraki et al., 2024). The LB issues are handled using the presented Modified Parallel Particle Swarm Optimization (MPPSO) method for task scheduling, reducing the time parameters in the cloud (Pradhan et al., 2024). The author has implemented ML algorithms like DT, RF, and KNN to provide solutions related to posture classification, achieving better accuracy and Interpretability (Rahimi et al., 2024). The SJF is hybridized with a Multi-Level Memory-Based Framework (SJF-MMFB) for enhanced cloud scheduling by incorporating multi-queues addressing starvation Issues and improving fairness in cloud systems (Rekha et al., 2024). Hybrid ML models are implemented for cryptocurrency predictions and handling real-time predictions with achieved profitability (Salehi et al., 2024). Big data Integrated cloud systems are used to improve decision-making and operational performance for the ERP systems (Saraswat et al., 2024). The challenges associated with the cloud are focused on ML techniques to improve LB, the system's efficiency, and decision-making (Zende et al., 2024). The authors have proposed a dual-objective algorithm to improve results concerning the scheduling domain (Asghari et al., 2023). Post validations across twelve scenarios, the presented algorithm provides better results when compared to the other state-of-the-art algorithms. The author has used the SVM scheme in the two-cloud platform to reduce computing and communication overheads (Hu et al., 2023). Modern statistical models are utilized to investigate the cloud performance in this work (Kumar et al., 2024). This work surveys task scheduling methods and compares them with various performance parameters for optimal task management (Nayak et al., 2023). This work demonstrates the use of ML to improve real-time scheduling algorithms for cluster environments, optimizing task scheduling for dynamic environments (Zhang et al., 2023). This work explores how ML impacts energy optimization, dynamic load balancing, task scheduling, and security in cloud computing, improving resource allocation, security, and VM migration (Kumar et al., 2022). The EAs are utilized for the prediction of diseases in cocoa crops, offering early and accurate predictions based on climatic parameters, and benefiting farmers with proactive disease management (Olofintuyi et al., 2022).

This work evaluates hybrid scheduling and allocation algorithms in cloud computing, showing that combining SJF with RR results in the best response time, while SJF with a novel length-wise allocation (LwA) provides the least CPU utilization (Sahkhar et al., 2022). The author has combined two optimization methods to strengthen the prominent cloud performance parameters and enhance the task scheduling process in the cloud (Verma, 2022). The authors have examined the use of RR and SJF algorithms for improving efficiency in cloud environments, suggesting an approach for optimal scheduling and resource allocation (Kumar et al., 2021). A detailed comparison is made with SJF and Longest Job First (LJF) with ML-based scheduling approaches using CloudSim, showing SJF's efficiency while highlighting ML's potential for further optimization in dynamic environments (Murad et al., 2021). This work reviews studies on ML in cloud security, identifying SVM as key and emphasizing hybrid methods for threat detection using metrics like True Positive Rate (Nassif et al., 2021). The authors propose hybrid SJF-Min-Min Best Fit (MMBF) and SJF-Extreme Learning Machine (ELM) scheduling models using Particle Swarm Optimization (PSO), optimizing job hosting and minimizing starvation in dynamic cloud settings (Rekha et al., 2021). This work combines SJF-MMBF with ELM energy-aware scheduling, addressing energy and security challenges (S Rekha et al., 2021).

This paper introduces a scheduling method for enhancing the task allocation process in cloud environments (Tanha, M. et al., 2021). The authors have designed an Enhanced SJF (ESJF) for stable task management by introducing ESJF with time-slicing between shortest and longest jobs, outperforming SJF in unstable environments by reducing starvation and delays (Younis et al., 2021). This work enhances VM scheduling using SJF-MMBF and SJF-ELM models with PSO, improving performance in high-load, dynamic cloud environments (Rekha et al., 2020). This work uses DRL and Long Short-Term Memory (LSTM) for scheduling, outperforming SJF, RR, and PSO by reducing CPU/RAM usage and task delays on real-world cloud workloads (Rjoub et al., 2020). A SJF-MMBF method is applied for efficient VM scheduling, combining SJF and MMBF with the Queue-Length MaxWeight policy to prevent starvation and improve job throughput (Guo et al., 2019). This work includes DRL used with LSTM to reduce CPU and RAM usage, outperforming SJF, RR, and PSO in big data task scheduling (Rjoub et al., 2019). CAs are discussed in this paper to ensure that better cloud system performance is achieved (Samie et al., 2019). A Q-learning-based method integrated with the heterogeneous earliest finish time (HEFT) is deployed to reduce makespan and response time, outperforming traditional scheduling algorithms in efficiency (Tong et al., 2019). This work compares heuristic SJF-MMBF with reinforcement-learning-based SJF-RL, showing SJF-RL excels in minimizing job delay and preventing starvation (Guo et al., 2018). This work includes an SRDQ algorithm combining SJF and RR with an adjustable time quantum to reduce time and avoid long task starvation (Elmougy et al., 2017).

The above-mentioned contributions from the researchers all around the globe who have presented their hybrid methodologies involving SJF. This work helped the author to understand how the classical and traditional SJF scheduling algorithms have been used with the modern ML methods. The next sub-section of the LR dives deeper into understanding the contributions of these authors and how they helped the author of this paper to contribute to improving cloud performance through hybridizations.

2.2. LR concerning Technical Overview

This sub-section includes the LR concerning Technical Overview, where Important aspects of author contributions are studied and represented in terms of the algorithm presented, its features, along with the enhancements it offers and limitations it possesses, presented in detail in Table 1.

Ref. No.	Algorithm	Features	Enhancements	Limitations
[1]	AI-enhanced framework	Dual-layer neural network	Adaptive scheduling	High training time, overhead
[2]	Lightweight sampling	Job migration, sampling	Boosts efficiency, fairness	Potential migration overhead
[3]	Meta-heuristic algorithms	Categorization, review	Identifies research gaps	Lacks novel implementation
[4]	Ensemble Learning	Combines multiple learners	Increases robustness, accuracy	Computationally intensive
[5]	AI-based hybrid models	AI decision-making cloud	Improves energy, fault handling	Complex adaptation to real-time
[6]	LB techniques	Real-time workload adaptation	Optimizes resource distribution	Hybrid LB in cloud
[9]	AC-ATP protocol	Aggregation and congestion	Reduces traffic, accelerates training	Straggler delays

[10]	Fuzzy Approach	IoT integration	Enhances healthcare predictions	Relies on expert labelling
[11]	EO-GNH	Optimized placement, parallelism	Boosts edge-cloud performance	Scaling complexity
[12]	K-means, HEFT	Task priority scheduling	Improves LB	Tuning required for K-values
[13]	Zero-Trust Framework	Identity/context verification	Fine-grained access, real-time policies	Legacy system updates needed
[14]	CBWO	Chaos theory, task distribution	Reduces energy usage	Needs parameter tuning
[15]	PH-DRL	DRL-based hybrid scheduler	Balances performance	Limited generalizability
[17]	Regression Models	Comparative feature study	Enhances prediction, tuning strategies	Risk of overfitting
[18]	General ML methods	Data retrieval, automation	Boosts decision-making, scalability	Infrastructure reliability
[19]	ESJFP	Prioritizes tasks	Reduces wait time	Lacks real-time adaptability
[20]	MO-GARL	Combines GAT with RL	Generalizable solutions obtained	Compute-heavy training
[21]	G-DDPM	Graph learning	Enhances accuracy	Resource-intensive
[22]	MPPSO	PSO for scheduling	Reduces delays, boosts throughput	Large-scale setting complexity
[23]	DT, RF, KNN	Time-frequency	Transparent feature influence	Inconsistent accuracy
[24]	SJF-MMFB with PSO	Integrates PSO, SJF	Balances load, prevents starvation	Overhead in large environments
[25]	ML methods	Applied to crypto price prediction	Combines fuzzy logic and ML	Some models over-fit
[26]	Big data with Cloud	Integration of big data in ERPs	Cost-effective	Lack of real-world case studies
[27]	ML methods	ML for LB	Improves cloud efficiency, security	Resource-demanding
[29]	SVM	Privacy-preserving ML	Resilience to malicious behaviour	Encryption complexity
[30]	Statistical Models	Survey of cloud apps	Insight into cloud app preferences	No qualitative analysis
[31]	Scheduling methods	Scheduling Methods Analysis	Optimize scheduling	Context-sensitive performance
[32]	FCFS, SJF	Predictive ML	Accurate burst time forecasts	Limited to real-time constraints
[33]	ML methods	Predictive ML	AI improves efficiency, and automation	Human expertise is still needed
[34]	Ensemble Learning	Time-series	High-accuracy forecasting	Region-specific data limitations
[35]	SJF, FCFS, RR, LwA	Various scheduling algorithms	Best response time with SJF+RR	Neglects other metrics
[36]	RR, SJF	Pre-emptive RR, SJF	Minimizes waiting time	High waiting time
[37]	SJF, LJF	SJF, LJF for Scheduling	Increases throughput	Starvation observed
[38]	SVM	Cloud security threats	Improve accuracy	High training time
[39]	PSO, SJF, MMBF, ELM	Hybrid queue scheduling	Prevents starvation	Scalability issues
[40]	SJF, MMBF, ELM	Hybrid scheduling	Improves efficiency	Complex hybrid approach

[42]	ESJF	Time-slice strategy	Reduces waiting times	Increased scheduling complexity
[43]	PSO, SJF, MMBF, ELM	Prevents starvation	Optimized scheduling performance	Computational overhead
[44]	RL, DQN, RNN-LSTM	LSTM	Reduces CPU/RAM usage	High computational expense
[45]	MMBF, SJF	SJF with QMW Scheduling	Improves fairness	Complexity in resource scheduling
[46]	DRL, LSTM	Predict VM assignment for tasks	Reduces resource consumption	Complex implementation
[47]	Classification Techniques	ML techniques for IoT	Improves data management	Scalability concerns
[48]	QL-HEFT	Combines Q-learning and HEFT	Improves makespan	High computational overhead
[49]	SJF-MMBF	Hybrid SJF with RL	Adapts to dynamic workloads	Adds computational complexity
[50]	SRDQ	Hybrid SJF and RR	Balances SJF and RR	Complexity in fine-tuning

Table 1. Literature Review concerning Technical Overview.

The conducted LR helped the author to identify the gaps and place the foundations for this research.

2.3. Gaps

The following gaps were identified from the above conducted LR:

- Existing studies have not fully developed hybrid scheduling methods that integrate a diverse range of ML algorithms beyond currently explored techniques, restricting cloud evolution.
- Current approaches lack the integration of BDT computations within ML-enhanced SJF, despite their need in the modern computing world.
- A comprehensive comparison between SJF and its improved versions is missing, particularly in assessing the entire spectrum of cloud performance metrics.
- Existing methods lead to increased energy consumption and expenses despite improved results.
- Contributing to a second level of hybridization using the best obtained multiple algorithms is absent in the same research contribution.

The identified gaps helped the author to develop the SJF-ML algorithms and make suitable contributions for evolving the cloud, further represented in section 3.

3. Contributions

The following includes the author's contributions from this research work:

- The SJF method has been enhanced through hybridization with four distinct and diverse categories of ML algorithms, resulting in the development of sixteen SJF-ML algorithms.
- Real-time BDTs were employed to evaluate the performance of the developed SJF-ML algorithms with baseline SJF, ensuring their applicability to evolve cloud environments.
- A comprehensive analysis was conducted to compare the SJF with hybridized SJF-ML methods, considering a broad spectrum of cloud performance metrics, including cost, time, energy, and LB.
- From the best-performing algorithm in each ML category, an architecture consisting of a second-level hybridization approach is provided to further refine scheduling efficiency and contrast its effectiveness against other SJF-ML variants.
- The proposed SJF-ML methodologies aim to enhance cloud performance, improve QoS, and contribute to more efficient, scalable scheduling solutions in high-demand computing environments.

The above contributions are made through this unbiased research work through extensive experiments in an open-source simulator where several real-time cloud components are deployed and BDTs are computed using all the developed SJF-ML algorithms.

4. Dataset, Experimental Setup and Performance Metrics

4.1. Dataset

Google's BDT was utilized by the classical SJF method along with the developed sixteen SJF-ML hybrid algorithms across four categories for computing. This dataset consists of nineteen input files across task lengths from one hundred tasks to one thousand tasks, each task having AT and CT as the significant features for the BDT. These features are fed to the SJF-ML algorithms to improve the cloud's scheduling decisions. The author has provided data analysis for the input BDTs in the form of descriptive statistics for these BDTs, including Mean (μ), Median (MDN), and Standard Deviation (σ) for a better understanding of the BDTs used. Table 2 includes the descriptive statistics for the BDTs.

Descriptive Statistics		**BDT-AT (s)**			**BGT-CT (s)**		
Sr. No.	**BDT Length**	**μ**	**MDN**	**σ**	**μ**	**MDN**	**σ**
1	100	543.1000	549.5000	273.6093	135.10	91.00	167.42
2	150	557.8733	587.0000	253.7945	147.87	97.00	185.55
3	200	573.3700	585.5000	275.2022	136.10	87.00	178.60
4	250	536.9560	537.5000	254.4490	119.91	89.00	152.95
5	300	545.1067	552.0000	265.7443	138.42	89.00	182.21
6	350	551.8571	527.0000	262.9433	135.05	95.00	162.92
7	400	552.8725	564.5000	263.0676	129.67	95.00	150.59
8	450	549.8956	560.5000	261.5785	123.83	91.00	151.95
9	500	556.3660	560.5000	258.0876	130.00	93.00	153.38
10	550	538.8127	527.5000	272.4410	130.29	93.00	168.00
11	600	542.4350	538.5000	258.2551	137.24	93.00	175.03
12	650	544.5800	549.5000	257.2584	136.16	93.00	168.88
13	700	550.8814	550.0000	255.6384	124.99	93.00	150.30
14	750	551.6360	545.5000	251.5622	136.07	93.00	171.25
15	800	560.9663	567.0000	256.9907	123.70	91.00	149.87
16	850	547.2741	537.5000	261.1188	125.12	91.00	149.47
17	900	542.2711	538.0000	256.2634	133.39	93.00	161.45
18	950	543.3053	544.5000	260.4067	124.61	94.00	150.46
19	1000	552.9770	552.5000	248.5605	129.66	93.00	162.63
	AVG	**549.6072**	**551.2895**	**260.3669**	**131.4305**	**92.3158**	**162.7847**

Table 2. Descriptive Statistics for the BDTs.

Table 2 shows that the descriptive statistics of the ATs possess a considerable amount of variability, making it difficult to predict the SJF and develop SJF-ML algorithms to be computed. On the other hand, the descriptive statistics of the BDA-CTs showcase a slight amount of stability with lower fluctuations, providing a fair chance to all the algorithms for computing. At a higher task length, both times showcase balanced behavior, suggesting there will be a steady execution with higher loads. These tasks were input into a real-time computing environment to the SJF and SJF-ML algorithms across several VM-based experimental scenarios to ensure unbiased results are obtained.

4.2. Working of SJF-ML hybrid algorithms

The SJF and ML algorithms integrate with each other to make the crucial scheduling decisions at runtime. The SJF-ML scheduling algorithms executed the BDTs based on the feature of BDT-CTs. The working mechanism includes loading the challenging BDTs in the cloud queue, followed by the feature extraction with the label defined by the BDT-CTs. Further, the BDTs are split into 80 % training and 20 % testing sets and fed to each of the presented sixteen SJF-ML algorithms for training and testing in all the experimental scenarios. Later, the BDTs are sorted considering CTs and scheduled with the ideal available VM at runtime. Figure 3 represents the general block diagram of the working of all the SJF-ML algorithms.

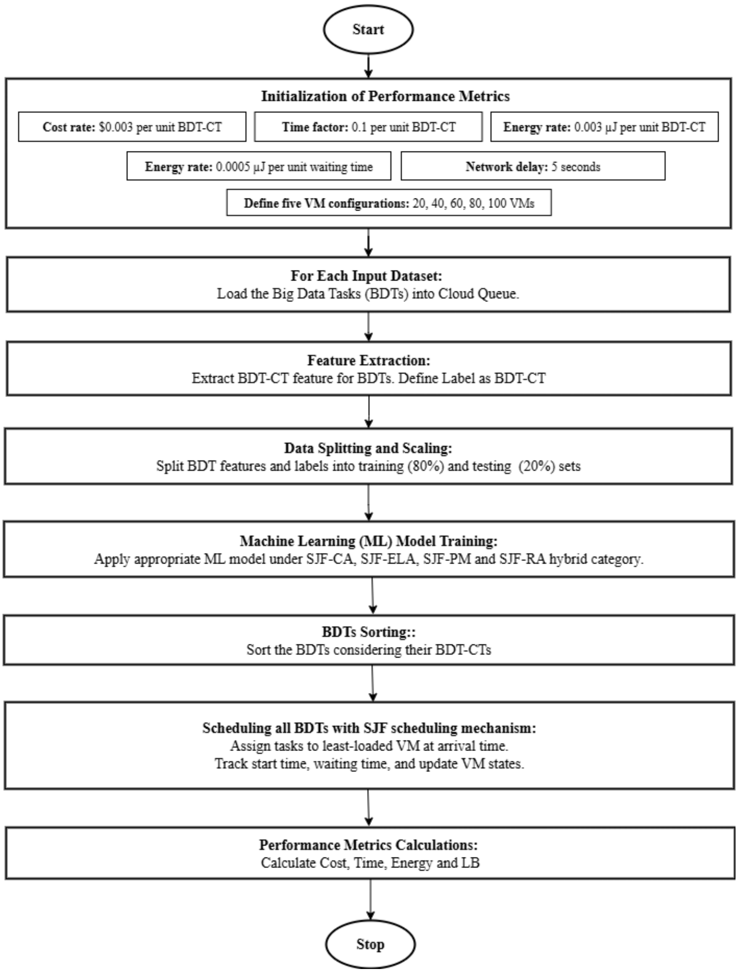

Figure 3. Block diagram of Working of the SJF-ML Algorithms.

5. Experimental Setup and Performance Metrics, and Architecture of Experiment

5.1. Experiment Setup and Performance Metrics

The experiment was set up in the open-source CloudSim environment where the classical SJF and sixteen developed SJF-ML hybrid algorithms across four SJF-ML categories were Incorporated In this simulation environment. The experiment simulates computing the BDTs in a multiple-VM environment. The performance metrics considered for the study are: cost ($), time (seconds 's'), energy (microJules 'µJ'), and LB. LB here is defined as a measurement that uses the coefficient of variation, considering workloads across all the VMs post-scheduling all the tasks, considering a certain scenario. All the SJF-ML algorithms provide dynamic LB, unlike the existing SJF, through enhancements of predictive intelligence mechanisms. The ML component during the scheduling process allocates VMs by considering their queue lengths, resource utilization patterns, and estimation of BDT-CTs, which is significant for SJF. Through continuous real-time feedback against fluctuating workloads, the SJF-ML ensures a stable distribution of all BDTs evenly among the considered VMs. The computing cost rate was set to \$0.003 per unit BDT-CT. The Execution Time Factor was set to 0.1 per unit BDT-CT. The Energy Consumption Rate and Idle Energy Consumption Rate were set to 0.003μJ per unit BDT-CT and 0.0005μJ per unit waiting time. With the possibility of errors and faults in the network, a constant network delay of 5s is considered. The experiment was conducted in five scenarios with the VM count as: 1: 20 VMs; 2:40 VMs; 3:60 VMs; 4:80 VMs, and 5:100 VMs. The main reason to choose these four metrics is that they cover the entire spectrum of parameters for the cloud. Time and Cost play a major role in scheduling in the cloud and are critical for judging and comparing algorithms. Energy ensures the cloud's resources are consumed optimally without having to sacrifice time and cost to do so. Lastly, each cloud resource needs to be utilized to the fullest, and an uneven workload allotment ensures high time, cost, and energy for the ones most used. Hence, LB as a metric is calculated in the form of a coefficient of variance which provides a balance to ensure appropriate energy is consumed with Ideal cost and time metrics. Together, these metrics form an Ideal comparative framework for the classical SJF to be compared with the developed SJF-ML hybrid algorithms. If a certain algorithm provides ideal results across all four metrics, then other metrics also significant to the cloud will be optimal.

5.2. Architecture of the Experiment

This sub-section provides the experiment's architecture with the deployed SJF-ML hybrid scheduling algorithms. The experiment was conducted where the BDTs were submitted for computing from the user environment to the cloud environment. Upon reaching the cloud's platform, the BDTs are added to the cloud queue, where they wait until their ATs are reached. This queue represents a storage from which the SJF-ML algorithm will choose the BDT for execution. This BDT's size acts like a main feature for the SJF-ML algorithms to select the tasks from the queue, considering its least CT. Here, the developed SJF-ML algorithms deployed will ensure appropriate pattern recognition through eradicating irregularities of the BDTs, classifying them according to their trends, thereby provide an intelligence mechanism before submitting the BDTs to the cloud VMs. From its initial submission until getting selected and assigned to being computed and finally getting completed and closed, a BDT undergoes several stages in its lifetime. Post all completion of BDT computations, the results are computed and provided to the user for the considered metrics. Figure 4 represents the architecture of the experiment with the deployment of SJF-ML hybrid scheduling algorithms.

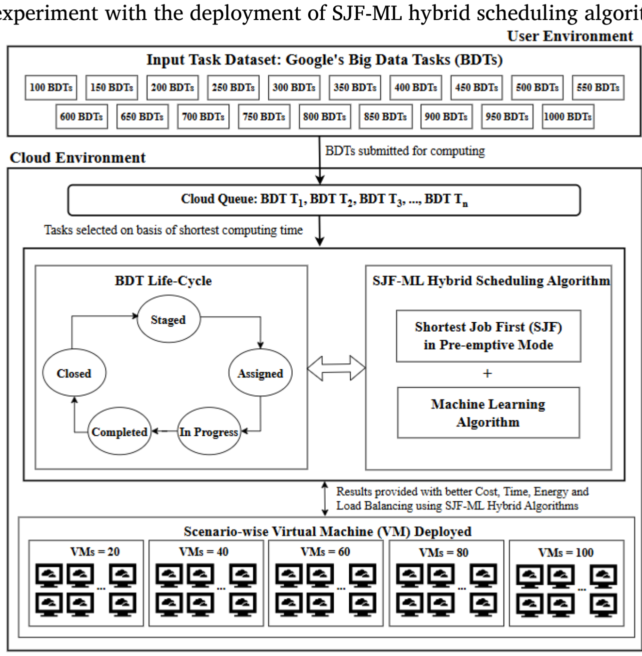

Figure 4. Architecture of the Experiment with SJF-ML Algorithms.

6. Results and Discussions

This section includes the results and discussions for the experiment conducted, presented in four sub-sections: 6.1, 6.2, 6.3, and 6.4, including cost, time, energy, and LB results compared in terms of average (AVG) and improvement percentage (IMPR %) compared against SJF AVG values. The performance rank of a certain algorithm is represented with a circled number, representing its performance rank within its respective SJF-ML hybrid category. Additionally, a green-circled asterisk is represented for the best performance, considering all SJF-ML algorithms and the SJF algorithm.

6.1. Cost Results

Table 3 includes the Cost Results for the developed SJF-ML hybrid algorithms compared to SJF.

Cost ($) Results		VM Counts					AVG	IMPR% vs SJF
Category	Algorithm	20	40	60	80	100		
Classical	SJF	3138.54	840.47	399.99	240.77	164.47	956.85	-
SJF-CAs vs SJF	SJF-DT④	5188.89	1280.00	569.58	328.71	212.67	1515.97	-58.4334
	SJF-KNN③	5028.80	1249.88	557.97	318.10	207.78	1472.51	-53.8914
	SJF-LDA①	3704.52	977.66	458.97	272.85	184.36	1119.67	-17.0163
	SJF-SVM②	3754.84	972.62	447.66	260.50	172.02	1121.53	-17.2106
SJF-ELAs vs SJF	SJF-AB④	4952.65	1236.87	557.00	318.68	209.15	1454.87	-52.0479
	SJF-LGB③	3704.52	977.66	458.97	272.85	184.36	1119.67	-17.0163
	SJF-RF②	3149.31	843.07	401.09	241.36	164.83	959.93	-0.3219
	SJF-XGB①	3138.55	840.46	399.99	240.77	164.47	956.85	0.0000
SJF-PMs vs SJF	SJF-BAY①	4532.35	1116.82	522.63	292.29	195.22	1331.86	-39.1921
	SJF-NAV②	4980.60	1240.91	555.99	317.71	207.68	1460.58	-52.6446
SJF-RAs vs SJF	SJF-LAS⑤	5069.30	1285.20	577.98	329.59	214.76	1495.37	-56.2805
	SJF-LN①⊛	3138.52	840.46	399.99	240.77	164.47	956.84	0.0010
	SJF-MLP④	3150.74	843.71	402.86	241.69	165.11	960.82	-0.4149
	SJF-PLY②	3138.56	840.47	399.99	240.77	164.47	956.85	0.0000
	SJF-RDG③	3142.46	841.42	400.39	240.99	164.60	957.97	-0.1171
	SJF-ROB②	3138.55	840.46	399.99	240.77	164.47	956.85	0.0000

Table 3. Cost Results for SJF-ML hybrid algorithms compared to SJF algorithm.

Table 3 shows the following cost comparison of the SJF-ML hybrid algorithms compared to the SJF algorithm:

- **SJF-CA vs SJF**: SJF > SJF-LDA > SJF-SVM > SJF-KNN > SJF-DT
- **SJF-ELA vs SJF**: [SJF ≈ SJF-XGB] > SJF-RF > SJF-LGB > SJF-AB
- **SJF-PM vs SJF**: SJF > SJF-BAY > SJF-NAV
- **SJF-RA vs SJF**: SJF-LN⊛ > SJF > [SJF-PLY ≈ SJF-ROB] > SJF-RDG > SJF-MLP > SJF-LAS

Figure 5 shows Cost IMPR % graph for SJF-ML Hybrid Algorithms compared to SJF Algorithm.

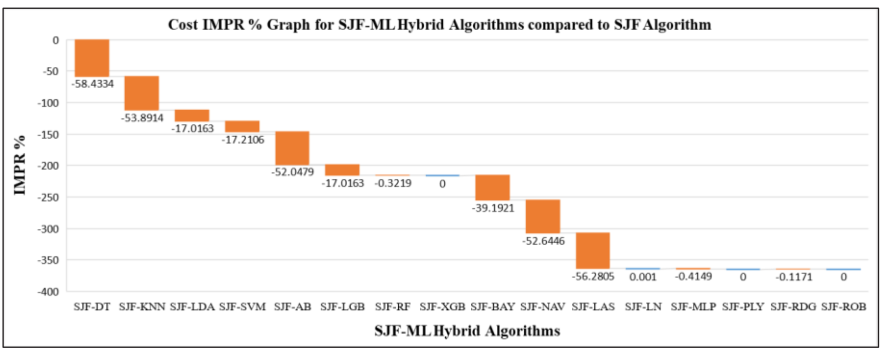

Figure 5. Cost IMPR % graph for SJF-ML Hybrid Algorithms compared to SJF Algorithm.

The cost results convey the SJF-LDA of the SJF-CA category, SJF-XGB of the SJF-ELA category, SJF-BAY of the SJF-PM category, and SJF-LN of the SJF-RA provide the best performance within their respective categories. Additionally, the SJF-LN hybrid algorithm of the SJF-RA category provides the best cost results across all the SJF-ML developed hybrid algorithms, along with the SJF algorithm.

6.2. Time Results

Table 4 includes the time comparison of SJF with presented SJF-ML Hybrid Algorithms.

Time (s) Results		**VM Counts**					**AVG**	**IMPR% vs SJF**
Category	Algorithm	20	40	60	80	100		
Classical	SJF⊛	104.62	28.02	13.33	8.03	5.48	31.90	-
SJF-CAs vs SJF	SJF-DT④	172.96	42.67	18.99	10.96	7.09	50.53	-36.8692
	SJF-KNN③	167.63	41.66	18.60	10.60	6.93	49.08	-35.0041
	SJF-LDA①	123.48	32.59	15.30	9.09	6.15	37.32	-14.5230
	SJF-SVM②	125.16	32.42	14.92	8.68	5.73	37.38	-14.6602
SJF-ELAs vs SJF	SJF-AB④	165.09	41.23	18.57	10.62	6.97	48.50	-34.2268
	SJF-LGB③	123.48	32.59	15.30	9.09	6.15	37.32	-14.5230
	SJF-RF②	104.98	28.10	13.37	8.05	5.49	32.00	-0.3125
	SJF-XGB①⊛	104.62	28.02	13.33	8.03	5.48	31.90	0.0000
SJF-PMs vs SJF	SJF-BAY①	166.02	41.36	18.53	10.59	6.92	48.68	-34.47
	SJF-NAV②	182.62	45.5	19.64	11.54	7.47	53.35	-40.2062
SJF-RAs vs SJF	SJF-LAS④	168.98	42.84	19.27	10.99	7.16	49.85	-36.0080
	SJF-LN①⊛	104.62	28.02	13.33	8.03	5.48	31.90	0.0000
	SJF-MLP③	105.02	28.12	13.43	8.06	5.50	32.03	-0.4059
	SJF-PLY①⊛	104.62	28.02	13.33	8.03	5.48	31.90	0.0000
	SJF-RDG②	104.75	28.05	13.35	8.03	5.49	31.93	-0.0940
	SJF-RBST①⊛	104.62	28.02	13.33	8.03	5.48	31.90	0.0000

Table 4. Time comparison of SJF with presented SJF-ML Hybrid Algorithms.

Table 4 shows the following time comparison of the SJF-ML hybrid algorithms compared to the SJF algorithm:

- **SJF-CA vs SJF**: SJF⊛ > SJF-LDA > SJF-SVM > SJF-KNN > SJF-DT
- **SJF-ELA vs SJF**: [SJF⊛ ≈ SJF-XGB⊛] > SJF-RF > SJF-LGB > SJF-AD
- **SJF-PM vs SJF**: SJF⊛ > SJF-BAY > SJF-NAV
- **SJF-RA vs SJF**: [SJF⊛ ≈ SJF-LN⊛ ≈ SJF-PLY⊛ ≈ SJF-RBST⊛] > SJF-RDG > SJF-MLP > SJF-LAS

Figure 6 shows Time IMPR % graph for SJF-ML Hybrid Algorithms compared to SJF Algorithm.

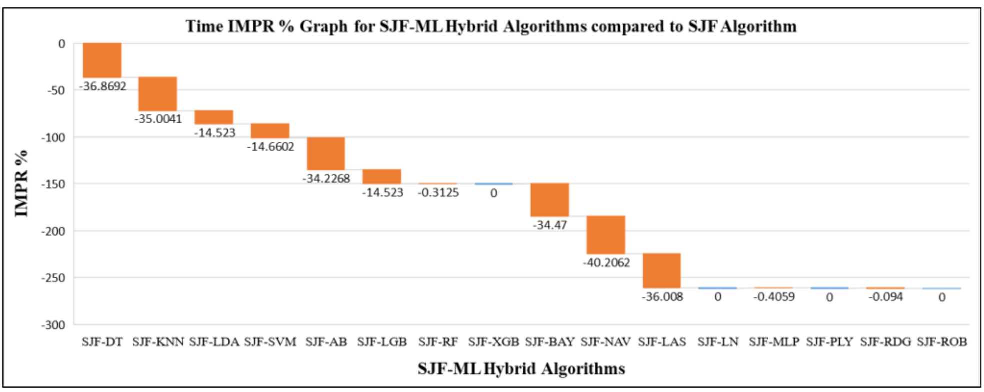

Figure 6. Time IMPR % graph for SJF-ML Hybrid Algorithms compared to SJF Algorithm.

The time results convey the SJF-LDA of the SJF-CA category, SJF-XGB of the SJF-ELA category, SJF-BAY of the SJF-PM category, and SJF-LN, SJF-PLY, and SJF-RBST algorithms of the SJF-RA provide the best performance within their respective categories. Additionally, the existing SJF algorithm, SJF-XGB of the SJF-SJF-ELA category, and SJF-LN, SJF-PLY, and SJF-RBST hybrid algorithms of the SJF-RA category provide the best time results across all the SJF-ML developed hybrid algorithms, along with the SJF algorithm.

6.3. Energy Results

Table 5 includes the energy comparison of SJF with presented SJF-ML Hybrid Algorithms.

Energy (μJ) Results		**VM Counts**					**AVG**	**IMPR% vs SJF**
Category	Algorithm	20	40	60	80	100		
Classical	SJF	835.24	296.15	170.71	118.17	89.84	302.02	-
SJF-CAs vs SJF	SJF-DT④	1176.96	369.41	198.98	132.82	97.87	395.21	-30.8556
	SJF-KNN③	1150.28	364.39	197.04	131.05	97.06	387.96	-28.4551
	SJF-LDA②	929.57	319.02	180.54	123.51	93.16	329.16	-8.9862
	SJF-SVM①⊛	845.06	271.73	147.69	98.23	72.52	287.05	4.9566
SJF-ELAs vs SJF	SJF-AB④	1137.59	362.22	196.88	131.15	97.29	385.03	-27.4849
	SJF-LGB③	929.57	319.02	180.54	123.51	93.16	329.16	-8.9862
	SJF-RF②	836.71	296.42	170.79	118.18	89.84	302.39	-0.1225
	SJF-XGB①	835.24	296.15	170.71	118.17	89.84	302.02	0.0000
SJF-PMs vs SJF	SJF-BAY①	1050.87	326.60	179.01	120.51	88.31	353.06	-16.8995
	SJF-NAV②	1142.25	362.89	196.71	130.99	97.04	385.98	-27.7995
SJF-RAs vs SJF	SJF-LAS④	1157.03	370.27	200.38	132.97	98.22	391.77	-29.7166
	SJF-LN①	835.23	296.15	170.71	118.17	89.84	302.02	0.0000
	SJF-MLP③	838.32	297.24	171.87	118.54	90.18	303.23	-0.4006
	SJF-PLY①	835.24	296.15	170.71	118.17	89.84	302.02	0.0000
	SJF-RDG②	835.89	296.31	170.78	118.2	89.86	302.21	-0.0629
	SJF-RBST①	835.24	296.15	170.71	118.17	89.84	302.02	0.0000

Table 5. Energy comparison of SJF with presented SJF-ML Hybrid Algorithms.

Table 5 shows the following energy comparison of SJF-ML hybrid algorithms compared to the SJF algorithm:

- **SJF-CA vs SJF**: SJF-SVM⊛ > SJF > SJF-LDA > SJF-KNN > SJF-DT
- **SJF-ELA vs SJF**: [SJF ≈ SJF-XGB] > SJF-RF > SJF-LGB > SJF-AB
- **SJF-PM vs SJF**: SJF > SJF-BAY > SJF-NAV
- **SJF-RA vs SJF**: [SJF ≈ SJF-LN ≈ SJF-PLY ≈ SJF-RBST] > SJF-RDG > SJF-MLP > SJF-LAS

Figure 7 shows Energy IMPR % graph for SJF-ML Hybrid Algorithms compared to SJF Algorithm.

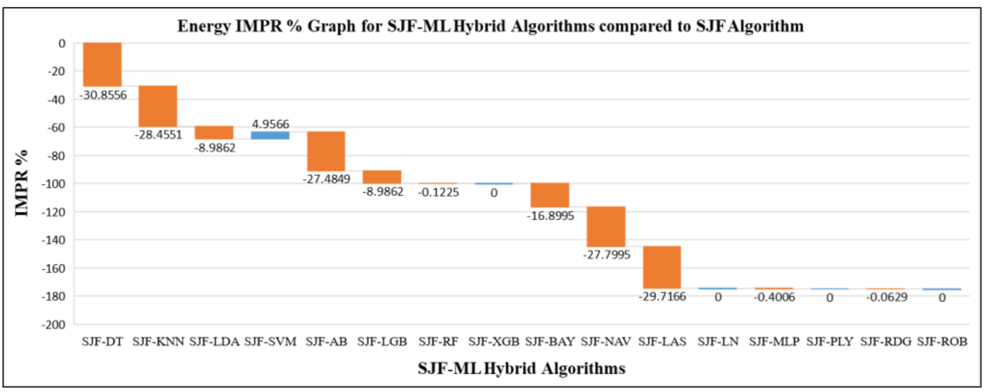

Figure 7. Energy IMPR % graph for SJF-ML Hybrid Algorithms compared to SJF Algorithm.

The energy results convey the SJF-SVM of the SJF-CA category, SJF-XGB of the SJF-ELA category, SJF-BAY of the SJF-PM category, and SJF-LN, SJF-PLY, and SJF-RBST hybrid algorithms of the SJF-RA provide the best performance within their respective categories. Additionally, the SJF-SVM hybrid algorithm of the SJF-CA category provides the best energy results across all the SJF-ML developed hybrid algorithms, along with the SJF algorithm.

6.4. LB Results

Table 6 includes the LB comparison of SJF with presented SJF-ML Hybrid Algorithms.

LB Results		**VM Counts**					**AVG**	**IMPR% vs SJF**
Category	Algorithm	20	40	60	80	100		
Classical	SJF	2.30	4.56	6.50	8.40	9.99	6.35	-
SJF-CAs vs SJF	SJF-DT②	1.48	3.08	4.8	6.31	7.83	4.70	25.9843
	SJF-KNN③	1.18	3.23	4.91	6.51	7.99	4.76	25.0394
	SJF-LDA④	2.21	4.5	6.51	8.48	10.16	6.37	0.3150
	SJF-SVM①⊛	0.23	0.75	1.24	1.62	1.14	1.00	84.2520
SJF-ELAs vs SJF	SJF-AB①	1.65	3.51	5.25	6.78	8.19	5.08	20.0000
	SJF-LGB④	2.21	4.5	6.51	8.48	10.16	6.37	0.3150
	SJF-RF②	2.29	4.54	6.47	8.37	9.96	6.33	0.3150
	SJF-XGB③	2.30	4.56	6.50	8.40	9.99	6.35	0.0000
SJF-PMs vs SJF	SJF-BAY①	1.48	3.28	5.03	6.61	8.04	4.89	16.3780
	SJF-NAV②	1.63	3.54	5.38	7.14	8.84	5.31	22.9921
SJF-RAs vs SJF	SJF-LAS①	1.23	2.79	4.44	6.07	7.70	4.45	29.9213
	SJF-LN④	2.30	4.56	6.50	8.40	9.99	6.35	0.0000
	SJF-MLP②	2.25	4.49	6.35	8.34	9.84	6.25	1.5748
	SJF-PLY④	2.30	4.56	6.50	8.40	9.99	6.35	0.0000
	SJF-RDG③	2.29	4.54	6.48	8.37	9.95	6.33	0.3150
	SJF-RBST④	2.30	4.56	6.50	8.40	9.99	6.35	0.0000

Table 6. LB comparison of SJF with presented SJF-ML Hybrid Algorithms.

Table 6 shows the following LB comparison of the SJF-ML hybrid algorithms compared to the SJF algorithm:

- **SJF-CA vs SJF**: SJF-SVM⊛ > SJF-DT > SJF-KNN > SJF > SJF-LDA
- **SJF-ELA vs SJF**: SJF-AB > SJF-RF > [SJF ≈ SJF-XGB] > SJF-LGB
- **SJF-PM vs SJF**: SJF-BAY > SJF-NAV > SJF
- **SJF-RA vs SJF**: SJF-LAS > SJF-MLP > SJF-RDG > [SJF ≈ SJF-LN ≈ SJF-PLY ≈ SJF-RBST]

Figure 8 shows LB IMPR % graph for SJF-ML Hybrid Algorithms compared to SJF Algorithm.

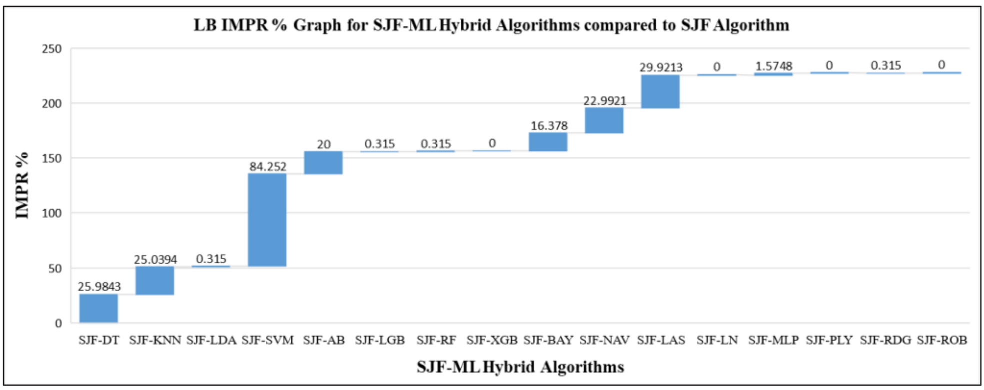

Figure 8. LB IMPR % graph for SJF-ML Hybrid Algorithms compared to SJF Algorithm.

The LB results convey the SJF-SVM of the SJF-CA category, SJF-AB of the SJF-ELA category, SJF-BAY of the SJF-PM category, and SJF-LAS of the SJF-RA provide the best performance within their respective categories. Additionally, the SJF-SVM hybrid algorithm of the SJF-CA category provides the best LB results across all the SJF-ML developed hybrid algorithms, along with the SJF algorithm.

7. Cumulative Results and Implications

This section includes filtering the better-performing SJF-ML hybrid algorithms obtained post the experiment along with their implications. Table 7 includes the comparative matrix representing the cumulative experimental results with performance of algorithms categorized as Best, Good, Average, Poor, or Worst.

Metric	Best	Good	Average	Poor	Worst
Cost	SJF SJF-XGB SJF-LN	SJF-BAY SJF-LDA SJF-PLY SJF-RF SJF-ROB	SJF-LGB SJF-NAV SJF-SVM SJF-RDG	SJF-KNN SJF-AB SJF-MLP	SJF-DT SJF-LAS
Time	SJF SJF-XGB SJF-LN SJF-PLY SJF-RBST	SJF-LDA SJF-RF SJF-BAY SJF-RDG	SJF-SVM SJF-LGB SJF-NAV SJF-MLP	SJF-KNN SJF-AB SJF-LAS	SJF-DT
Energy	SJF-SVM SJF SJF-XGB SJF-LN SJF-PLY SJF-RBST	SJF SJF-RF SJF-BAY SJF-RDG	SJF-LDA SJF-LGB SJF-NAV SJF-MLP	SJF-KNN SJF-AB SJF-LAS	SJF-DT
LB	SJF-SVM SJF-AB SJF-BAY SJF-LAS	SJF-DT SJF-RF SJF-NAV SJF-MLP	SJF-KNN SJF SJF-XGB SJF SJF-RDG	SJF SJF-LGB SJF-LN SJF-PLY SJF-RBST	SJF-LDA

Table 7. Comparative Matrix representing cumulative results for the experiment.

Table 7 shows that the developed SJF-ML hybrid algorithms, such as existing SJF, SJF-XGB of the SJF-ELA category, and SJF-LN of the SJF-RA category, provide the best cost results. The hybrid developed algorithms SJF-DT of the SJF-CA category and the SJF-LAS of the SJF-RA category are least suitable for the cost and can be avoided for hybridizations and implementations in the cloud. In terms of time performance metric, the existing SJF, along with SJF-XGB of the SJF-ELA category, SJF-LN, SJF-PLY, and SJF-RBST algorithms of the SJF-RA category provide the best results concerning time. Concerning time, the hybrid-developed SJF-DT algorithm of the SJF-CA category can be avoided for implementation since it provides poor results concerning time. In terms of energy, SJF-SVM of the SJF-CA category, existing SJF, SJF-XGB of the SJF-ELA category, and SJF-LN, SJF-PLY, SJF-RBST algorithms of the SJF-RA category provide the best results concerning energy. Like the time performance metric, the SJF-DT performance metric provides poor results concerning energy. Lastly, the SJF-SVM of the SJF-CA category, SJF-AB of the SJF-ELA category, SJF-BAY of the SJF-PM category, and SJF-LAS of the SJF-RA category provide the best results concerning LB. Here, the SJF-LDA of the SJF-CA category provides poor LB results. All the other developed hybrid algorithms provide either good, average, or poor performance compared within their SJF-ML categories, along with the SJF algorithm. Among the best SJF-ML performers, the SJF-XGB and SJF-LN algorithms perform better across maximum parameters, and the cloud can deploy these SJF-ML hybrid algorithms to ensure better results are obtained against the baseline of the SJF algorithm. Thus, the integration of ML with SJF provides fruitful results, improving the required cost, time, energy, and LB-considered performance metrics, and can be used to provide a better decision-making ability to the cloud to evolve itself from existing versions and provide better QoS.

8. Conclusion and Future Research Direction

This research paper provided hybrid scheduling methods by combining the currently existing best Shortest Job First (SJF) scheduling algorithm with sixteen Machine Learning (ML) algorithms. The ML algorithms within the categories of Classification Algorithms (CA), Ensemble Learning Algorithms (ELA), Probabilistic Models (PM), and Regression Algorithms (RA) were used to develop SJF-CA, SJF-ELA, SJF-PM, and SJF-RA as the four different categories of SJF-ML hybrid algorithms. From these sixteen SJF-ML hybrid categories, SJF-ML hybrid algorithms including SJF-AB, SJF-BAY, SJF-DT, SJF-KNN, SJF-LASSO, SJF-LDA, SJF-LGBM, SJF-LIN, SJF-MLPNN, SJF-NAVBAY, SJF-POLY, SJF-RDGE, SJF-RF, SJF-ROBST, SJF-SVM, and SJF-XGB are developed across the four SJF-ML categories. The main aim was to embed the cloud's scheduling process with ML techniques and improve the scheduling by enhancing the considered performance metrics such as cost, time, energy, and LB. The existing SJF algorithm was used as a baseline to compare it with all the developed SJF-ML algorithms across all the mentioned metrics. The real-time Google big data tasks were computed using all the SJF-ML hybrid algorithms across five different experimental scenarios of VM counts. From the experiment conducted, it can be conveyed that the SJF-XGB of the SJF-ELA category and the SJF-LN of the SJF-RA category provided better results across the maximum considered parameters. The cloud can deploy these variants for scheduling, rather than the SJF algorithm. A major limitation of the presented SJF-ML hybrid algorithms is the complexity of integration, where the coordination of both algorithms needs to be carefully executed. An additional limitation is the overhead of time spent in training the SJF-ML algorithms, which adds to the unwanted latency and energy consumption. The experiment findings conclude that the development of ML algorithms with scheduling approaches in the cloud provides better results through their improved decision-making, pattern recognition of BDTs, and overall allocation of resources, therefore providing a better, enhanced, and evolved version of itself. As a part of future research direction, the author aims to develop a second-level hybridization where the best performers, SJF-XGB and SJF-LN algorithms, will be integrated to ensure further improvements can be made for the cloud evolution. This second-level algorithm will be named SJF-XL, a combination of SJF-XGB and SJF-LN, where the individual first-level hybrid algorithm characteristics will be used to provide better and enhanced scheduling results to the first-level evolved cloud. Figure 9 represents the architecture where the SJF-XL second-level hybrid algorithm will be used by the cloud.

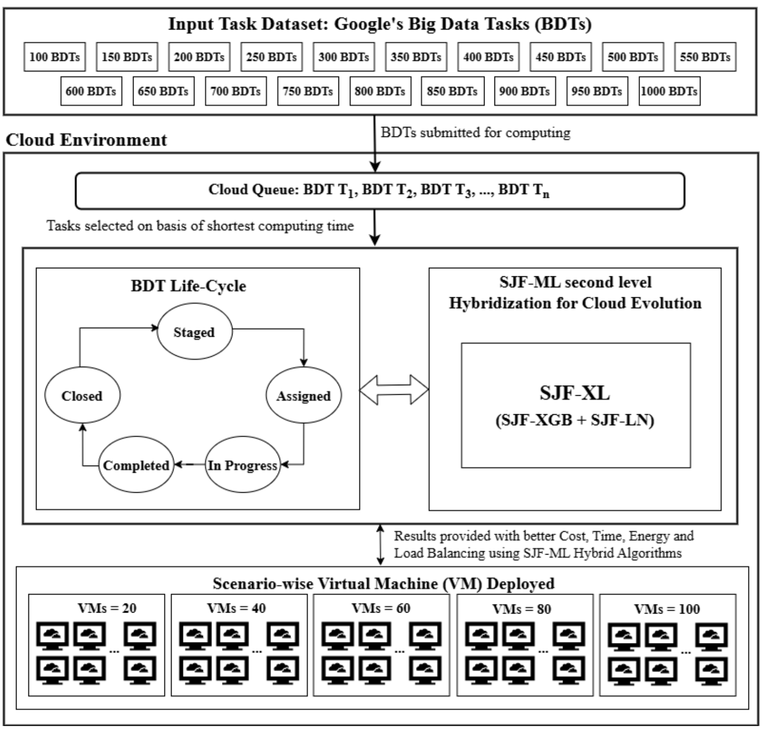

Figure 9. Second-Level SJF-ML Architecture.

References

Chaudhary, H., Sharma, G., Nishad, D. K., & Khalid, S. (2025). AI-enhanced modelling of queueing and scheduling systems in cloud computing. Deleted Journal, 7(4).

Chung, W., Tong, J., & Chen, Z. (2025). A fine-grained GPU sharing and job scheduling for deep learning jobs on the cloud. The Journal of Supercomputing, 81(2).

Kathole, A. B., Vhatkar, K., Lonare, S., & Kshirsagar, A. P. (2025). Optimization-based resource scheduling techniques in cloud computing environment: A review of scientific workflows and future directions. Computers & Electrical Engineering, 123, 110080.

Liu, Z. “. (2025). Ensemble Learning. In Springer (pp. 221–242).

Sanjalawe, Y., Al-E’mari, S., Fraihat, S., & Makhadmeh, S. (2025). AI-driven job scheduling in cloud computing: a comprehensive review. Artificial Intelligence Review, 58(7).

Sonia, N., & Nath, R. (2025). A systematic review of various load balancing approaches in cloud computing utilizing machine learning and deep learning. International Journal of Data Science and Analytics.

Verma, G. (2025). Sustainable Cost-Energy Aware load balancing in cloud environment using intelligent optimization. Sustainable Computing Informatics and Systems, 101115.

Ye, J., Peng, Y., Li, Y., Li, Z., & Huang, J. (2025). Asynchronous Control based Aggregation Transport Protocol for Distributed Deep Learning. IEEE Transactions on Computers, 1–14.

Ali, A. A., Umamaheswari, S., Khan, A. B. F., & Jayabrabu, R. (2024). Fuzzy rules-based Data Analytics and Machine Learning for Prognosis and Early Diagnosis of Coronary Heart Disease. Journal of Information and Organizational Sciences, 48(1), 167–181.

Almurshed, O., Kaushal, A., Meshoul, S., Muftah, A., Almoghamis, O., Petri, I., Auluck, N., & Rana, O. (2024). Enhancing performance of machine learning tasks on edge-cloud infrastructures: A cross-domain internet of things based framework. Future Generation Computer Systems, 166, 107696.

Alsubaei, F. S., Hamed, A. Y., Hassan, M. R., Mohery, M., & Elnahary, M. K. (2024). Machine learning approach to optimal task scheduling in cloud communication. Alexandria Engineering Journal, 89, 1–30.

Bartakke, J., & Kashyap, R. (2024). The usage of clouds in Zero-Trust security strategy. Journal of Information and Organizational Sciences, 48(1), 149–165.

Hayyolalam, V., & Özkasap, Ö. (2024). CBWO: A novel multi-objective load balancing technique for cloud computing. Future Generation Computer Systems, 107561.

He, H., Gu, Y., Liu, Q., Wu, H., & Cheng, L. (2024). Job scheduling in hybrid clouds with privacy constraints: A deep reinforcement learning approach. Concurrency and Computation Practice and Experience.

Hosseini, S. M., Shirvani, M. H., & Motameni, H. (2024). Multi-objective discrete Cuckoo search algorithm for optimization of bag-of-tasks scheduling in fog computing environment. *Computers & Electrical Engineering, 119*, 109480.

Huymajer, M., Filzmoser, P., Mazak-Huemer, A., Winkler, L., & Kraxner, H. (2024). Opportunities and pitfalls of regression algorithms for predicting the residual value of heavy equipment — A comparative analysis. Engineering Applications of Artificial Intelligence, 141, 109599.

Kanchetti, D., Munirathnam, R., & Thakkar, D. (2024). Integration of Machine Learning Algorithms with Cloud Computing for Real-Time Data Analysis. Journal for Research in Applied Sciences and Biotechnology, 3(2), 301–306.

Laha, J., Pattnaik, S., Chaudhury, K. S., & Palai, G. (2024). Reducing makespan and enhancing resource usage in cloud computing with ESJFP Method: a new Dynamic Approach. Internet Technology Letters.

Li, Y., Zhong, W., & Wu, Y. (2024). Multi-objective flexible job-shop scheduling via graph attention network and reinforcement learning. The Journal of Supercomputing, 81(1).

Miraki, A., Parviainen, P., & Arghandeh, R. (2024). Probabilistic forecasting of renewable energy and electricity demand using Graph-based Denoising Diffusion Probabilistic Model. Energy and AI, 19, 100459.

Pradhan, A., Das, A., & Bisoy, S. K. (2024). Modified parallel PSO algorithm in cloud computing for performance improvement. Cluster Computing, 28(2).

Rahimi, N., Kamankesh, A., Amiridis, I. G., Daneshgar, S., Sahinis, C., Hatzitaki, V., & Enoka, R. M. (2024). Distinguishing among standing postures with machine learning-based classification algorithms. Experimental Brain Research, 243(1).

Rekha, S., Mohanapriya, D., Selvi, S., Banupriya, C., & Gobi, M. (2024). A Reinforced PSO Algorithm With SJF-MMBF For Allocating Work In Cloud Computing. IEEE, 1–6.

Salehi, S. (2024). Comparative analysis of machine learning techniques for cryptocurrency price prediction. Journal of Information and Organizational Sciences, 48(2), 341–352.

Saraswat, J. K., & Choudhari, S. (2024). Integrating big data and cloud computing into the existing system and performance impact: A case study in manufacturing. Technological Forecasting and Social Change, 210, 123883.

Zende, S., Singh, T., & Suryavanshi, M. (2024). Comprehensive review on Machine learning applications in Cloud Computing. International Journal of Innovative Research in Computer Science & Technology, 12(4), 16–24.

Asghari Alaie, Y., Hosseini Shirvani, M. & Rahmani, A.M. (2023). A hybrid bi-objective scheduling algorithm for execution of scientific workflows on cloud platforms with execution time and reliability approach. J Supercomput 79, 1451–1503.

Hu, C., Zhang, C., Lei, D., Wu, T., Liu, X., & Zhu, L. (2023). Achieving Privacy-Preserving and verifiable support vector machine training in the cloud. IEEE Transactions on Information Forensics and Security, 18, 3476–3491.

Kumar, R. (2023). Usage of Cloud Computing Application by Students from Kurukshetra University. Journal of Information and Organizational Sciences, 47(1), 189–201.

Nayak, A. A., & Shetty, S. (2023). A Systematic Analysis on Task Scheduling Algorithms for Resource Allocation of Virtual Machines on Cloud Computing Environments. IEEE Xplore, 1–6.

Zhang, R., Liu, Z., Tian, G., Lu, Y., & Sujatha, K. (2023). A Study of Real-Time Scheduling Algorithms in Cluster Environment Based on Machine Learning. IEEE, 682–686.

Chauhan, N., & Agrawal, R. (2022). Probabilistic Optimized Kernel Naive Bayesian Cloud Resource Allocation System. Wireless Personal Communications, 124(4), 2853–2872.

Kumar, Y., Kaul, S., & Hu, Y. (2022). Machine learning for energy-resource allocation, workflow scheduling and live migration in cloud computing: State-of-the-art survey. Sustainable Computing Informatics and Systems, 36, 100780.

Olofintuyi, S. S. (2022). Early Cocoa Blackpod Pathogen Prediction with Machine Learning Ensemble Algorithm based on Climatic Parameters. Journal of Information and Organizational Sciences, 46(1), 1–14.

Sahkhar, L., Balabantaray, B. K., & Yadav, S. S. (2022). Efficient cloudlet allocation to virtual machine to impact cloud system performance. International Journal of Information System Modeling and Design, 13(6), 1–21.

Talwani, S., Singla, J., Mathur, G., Malik, N., Jhanjhi, N. Z., Masud, M., & Aljahdali, S. (2022). Machine-Learning-Based Approach for Virtual Machine Allocation and Migration. Electronics, 11(19), 3249.

Verma, G. (2022). Hybrid optimization model for secure task scheduling in Cloud: Combining Seagull and Black Widow Optimization. *Cybernetics & Systems*, *55*(8), 2489–2511.

Kumar, S., & Dumka, A. (2021). Load Balancing with the Help of Round Robin and Shortest Job First Scheduling Algorithm in Cloud Computing. Springer Nature, 213–223.

Murad, S. A., Azmi, Z. R. M., Muzahid, A. J. M., & Al-Imran, M. (2021). Comparative Study on Job Scheduling Using Priority Rule and Machine Learning. IEEE Xplore, 1–8.

Nassif, A. B., Talib, M. A., Nasir, Q., Albadani, H., & Dakalbab, F. M. (2021). Machine Learning for Cloud Security: A Systematic Review. IEEE Access, 9, 20717–20735.

Rekha, S., & Kalaiselvi, C. (2021). Load balancing using SJF-MMBF and SJF-ELM approach. International Journal of Scientific Research in Computer Science Engineering and Information Technology, 74–86.

S, R., & C, K. (2021). Secure And Energy Aware Task Scheduling In Cloud Using Deep Learning And Cryptographic Techniques. ICTACT Journal on Communication Technology, 12(2), 2434–2441.

Tanha, M., Shirvani, M. H., & Rahmani, A. M. (2021). A hybrid meta-heuristic task scheduling algorithm based on genetic and thermodynamic simulated annealing algorithms in cloud computing environments. *Neural Computing and Applications*, *33*(24), 16951–16984.

Younis, M. F. (2021). ESJF algorithm to improve cloud environment. Iraqi Journal of Science, 4171–4180.

Rekha, S., & Kalaiselvi, D. (2020). Multi Level Queue Scheduling With Particle Swarm Optimization (Mlqs-Pso) Of Vms in Queueing Heterogeneous Cloud Computing Systems. International Journal of Recent Technology and Engineering (IJRTE), 8(5), 664–672.

Rjoub, G., Bentahar, J., Wahab, O. A., & Bataineh, A. S. (2020). Deep and reinforcement learning for automated task scheduling in large-scale cloud computing systems. Concurrency and Computation Practice and Experience, 33(23).

Guo, M., Guan, Q., Chen, W., Ji, F., & Peng, Z. (2019). Delay-Optimal Scheduling of VMs in a Queueing Cloud Computing System with Heterogeneous Workloads. IEEE Transactions on Services Computing, 15(1), 110–123.

Rjoub, G., Bentahar, J., Wahab, O. A., & Bataineh, A. (2019). Deep Smart Scheduling: A Deep Learning Approach for Automated Big Data Scheduling Over the Cloud. IEEE Xplore, 189–196.

Samie, F., Bauer, L., & Henkel, J. (2019). From Cloud Down to Things: An overview of Machine learning in Internet of Things. IEEE Internet of Things Journal, 6(3), 4921–4934.

Tong, Z., Deng, X., Chen, H., Mei, J., & Liu, H. (2019). QL-HEFT: a novel machine learning scheduling scheme base on cloud computing environment. Neural Computing and Applications, 32(10), 5553–5570.

Guo, M., Guan, Q., & Ke, W. (2018). Optimal scheduling of VMs in queueing cloud computing systems with a heterogeneous workload. IEEE Access, 6, 15178–15191.

Elmougy, S., Sarhan, S., & Joundy, M. (2017). A novel hybrid of Shortest job first and round Robin with dynamic variable quantum time task scheduling technique. Journal of Cloud Computing Advances Systems and Applications, 6(1).

Mo, Y., Kontonatsios, G., & Ananiadou, S. (2015). Supporting systematic reviews using LDA-based document representations. Systematic Reviews, 4(1).

Reiss, C., Tumanov, A., Ganger, G. R., Katz, R. H., & Kozuch, M. A. (2012). Heterogeneity and dynamicity of clouds at scale. *ACM*.

Journal of Information and Organizational Sciences
Volume 49, Number 2 (2025)
Journal homepage: jios.foi.hr

JIOS

DOI: 10.31341/jios.49.2.4
Open Access

UDC 37.091.3:004-047.37(497.5)
Original Scientific Paper

Validation of the TPACK Instrument in the Croatian Primary School Context

Blanka Runtić[1], Matilda Karamatić Brčić[2] and Goran Kuvačić[3*]

[1]Primary School Kman-Kocunar, Split, Croatia
[2]Department of Pedagogy, University of Zadar, Zadar, Croatia
[3]Faculty of Kinesiology, Universtity of Split, Split, Croatia
[*]Correspondence: goran.kuvacic@kifst.eu

PAPER INFO

Paper history:
Received 19 April 2025
Accepted 26 July 2025

Citation:
Runtić, B., Karamatić Brčić, M & Kuvačić, G. (2025). Validation of the TPACK Instrument in the Croatian Primary School Context. In Journal of Information and Organizational Sciences, vol. 49, no. 2, pp. 213-234

ABSTRACT

This study aimed to validate a generic TPACK questionnaire within the Croatian educational context, involving 609 in-service subject teachers from five regions. The survey-based research included exploratory and confirmatory factor analyses. The resulting structure diverged from the original seven-factor model, yielding five factors: pedagogical knowledge (PK), technological knowledge (TK), content knowledge (CK), technological content knowledge (TCK), and technological pedagogical content knowledge (TPACK). Some items were excluded to enhance validity. Pedagogical and pedagogical content knowledge merged, as did technological pedagogical and technological pedagogical content knowledge. The final five-factor model exhibited strong psychometric properties, characterized by high internal consistency and a good fit to the data. Evidence of convergent and discriminant validity supports the distinct yet interconnected nature of the identified knowledge domains. Significant positive correlations were found between all constructs, with the strongest correlations between TPACK and TCK, and also between TPACK and TK. The findings underscore the contextuality of the TPACK framework, and therefore, they were discussed in relation to recent national educational initiatives as well as international indicators on digital competencies. These results contribute to understanding the TPACK framework in the Croatian context, supporting future research on teacher competencies in integrating technology into education.

Keywords: in-service teachers, educational technology integration, factor analysis, teacher competencies, measurement validation

1. Introduction

In preparing students for the labor market, teachers are expected to effectively integrate information and communication technology (ICT) into the teaching process to enhance learning and contribute to the achievement of educational goals. To apply technology effectively in the classroom, teachers must connect knowledge from their subject area, the teaching methods they use in their specific classroom context, and technology's role in 21st-century learning. This concept is best explained through the Technological Pedagogical Content Knowledge (TPACK) model (Mishra & Koehler, 2006) in which the authors identify seven forms of teacher knowledge that emerge from the interaction of content, pedagogical, and technological knowledge. It is often visualized as a Venn diagram with three overlapping circles representing the types of knowledge and an outer circle symbolizing the context.

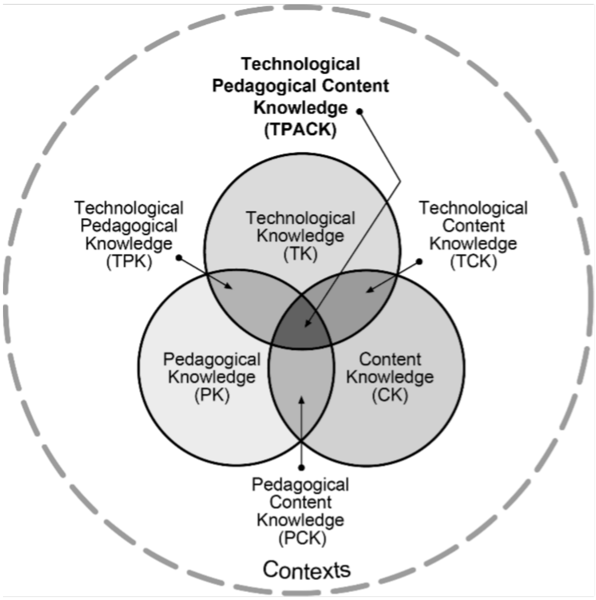

Figure 1. TPACK framework (Reproduced by permission of the publisher, © 2012 by tpack.org)

Building upon Shulman's earlier model of Pedagogical Content Knowledge (PCK) (Shulman, 1986, 1987), the authors (Koehler & Mishra, 2009; Mishra & Koehler, 2006) added the dimension of technology alongside content and pedagogy, and explained seven types of knowledge: content knowledge (CK) – knowledge of the subject area, including key concepts, theories, ideas, and organizational frameworks; pedagogical knowledge (PK) – knowledge of teaching and learning processes, including teaching methods, assessment, classroom management, and understanding of learning theories; technological knowledge (TK) – knowledge about using technology and digital tools, and the ability to adapt to technological changes; pedagogical content knowledge (PCK) – knowledge of teaching specific content adapted to the needs, prior knowledge, and interests of students; technological content knowledge (TCK) – understanding the relationship between content and technology and how technology can enhance the presentation and understanding of specific content; technological pedagogical knowledge (TPK) – knowledge about applying technology to improve teaching and learning while considering specific methods and strategies; and technological pedagogical content knowledge (TPACK) – knowledge that connects content, pedagogy, and technology for effective teaching in specific contexts. Although TPACK is often used as a model for the effective integration of technology into the teaching process, it is important to emphasize that, in fact, two knowledge dimensions within this model refer specifically to the integration of technology in teaching, TPK and TPACK (Graham, 2011). Together with the remaining dimensions, they provide insight into the complexity and dynamic nature of teacher knowledge.

Recently, an eighth dimension of contextual knowledge (XK) was proposed in the conceptualization of the model to encompass the specific knowledge required for teaching in particular contexts (Mishra, 2019; Petko et al., 2025). This dimension extends the original TPACK model by recognizing that effective teaching with technology is not universally applicable but depends on social, cultural, institutional, and situational factors. Although the TPACK model has been widely accepted and globally used as a theoretical framework in many studies, it has yet to gain significant recognition within the broader scientific community in the Republic of Croatia. In fact, only two TPACK-related studies have been conducted, both involving pre-service teachers (Dobi Barišić et al., 2019; Dobi Barišić & Brust Nemet, 2023). The curriculum reform in Croatia has further emphasized the role of technology through the adoption of the Cross-Curricular Topic Curriculum on the *Use of Information and Communication Technology* in primary and secondary schools, which promotes appropriate, responsible, and creative use of ICT across all educational domains (Ministry of Science and Education, 2019a). This has created a framework for applying the TPACK model in research within the Croatian educational context. Therefore, this study is primarily focused on validating the TPACK questionnaire in order to determine its applicability to a sample of in-service teachers within the Croatian education system.

1.1. Literature overview

TPACK research has employed quantitative, qualitative, and mixed-methods research approaches and has been conducted on samples of pre-service teachers, in-service teachers, university professors, online teachers, and teachers in professional development education. In quantitative research, various questionnaires have been employed, including general TPACK questionnaires, questionnaires for specific technologies, teaching methods, and content areas (Chai et al., 2016). The constructed questionnaires examined the interrelationships among TPACK factors, as well as the interrelationships with demographic variables, teacher beliefs, self-regulated learning, learning approaches and learning environment. Qualitative methods used in TPACK research can be classified into the following categories: observation, performance evaluation, open-ended questionnaires, and interviews (Koehler et al., 2012).

Previous validations of the TPACK questionnaire have focused on testing the reliability and validity of this self-assessment tool for teacher knowledge, as well as its adaptation to different contexts and cultures. Results indicate that the TPACK framework is widely accepted; however, challenges remain regarding its universal applicability. Using one or more statistical methods, such as exploratory factor analysis, confirmatory factor analysis, and structural equation modeling, several studies have confirmed the seven-factor structure in line with the theoretical model (Baser et al., 2016; Chai, Ng, et al., 2013; Deng et al., 2017; Dong et al., 2015; Lin et al., 2013; Pamuk et al., 2015; Sahin, 2011; Schmid et al., 2020; Schmidt et al., 2009; Valtonen et al., 2017). In some questionnaires used in these studies, more than seven factors were identified, as content knowledge was subdivided into multiple subscales depending on the sample (Chai et al., 2011; Schmidt et al., 2009). However, numerous studies have identified different numbers of factors in general TPACK questionnaires, TPACK questionnaires for specific subject areas, particular teaching methods, or the application of specific technologies: three factors (Archambault & Barnett, 2010; Luik et al., 2018), four (Chai et al., 2010; Jang & Tsai, 2012; Kabakci Yurdakul et al., 2012; Zelkowski et al., 2013), five (Karadeniz & Vatanartıran, 2013; M. H. Lee & Tsai, 2010; Semiz & Ince, 2012), six (Bostancıoğlu & Handley, 2018; Liang et al., 2013; Valtonen et al., 2017), eight (Canbazoğlu Bilici et al., 2013; Sang et al., 2016; Shinas et al., 2013) and nine factors (Ritzhaupt et al., 2016).

In Croatia, Dobi Barišić et al. (2019) validated the TPACK questionnaire developed by Schmidt et al. (2009) The *Survey of Pre-service Teachers' Knowledge of Teaching and Technology* (SPTKTT) on a sample of early childhood and primary education students. The factor structure was generally confirmed in line with the original model, although factor analysis results indicated certain differences in the distribution of individual items compared to the original validation, as well as contextual differences resulting from the organization of subject content in Croatian primary education (Dobi Barišić et al., 2019).

Considering the limited research on the knowledge required for successful technology integration in teaching in the Republic of Croatia, the contextual nature of TPACK practices, and the globally inconsistent results of TPACK questionnaire validations, it is necessary to validate the TPACK questionnaire on a sample of Croatian in-service teachers. This will enable further research on TPACK dimensions of knowledge within the Croatian educational context, which is especially important given the ongoing reform of the education system in Croatia and the adoption of subject and cross-curricular curricula (Ministry of Science and Education, 2019b). Since the 2021/2022 school year, all school classes and teachers in primary and secondary schools have been included in the education reform and are required to meet the educational expectations of the cross-curricular topic *Use of Information and Communication Technology*.

Therefore, this study aimed to validate the TPACK questionnaire to determine its applicability to a sample of in-service subject teachers in the Croatian education system. Through the validation of the instrument, the goal was to examine the factor structure of the TPACK model and to determine the reliability and validity of the measurement tool used, enabling future research into teacher competencies in the area of technology integration in the educational process.

2. Materials and Methods

2.1. Research Design

This cross-sectional study employed a quantitative research approach using a survey method. Surveying is considered the most appropriate method for assessing trends or characteristics of a specific population, determining relationships between variables, and comparing groups of respondents. It also enables the economical collection of data from geographically dispersed populations (Creswell, 2015). Additionally, it allows for rapid data collection from a large number of respondents while ensuring anonymity and facilitating the quantification of the collected data.

2.2. Participants

The study's target population consisted of primary school subject teachers in the Republic of Croatia. A cluster sampling strategy was employed considering the population's size and distribution, as well as the data collection method (Cohen et al., 2007). The study sample included subject teachers from randomly selected schools in five Croatian regions: Slavonia, Northern Croatia, the City of Zagreb, Istria and the Croatian Littoral with its Hinterland, and Dalmatia. In accordance with the available data, schools in both urban and rural areas, including island settlements, were included, with schools in each area selected randomly.

Due to the anticipated lower response rate in online surveys, a larger number of schools was selected than would be needed if all teachers responded. From the list of primary schools provided by the Ministry of Science and Education, 70 schools were randomly selected out of a total of 862 main schools and 1,101 branch schools in the 2021/2022 school year. A total of 651 teachers responded, which constitutes a representative sample for the target population of 16,518 subject teachers in Croatia (Primary and Secondary Schools, Croatian Bureau of Statistics). Further analysis of the dataset resulted in the removal of incomplete responses, resulting in a total of 632 participants. A statistical analysis was then conducted to remove multivariate outliers where 23 participants exceeded the critical value in the Mahalanobis distance analysis ($\chi2$ (42) = 74.576, p = 0.001). Consequently, the final dataset used for statistical analysis included 609 teachers.

2.3. Instrument

In this study, teachers' self-assessments of their knowledge were examined in accordance with the components of the TPACK model, which include the following dimensions: (1) pedagogical knowledge, (2) content knowledge, (3) technological knowledge, (4) pedagogical content knowledge, (5) technological content knowledge, (6) technological pedagogical knowledge, and (7) technological pedagogical content knowledge. The results for each of these dimensions were calculated as the arithmetic mean of the scores of the respective items within each subscale. Since responses to the items were rated on a scale from 1 to 5, the values of each dimension also range within this interval.

In line with the study's aim and the operationalization of the research tasks, a questionnaire was constructed, consisting of two parts: sociodemographic data of the teachers and the TPACK scale. The part of the questionnaire related to sociodemographic data included nine multiple-choice questions covering gender, age, years of work experience in the profession, school location, initial education, educational field, professional advancement, and participation in professional development.

The TPACK scale was translated and adapted from a questionnaire developed by Schmid et al. (Schmid et al., 2020). To enable broader studies that would include teachers from various subject areas and educational levels, the authors constructed a scale in which most items were adapted or taken from previously validated questionnaires (Chai et al., 2011; Schmidt et al., 2009; Valtonen et al., 2017), while the authors themselves created some items based on existing literature and definitions. Earlier versions of TPACK instruments were generally applicable to specific populations of students or teachers who taught individual subjects and were focused on specific technologies or teaching methods. Therefore, the authors designed generically worded items to create a shortened version of the TPACK scale (28 items), aiming to reduce respondents' cognitive fatigue and facilitate the combination with other scales and applicability across a diverse teacher population.

Considering the specific characteristics of the sample within the Croatian education system and the fact that the TPACK questionnaire was being applied to teachers from various subject areas for the first time without being combined with other scales, a longer version of this generic questionnaire was selected. This version contains 42 items instead of 28. It was assumed that the longer version would provide better coverage of all TPACK construct dimensions, greater measurement reliability and validity, and better potential for factor analysis to effectively examine the structure of the measured construct in the national Croatian context. The longer version of the questionnaire contains 42 statements distributed across seven subscales: pedagogical knowledge (7 items), content knowledge (6 items), technological knowledge (7 items), pedagogical content knowledge (6 items), technological pedagogical knowledge (5 items), technological content knowledge (6 items), and technological pedagogical content knowledge (5 items). Each statement was rated on a five-point Likert scale, where respondents assessed their level of agreement with the statements (1 – strongly disagree; 2 – mostly disagree; 3 – neither agree nor disagree; 4 – mostly agree; 5 – strongly agree).

The content validity of the TPACK questionnaire was established during the translation and adaptation process. A university professor of pedagogy, two English language professors, and a subject teacher participated in the translation of the TPACK scale into Croatian. After harmonizing the translation, a pilot test was conducted with a small group of respondents. The aim was to obtain feedback on the clarity of questions and items to eliminate potential ambiguities and assess the time required to complete the questionnaire (Cohen et al., 2007). The pilot included eight subject teachers from the educational institution where one of the author works. Based on participant feedback, no major content ambiguities were identified. Two respondents suggested adding the word “digital” to the term “technology” in statements related to technological knowledge to clarify the concept. Although it was stated in the introductory note of the technological knowledge subscale that the statements referred to digital technologies (computers, tablets, mobile phones, the internet, etc.), the suggestion was accepted to include the word “digital” in the item wording for greater clarity. A professor of Croatian language also provided several stylistic improvement suggestions, which were incorporated into the final version of the questionnaire. A professor of Croatian language and literature proofread the translated and revised questionnaire.

2.4. Procedures

The survey was conducted online, considering the advantages and limitations of this data collection method. Online questionnaires enable the rapid and efficient collection of a large amount of data; however, they may also result in lower response rates and are more likely to reach respondents who have internet access and possess digital technology skills (Menon & Muraleedharan, 2020). Nevertheless, the use of an anonymous online questionnaire increases the level of confidentiality, as no third parties are involved in the distribution or collection process, unlike with traditional paper-based versions. The questionnaire was created using the digital application Microsoft Forms, which enabled easy distribution and systematic collection of participant responses.

In May and June of 2022, contact was established with principals of the 70 selected primary schools in five regions of the Republic of Croatia. According to the data available on the official website of the Ministry of Science and Education, initial telephone conversations were held with the principals to explain the purpose, goal, and procedures of the research. Principals of 68 schools agreed to participate in the scientific research, committing either personally or through school pedagogues to forward the online survey to subject teachers via the school’s established online communication channels (e-mail, Microsoft Teams, or WhatsApp). The principals of two contacted schools declined participation.

Following the initial phone call, an email was sent to those who agreed to participate, containing brief information about the research and a link to the online questionnaire. The Ethics Committee of the University of Zadar issued approval for conducting this study, confirming its compliance with ethical standards and the appropriateness of the methodological procedures used during data collection and analysis.

2.5. Statistical analysis

Descriptive statistical parameters were determined for all variables used in this study (arithmetic mean [M], standard deviation [SD], minimum [Min] and maximum [Max] measurement values, skewness, and kurtosis of the distribution). To determine the factor structure of the TPACK questionnaire, exploratory factor analysis (EFA) was conducted on the first random subsample ($n = 312$). The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett’s test of sphericity (BTS) were applied prior to factor extraction to ensure that the data were suitable for EFA (Field, 2009). Multivariate outliers were identified by Mahalanobis distance, while multicollinearity was assessed through the correlation matrix and the Variance Inflation Factor (VIF), with values <5 considered acceptable. Principal Axis Factoring (PAF) was used for factor extraction, while the number of factors was determined using the Guttman-Kaiser criterion (eigenvalue > 1), Scree test (Cattell, 1966) and parallel analysis (Horn, 1965). While the Scree test is a visual method, parallel analysis establishes a criterion for factor extraction by comparing the eigenvalues from the actual data with those from randomly generated data sets with the same parameters (O’Connor, 2000). In this case, SPSS syntax was used: the factor was retained if the eigenvalue from the actual data was greater than the corresponding eigenvalue from the random data. Due to the assumption that the factors would be correlated, oblique rotation (Promax) was applied to enable a more realistic interpretation. Pattern coefficients from the Promax rotation were used to evaluate item-factor relationships, with items excluded if communalities were < 0.30, loadings < 0.40, or cross-loadings > 0.30 (Hair et al., 2010). Confirmatory Factor Analysis (CFA) using the Maximum Likelihood Estimates (ML) technique was conducted on the second random subsample ($n = 297$) to examine the stability of the factors obtained through EFA. A minimum standardized factor loading of 0.50 was used as the threshold for retaining items in the CFA model (Hair et al., 2010) Standard goodness-of-fit indices and their recommended values were used: $\chi^2/df < 3$, RMSEA (Root Mean Square Error of Approximation) ≤ 0.06, SRMR (Standardized Root Mean Square Residual) ≤ 0.05, TLI (Tucker–Lewis Index) ≥ 0.90, and CFI (Comparative Fit Index) ≥ 0.90 (Hooper et al., 2008). Convergent and discriminant validity were assessed through reliability coefficients and the following thresholds: AVE (Average Variance Extracted) ≥ 0.5, CR (Composite Reliability) > 0.7, and CR > AVE, indicating acceptable convergent validity. Discriminant validity was considered satisfactory if the correlation between factor scores was significant and if the correlation coefficient was lower than the square root of the corresponding AVE (Fornell-Larcker criterion). Additionally, the Heterotrait-Monotrait ratio (HTMT) ≤ 0.85 was used. To further support discriminant validity, the maximum shared variance (MSV) and average shared variance (ASV) were calculated, with the criterion that AVE should exceed both MSV and ASV for each construct. Internal consistency reliability in both EFA and CFA was determined using two coefficients: Cronbach's alpha (α) > 0.7 and McDonald's omega (ω) coefficient > 0.7 (McDonald, 1999). Since Cronbach's alpha is often criticized for underestimating internal consistency, omega was also calculated as an alternative. To determine the correlations between knowledge components according to the TPACK model, Pearson's correlation coefficient was used. The strength of correlations was evaluated using the following scale: 0.1, trivial; 0.1–0.3, small; 0.3–0.5, moderate; 0.5–0.7, large; 0.7–0.9, very large; 0.9, nearly perfect. The level of statistical significance was set at 0.05. All data were analyzed using SPSS 28.0 and AMOS 26.0 (SPSS, Chicago, IL, USA).

3. Results

3.1. Sociodemographic Characteristics

The study was conducted on a sample of 609 participants, with the majority being female (82.9% women and 17.1% men). Regarding age, the most represented categories were between 30 and 39 years (32.5%) and between 40 and 49 years (30.7%), followed by the 50- to 59-year-old age group (23.2%). Participants under the age of 30 made up 9.4% of the sample, while those over 60 were the least represented (4.3%). In terms of work experience, most participants had between 11 and 20 years of experience (32.2%), followed by those with 6 to 10 years (20.5%) and 21 to 30 years (20.5%). A total of 16.91% had less than five years of experience, and the smallest group included those with more than 30 years of teaching experience (9.9%). With respect to educational fields, most participants taught in the language and communication area (34%), followed by the social and humanistic area (12.6%), natural sciences (11%), interdisciplinary field (10.8%), mathematics (9.2%), technical and ICT field (11.2%), the arts (6.1%), and the least represented was the physical and health education field (5.1%).

3.2. Factor Structure – Exploratory Factor Analysis (EFA)

The skewness and kurtosis coefficients for all items range between -2 and $+2$, indicating the absence of univariate outliers in the data (see Appendix A). Regarding multivariate outliers, Mahalanobis distance analysis identified a total of 23 participants who exceeded the critical value ($\chi^2(42) = 74.576$, $p = 0.001$). Consequently, these participants were excluded from further analysis. Regrading multicollinearity, all correlation coefficients were below 0.9 (ranging from 0.07 to 0.8), with a large number above 0.3 ($n = 644$). Therefore, there is no indication of multicollinearity, and the data can be considered suitable for EFA. The Kaiser–Meyer–Olkin (KMO) measure was 0.949, and Bartlett's Test of Sphericity (BTS) yielded $\chi^2(861) = 9554.29$, $p < 0.001$. These results indicate the presence of a common structure among the variables, and thus, further analysis may be conducted.

Table 1 presents the results of factor extraction using the Principal Axis Factoring method, along with the corresponding total explained variance and initial eigenvalues. According to the Guttman-Kaiser criterion, a total of six factors with eigenvalues greater than 1 were extracted, accounting for 64.87% of the total explained variance. The scree plot did not indicate a clear point of inflection. However, although considered simple and widely used, the scree test relies on identifying sharp drops between eigenvalues and may sometimes suggest more than one point of inflection. Furthermore, the reliability of interpreting the scree plot is regarded as low. Additionally, the Guttman-Kaiser criterion overlooks sampling variability and is prone to overestimating the number of factors (Fabrigar et al., 1999). Consequently, the number of factors retained for further analysis was determined using the parallel analysis criterion, with Principal Axis Factoring as the extraction method. The parallel analysis revealed that the eigenvalues of the actual data exceeded those of the 95th percentile of randomly generated data for five factors, as shown in Figure 2. Accordingly, a five-factor solution was retained for further analysis.

total variance explained						
factor	initial eigenvalues			extracted sums of squared loadings		
	total	% of variance	cumulative %	total	% of variance	cumulative %
1	17.28	41.14	41.14	16.90	40.23	40.23
2	4.45	10.59	51.74	4.04	9.63	49.85
3	1.88	4.47	56.21	1.54	3.66	53.51
4	1.44	3.42	59.62	1.03	2.46	55.97
5	1.16	2.77	62.39	0.75	1.80	57.77
6	1.04	2.48	64.87	0.62	1.47	59.23

Table 1. Total explained variance and initial eigenvalues obtained using the Principal Axis Factoring Extraction method.

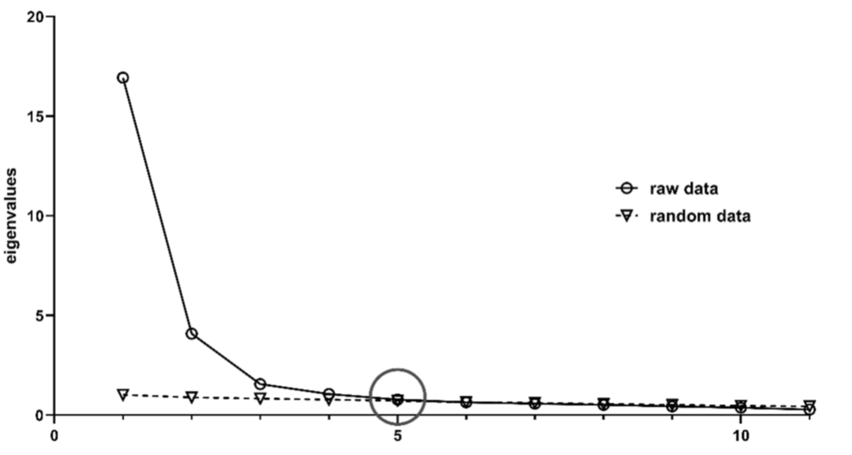

Figure 2. Cattell's Scree plot based on the results of the Parallel Analysis.

Table 2 presents the results of the EFA after Promax rotation. All retained items had communalities above 0.30. However, two items, TPK4 and TPK5, had factor loadings below 0.4 and were removed. Additionally, three items PCK2, CK2, and TCK6 were excluded due to cross-loadings on two or more factors (>0.3). As a result, a five-factor solution with 37 items was obtained, explaining a total of 64.22% of the variance. The first factor consisted of 12 items (explaining 41.57% of the variance) and was labelled PK. The second factor consisted of 7 items (11.01% of the variance) and was labelled TK. The third factor comprises eight items (4.94% of the variance) and was labelled TPACK. The fourth factor includes five items (3.70% of the variance) and was labelled TCK. Finally, the fifth factor consisted of 5 items (3% of the variance) and was labelled CK. Cronbach's alpha (α) and McDonald's omega (ω) coefficients indicated very good internal consistency reliability for the first four factors (ranging from 0.91 to 0.93). Slightly lower values were recorded for the CK factor (0.80).

item	factor				
	PK	TK	TPACK	TCK	CK
PK1	0.83				
PK5	0.74				
PK2	0.74				
PK7	0.68				
PCK5	0.65				
PK4	0.65				

PCK4	0.63				
PCK6	0.62				
PK3	0.60				
PK6	0.51				
PCK1	0.46				
PCK3	0.45				
TK7		0.96			
TK4		0.94			
TK5		0.93			
TK6		0.82			
TK2		0.75			
TK3		0.62			
TK1		0.58			
TPACK3			1.01		
TPACK2			0.88		
TPACK4			0.81		
TPACK1			0.72		
TPK1			0.61		
TPK2			0.54		
TPACK5			0.54		
TPK3			0.40		
TCK2				0.85	
TCK3				0.70	
TCK1				0.69	
TCK4				0.64	
TCK5				0.56	
CK6					0.72
CK5					0.59
CK4					0.58
CK3					0.56
CK1					0.47
α	0.91	0.93	0.93	0.91	0.80
ω	0.91	0.93	0.93	0.91	0.80

Legend: α – Cronbach's alpha; ω – McDonald's omega coefficient.

Table 2. Pattern coefficients and internal consistency reliability of subscales in the final model obtained through EFA analysis with promax rotation.

Table 3 shows the correlations between the factors obtained through Promax rotation. None of the correlations exceeded 0.7, which is favourable, as high correlations would suggest a strong linear relationship that could ultimately hinder the identification of distinct factors.

correlation matrix					
factor	**PK**	**TK**	**TPACK**	**TCK**	**CK**
PK	-				
TK	0.39	-			
TPACK	0.61	0.69	-		
TCK	0.37	0.61	0.70	-	
CK	0.62	0.40	0.60	0.47	-

Table 3. Interfactor correlation matrix of TPACK dimensions

3.3. Construct Validity – Confirmatory Factor Analysis (CFA)

Descriptive statistical parameters for all items retained after the EFA were examined prior to conducting the CFA. The skewness coefficients ranged from -3 to $+3$, and the kurtosis coefficients ranged from -7 to $+7$, indicating univariate normality of the distribution (see Appendix B). Regarding multivariate outliers, no participant exceeded the critical value in the Mahalanobis distance analysis. Thus, there were no multivariate outliers, and all participants were retained for further analysis. An inspection of the factor correlation matrix showed that all coefficients were below 0.9 (ranging from 0.51 to 0.84), indicating that multicollinearity was not a concern.

Table 4 presents the standardized factor loadings in the CFA. It is evident that all items have loadings above 0.5.

item	**factor**				
	1	**2**	**3**	**4**	**5**
PK1	0.68				
PK5	0.72				
PK2	0.68				
PK7	0.68				
PCK5	0.69				
PK4	0.72				
PCK4	0.74				
PCK6	0.73				
PK3	0.69				
PK6	0.57				
PCK1	0.75				
PCK3	0.73				
TK7		0.88			
TK4		0.85			
TK5		0.82			
TK6		0.82			
TK2		0.68			
TK3		0.81			
TK1		0.80			
TPACK3			0.87		
TPACK2			0.84		
TPACK4			0.88		
TPACK1			0.84		
TPK1			0.82		
TPK2			0.83		
TPACK5			0.80		
TPK3			0.83		
TCK2				0.77	
TCK3				0.79	
TCK1				0.71	
TCK4				0.82	
TCK5				0.84	
CK6					0.62
CK5					0.78
CK4					0.68
CK3					0.73
CK1					0.61

Table 4. Standardized Factor Loadings in CFA.

Table 5 presents the fit indices and χ^2 values for the five-factor model. The table includes three models. The first, original model shows a χ^2/df value below 3; however, the RMSEA, TLI, and CFI values are not satisfactory. Based on the analysis of modification indices, covariances were added between the items PK4 and PK5, TPACK2 and TPACK3, and TPK1 and TPK2. The model with added covariances shows acceptable values for nearly all parameters except for SRMR (> 0.05). Further analysis of the standardized residual covariances led to the exclusion of items CK1, TPACK5, PCK5, and PK6 from the analysis. As a result, a final five-factor model was obtained with all fit indices at satisfactory levels. The final five-factor model includes a total of 33 items (see Appendix C and Figure 3). Additionally, convergent and discriminant validity were assessed using reliability coefficients for the five-factor model.

Model	**$\chi2$**	**df**	**$\chi2$/df ≤ 3**	**RMSEA**			**SRMR**	**TLI**	**CFI**
				≤ **0.06**	**LO90**	**HI 90**	≤ **0.05**	≥ **0.90**	≥ **0.90**
original	1438.76	619	2.324	0.067	0.062	0.071	0.054	0.898	0.898
added covariances	1181.16	615	1.921	0.056	0.051	0.061	0.052	0.924	0.930
excluded CK1, TPACK5, PCK5, PK6	935.328	482	1.941	0.056	0.051	0.061	0.048	0.931	0.937

Legend: χ^2 – Chi-square; df – Degrees of freedom; RMSEA – Root Mean Square Error of Approximation; LO 90 – Lower bound of the 90% confidence interval for RMSEA; HI 90 – Upper bound of the 90% confidence interval for RMSEA; SRMR – Standardized Root Mean Square Residual; TLI – Tucker–Lewis Index; CFI – Comparative Fit Index

Table 5. Fit indices and χ^2 values for the five-factor model.

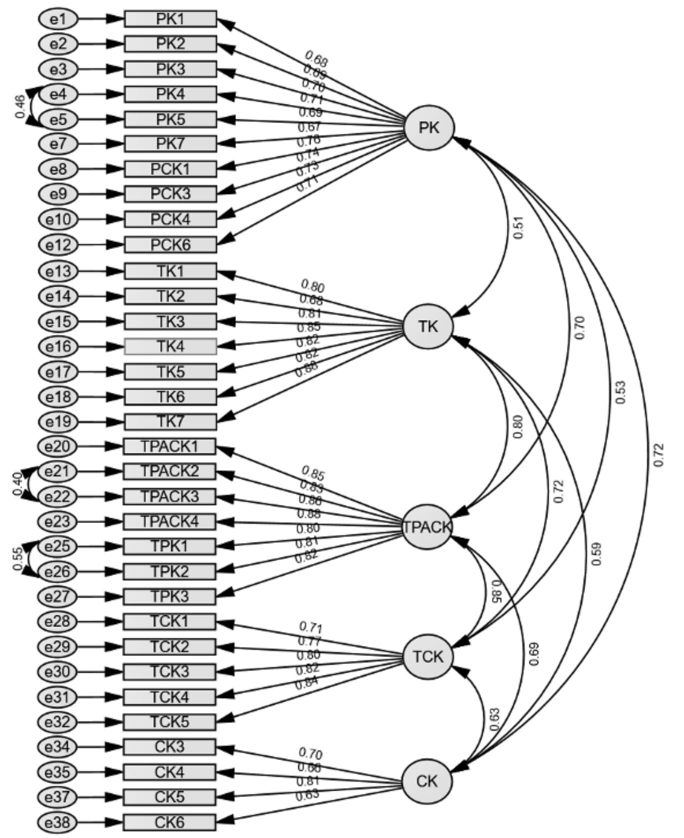

Figure 3. Confirmed five-factor model with 33 items based on CFA

3.4. Convergent and Discriminant Validity

Table 6 presents the results of the convergent and discriminant validity analyses along with the corresponding values. The average variance extracted (AVE) values for all factors were above 0.5, except for CK (0.49). Composite reliability (CR) values for each factor exceeded 0.7 and were greater than the respective AVE values. In addition, all Heterotrait–Monotrait ratio (HTMT) values were below the reference threshold of 0.85. To further support discriminant validity, MSV and ASV were also calculated. For four out of the five factors (PK, TK, TPACK, and CK), the AVE exceeded both MSV and ASV, supporting adequate discriminant validity. Although the AVE for the TCK factor was equal to its MSV, and its square root was slightly lower than the correlation with TPACK, this overlap is minor and theoretically expected. TCK and TPACK are closely related within the TPACK framework, where TCK is a core component of TPACK, which combines technological, pedagogical, and content knowledge. Importantly, the HTMT value between TCK and TPACK was well below the conservative 0.85 threshold, which supports acceptable discriminant validity. Additionally, all ASV values remained lower than their respective AVEs, further confirming discriminant validity across the model. Finally, the values of both internal consistency reliability coefficients were above 0.7, which is considered satisfactory.

factor	CR	AVE	MSV	ASV	α	ω	PK	TK	TPACK	TCK	CK
PK	0.91	0.50	0.48	0.36	0.91	0.89	**0.71**	0.46	0.64	0.47	0.60
TK	0.93	0.66	0.64	0.42	0.93	0.93	0.50	**0.81**	0.76	0.66	0.50
TPACK	0.94	0.71	0.64	0.55	0.95	0.95	0.69	0.80	**0.84**	0.77	0.60
TCK	0.89	0.62	0.62	0.40	0.89	0.89	0.49	0.66	0.79	**0.78**	0.55
CK	0.79	0.49	0.48	0.39	0.78	0.78	0.69	0.59	0.68	0.53	**0.70**

Legend: CR – Composite Reliability; AVE – Average Variance Extracted; MSV - maximum shared variance; ASV - average shared variance; square root of AVE on the diagonal (in bold); α – Cronbach's alpha; ω – McDonald's omega coefficient. The values above the diagonal are values from the heterotrait–monotrait (HTMT) validity analysis, while the values below the diagonal represent latent construct correlations (Fornell–Larcker matrix).

Table 6. Convergent and discriminant validity with reliability coefficients for the five-factor model.

3.5. Correlations between TPACK Model Constructs

Descriptive statistical parameters for the final model of the TPACK questionnaire are presented in Table 7. The results of the mean values show that respondents scored highest in PK and lowest in TCK. Correlations between the TPACK constructs are shown in Table 8. It is evident that statistically significant correlations exist between all constructs ($p < 0.01$).

construct	AS	(SD)	min.	max.	skew	kurt
PK	4.52	(0.42)	2.7	5	-0.60	-0.38
TK	4.12	(0.66)	1.86	5	-0.47	-0.35
TPACK	4.27	(0.55)	2.57	5	-0.36	-0.37
TCK	4.00	(0.66)	2.2	5	-0.08	-0.74
CK	4.47	(0.47)	2.75	5	-0.78	0.19

Legenda: M – Mean; SD – Standard Deviation; Min – Minimum score; Max – Maximum score; Skew – Skewness coefficient; Kurt – Kurtosis coefficient.

Table 7. Descriptive statistical parameters of TPACK questionnaire constructs ($n = 609$).

construct	PK	TK	TPACK	TCK	CK
PK	-				
TK	0.42**	-			
TPACK	0.63**	0.72**	-		
TCK	0.46**	0.64**	0.75**	-	
CK	0.61**	0.44**	0.58**	0.59**	-

Legend: ** – significant correlations at $p < 0.05$

Table 8. Pearson Product-Moment correlation between TPACK constructs.

4. Discussion

4.1. Factor analysis of the TPACK questionnaire

This study aimed to validate the TPACK questionnaire with a sample of in-service subject teachers within the Croatian educational system to confirm its factor structure and examine the reliability and validity of the applied measurement instrument. Unlike the findings of Dobi Barišić et al. (2019), who validated the SPTTK questionnaire (Schmidt et al., 2009) on a sample of Croatian pre-service teachers and early childhood educators and identified a nine-factor structure, this study, using the questionnaire developed by Schmid et al. (2020), which largely includes items from the original SPTTK, revealed a five-factor structure. These differences in factor structure can likely be attributed to the use of different instruments and samples (i.e., pre-service teachers and preschool teachers versus in-service subject teachers).

The merging of PK and PCK and TPK and TPACK constructs suggests a lack of clear distinction between these dimensions in how the items were formulated. Croatian subject teachers do not differentiate their general pedagogical knowledge (PK) from pedagogical content knowledge (PCK), i.e., knowledge specific to teaching particular subject matter. Similarly, they do not distinguish between TPK (knowledge of how to use technological tools to enhance teaching and learning) and TPACK (knowledge of how to integrate technology in teaching specific subject content effectively and constructively). This reflects how Croatian subject teachers perceive and apply pedagogical and technological pedagogical knowledge in an integrated and context-specific manner. As Croatian subject teachers are mainly trained within strong disciplinary context (e.g., science faculties, humanities faculties…), their perception of teaching knowledge may not clearly separate general pedagogical principles (PK) from pedagogical content knowledge (PCK) and general pedagogical technological principles (TPK) from its application in subject-specific context (TPACK). These results can be interpreted through cultural and systemic factors in Croatian education. In teaching subject methodologies, which are integral part of teacher education, integrated knowledge application is emphasized more than strict conceptual boundaries. Additionally, the findings may indicate theoretical limitations and ambiguities within the original TPACK framework, issues that have been highlighted by several scholars (Angeli & Valanides, 2009; Bos, 2011; Cox & Graham, 2009; Graham, 2011; Kimmons, 2015; Ritzhaupt et al., 2016). The lack of distinction between PK and PCK, as well as between TPK and TPACK, was also observed by Ritzhaupt et al. (2016) who validated Schmidt et al.'s (Schmidt et al., 2009) questionnaire on a sample of American pre-service teachers.

The merging of PK and PCK can also be interpreted within the context of limitations in Shulman's original PCK model (1986), upon which the TPACK framework is built. The boundaries between PK and PCK have been conceptualized differently over time and remain somewhat unclear (Gess-Newsome, 1999; Gess-Newsome et al., 2017; Lee & Luft, 2008). Difficulties distinguishing PK and PCK were also found in a study by Lee & Tsai (2010), conducted on a sample of Taiwanese primary and secondary school teachers using the TPAC-W questionnaire, which measured teachers' TPACK self-assessments in the context of web technologies. The merging of PK and PCK was similarly observed in studies involving pre-service teachers in the U.S. (Shinas et al., 2013), Singapore (Koh et al., 2010), Turkey (Semiz & Ince, 2012), and among 542 EFL teachers from nine countries (Bostancıoğlu & Handley, 2018). Zelkowski et al. (2013) encountered issues with the PCK construct in their exploratory factor analysis of a TPACK questionnaire among future mathematics teachers in the U.S. This construct, along with TPK and TCK, was entirely removed from the analysis due to cross-loading across different factors.

In the present study, validation of the questionnaire in the Croatian context revealed a clearly defined content knowledge (CK) factor despite removing several original CK items. This reflects the way subject teachers are educated at faculties focused on specific disciplines (e.g., language teachers at faculties of humanities, physics teachers at faculties of science, physical education teachers at faculties of kinesiology). In recent years some subject teachers have also been educated at teacher education faculties (classroom teachers with one subject specialization). In contrast to this finding, CK was combined with PK and PCK in some previous studies, such as those involving American online teachers (Archambault & Barnett, 2010) and pre-service teachers in Estonia (Luik et al., 2018). The persistence of a distinct CK factor in our study, contrasting with its merging with PK and PCK in U.S. and Estonian contexts may reflect systemic differences in teacher training priorities. Croatia's emphasis on traditional subject orientation could reinforce clear boundaries between content and pedagogical knowledge, whereas systems in the U.S. and Estonia that promote integrated and interdisciplinary teaching roles (Eisenschmidt et al., 2023; Mansilla & Lenoir, 2010) may blur these distinctions. Furthermore, the differences between Croatian and American results can be attributed to the contrasting educational traditions: Croatia aligns with the European, predominantly Germanic tradition of didactics, while the United States follows the Anglo-American tradition of curriculum and instruction (Matanović, 2017). Croatia was strongly influenced by the Germanic tradition that emphasizes professional subject competencies. On the other hand, the American educational paradigm is criticized for neglecting factual knowledge (Kansanen, 2002).

The technological knowledge (TK) construct also identified a clear distinction. Croatian teachers differentiate their knowledge of using specific digital technologies at the user level. In the Croatian educational system, toward the end of the 20th and the beginning of the 21st century, alongside the growing development of computer technologies, educators in many Croatian schools were offered opportunities for training and to obtain the European Computer Driving Licence (ECDL), an internationally recognized certification of digital literacy. Additionaly, significant contributions to the development of digital competences of teachers in Croatia were made by the e-School projects, implemented in two phases from 2018 to 2023 (CARNET, 2019), and the School for Life curricular reform implemented from 2018 to 2020. The e-School project facilitated comprehensive and systematic education in the application of digital tools and the integration of ICT in teaching, resulting in an increase in digital literacy and readiness for innovative pedagogical practices. The curricular reform further emphasized the importance of applying digital technologies through changes in teaching content and methods, encouraging the development of competences focused on active and creative learning (Ministry of Science and Education, 2019a). These contemporary efforts are supported by the research findings of this study, which indicates the strengthening of teacher knowledge dimensions, but also highlight challenges in their clear differentiation, confirming the need for continuous professional development and support within the educational system. Some earlier international studies indicate the need for further professional development of Croatian teachers. For example, TALIS (*Teaching and Learning International Survey*) results show that Croatian teachers express a need for further training in integrating technology into teaching (OECD, 2019), while DESI (Digital Economy and Society Index) reports indicate that Croatia, despite progress in basic digital skills, is still lagging in systematic digital transformation and development of ICT capacities in education (European Commission, 2023). These findings may partly explain the merging of the TPK and TPACK dimensions in the validation of the TPACK questionnaire. Although certain positive developments have been made in teacher education, it seems that some shortcomings in systematic education detected by TALIS and DESI research may make it difficult for teachers to clearly distinguish general technological pedagogical knowledge from its appropriate integration into subject content. The extraction of the TCK (Technological Content Knowledge) construct in this study is in contrast with some previous research. In several studies that demonstrated a departure from the original seven-factor structure, TCK was merged during factor analysis with other technology-related dimensions, for example, with the TPK (Liang et al., 2013), TPACK (Shinas et al., 2013), TPK and TPACK (Archambault & Barnett, 2010; Karadeniz & Vatanartıran, 2013; Koh et al., 2010; Liu et al., 2015), and with TK, TPK, and TPACK dimensions (Luik et al., 2018). This merging of TPK items with TPACK was also observed in studies by Koh et al. (2010), Karadeniz and Vatanartıran (2013), and Shinas et al. (2013). Research indicating the merging of technological dimensions may reflect highly integrated curricula that treat technology as an integral part of all subjects rather than a separate competency. The findings of this study, which indicate the extraction of the TCK construct and the merging of items from the TPK and TPACK subscales, are most similar to those reported by Lee and Tsai (2010).

In summary, the results of the TPACK questionnaire validation, which deviated from the original seven-factor theoretical model, are consistent with numerous international studies that identified either more or fewer factors than originally proposed. This suggests the presence of certain cultural and contextual differences in teachers' self-assessment of knowledge. Furthermore, the difficulties in distinguishing PCK from PK, as well as TK from other technology-related dimensions, observed in this study, are in line with findings from earlier studies conducted in American (Archambault & Barnett, 2010; Shinas et al., 2013; Zelkowski et al., 2013), Estonian (Luik et al., 2018), Singaporean (Koh et al., 2010), Taiwanese (Lee & Tsai, 2010) and

Turkish contexts (Karadeniz & Vatanartıran, 2013; Semiz & Ince, 2012). Given the subject-specific nature of TPACK knowledge, it is possible that the difficulty in confirming the original seven-factor model partially stems from the lack of attention to subject-area specificity. That is, the use of generically formulated questionnaire items, applied across samples of teachers from various disciplines, may not offer respondents a sufficient basis for clearly distinguishing between domains of knowledge.

4.2. Correlation analysis between TPACK constructs

All correlations between the TPACK knowledge dimensions were positive and statistically significant, consistent with the theoretical framework of the model. Weak correlations between the core knowledge dimensions PK and TK, as well as CK and TK suggest that these domains serve distinct roles and focus on different aspects of the teaching context. PK and CK showed a moderate correlation with each other (higher than with TK), reflecting the traditional interconnection between pedagogy and content during teacher education and professional development. In contrast, technology is perceived as a new variable and a relatively separate component. These results align with the findings of previous studies conducted on in-service teachers (Dong et al., 2015; Koh et al., 2014; Mohammad-Salehi & Vaez-Dalili, 2022), as well as on pre-service teacher education students (Koh & Chai, 2011; Valtonen et al., 2017). The weak correlations between PK and TK, and between CK and TK, are further supported by other research (Koh & Chai, 2011; Roig et al., 2015; Valtonen et al., 2017) which reported small to negligible associations between these constructs. In contrast, Soto and Herrera (2022) and Sofyan et al. (2023) emphasize a stronger correlation between CK and PK. It can be observed that correlations within the technological dimensions are generally higher than their correlations with the non-technological knowledge dimensions, except for TK and TCK, which showed a moderate correlation. This finding is consistent with the findings of Koh, Chai, and Tsai (2014) and Mohammad-Salehi and Vaez-Dalili (2022). However, previous research results in this area suggest a certain degree of variability, as some studies report a strong correlation between these two dimensions of knowledge (Soto & Herrera, 2022; Roig et al., 2015; Sofyan et al., 2023).

The relationship between the TCK and TPACK dimensions appears to be clearer, as many studies, including the present one, have found a strong correlation between them (Chai, Chin, et al., 2013; Dobi Barišić et al., 2019; Koh et al., 2014; Mohammad-Salehi & Vaez-Dalili, 2022; Soto & Herrera, 2022; Roig et al., 2015; Sofyan et al., 2023). These findings align with the transformative perspective on the TPACK framework, which suggests that hybrid dimensions have a greater influence on integrated teacher knowledge than the core components. The high correlation may stem from the fact that integrating technology into teaching requires knowledge of content-specific tools and that teachers with well-developed TCK are more motivated to experiment with technology in the classroom. This, in turn, can enhance their TPACK through reflection and experience. Although correlations are generally consistent with prior studies, inconsistencies in the strength of relationships among certain dimensions highlight the need for further research into contextual factors.

4.3. Recommendations and limitations

The results of the TPACK questionnaire validation highlight the challenges in developing a universal instrument capable of assessing all seven domains of knowledge defined by the TPACK model, applicable across diverse geographic and subject-specific contexts. Therefore, future research is encouraged to consider the previously discussed dimension of contextual knowledge. Specifically, the applied questionnaire did not include contextual knowledge, which Mishra (2019) later added to the original model. Future studies on teacher knowledge would benefit from incorporating this contextual dimension to gain a deeper understanding of the dynamics of teacher knowledge development, while also acknowledging the specific characteristics of the environment, students, curriculum, and resources. This, ultimately, may lead to more effective educational solutions. This study's main limitation lies in using self-assessment through the TPACK questionnaire. Self-assessments rely on participants’ personal judgments and perceptions, which may introduce subjectivity and bias. Teachers may overestimate or underestimate their knowledge. Therefore, future research should combine self-assessment data with observations of how TPACK dimensions are applied in educational practice, as well as instruments that objectively measure teacher knowledge.

5. Conclusion

The validation of the TPACK questionnaire confirms that the framework is a useful tool for understanding and assessing the competencies of Croatian in-service teachers related to technology integration in teaching. The findings support a five-factor structure, pointing to localized interpretations of teacher knowledge domains. This outcome highlights the importance of adapting the TPACK theoretical construct to reflect educational traditions and teacher education pathways within the Croatian context. By integrating recent reforms, professional development initiatives and relevant international indicators, this study situates its results within current educational policy. The validation of the TPACK questionnaire in the Croatian educational context has opened the door to further research on TPACK knowledge dimensions among in-service teachers and to the integration of this instrument with other measurement tools and research approaches.

Appendix A: Descriptive statistical parameters for all items of the applied questionnaire (n = 312).

item	M	(SD)	min.	max.	Skew	Kurt
PK1	4.60	(0.52)	3	5	-0.74	-0.73
PK2	4.45	(0.58)	3	5	-0.48	-0.70
PK3	4.48	(0.55)	3	5	-0.42	-0.89
PK4	4.50	(0.59)	3	5	-0.72	-0.44
PK5	4.51	(0.58)	3	5	-0.70	-0.49
PK6	4.37	(0.63)	2	5	-0.54	-0.22
PK7	4.52	(0.58)	3	5	-0.71	-0.48
CK1	4.71	(0.46)	4	5	-0.91	-1.17
CK2	4.66	(0.48)	3	5	-0.76	-1.20
CK3	4.79	(0.41)	4	5	-1.45	0.11
CK4	4.38	(0.70)	2	5	-0.86	0.13
CK5	4.53	(0.59)	2	5	-0.96	0.41
CK6	4.21	(0.73)	2	5	-0.59	-0.11
TK1	4.32	(0.70)	3	5	-0.54	-0.84
TK2	3.85	(1.00)	1	5	-0.44	-0.64
TK3	4.27	(0.70)	2	5	-0.47	-0.63
TK4	4.24	(0.74)	2	5	-0.51	-0.69
TK5	3.88	(0.83)	2	5	-0.28	-0.56
TK6	4.21	(0.77)	2	5	-0.51	-0.76
TK7	4.07	(0.81)	2	5	-0.41	-0.68
PCK1	4.42	(0.58)	3	5	-0.37	-0.75
PCK2	4.34	(0.62)	3	5	-0.37	-0.66
PCK3	4.50	(0.57)	3	5	-0.57	-0.69
PCK4	4.60	(0.55)	3	5	-0.95	-0.16
PCK5	4.67	(0.47)	4	5	-0.71	-1.51
PCK6	4.60	(0.52)	3	5	-0.76	-0.71
TPK1	4.25	(0.66)	3	5	-0.32	-0.76
TPK2	4.28	(0.64)	3	5	-0.31	-0.68
TPK3	4.30	(0.65)	2	5	-0.46	-0.37
TPK4	4.37	(0.65)	3	5	-0.56	-0.66
TPK5	3.94	(0.94)	1	5	-0.59	-0.11
TCK1	4.18	(0.75)	2	5	-0.49	-0.53
TCK2	3.93	(0.82)	2	5	-0.22	-0.77
TCK3	3.91	(0.78)	2	5	-0.13	-0.70
TCK4	3.90	(0.82)	2	5	-0.18	-0.77
TCK5	4.16	(0.73)	2	5	-0.36	-0.70
TCK6	4.39	(0.66)	2	5	-0.70	-0.25
TPACK1	4.16	(0.66)	2	5	-0.32	-0.23
TPACK2	4.36	(0.62)	3	5	-0.43	-0.66
TPACK3	4.31	(0.61)	3	5	-0.27	-0.63

TPACK4	4.23	(0.67)	2	5	-0.36	-0.53
TPACK5	3.95	(0.87)	2	5	-0.41	-0.59

Legend: M – Mean; SD – Standard Deviation; Min – Minimum score; Max – Maximum score; Skew – Skewness coefficient; Kurt – Kurtosis coefficient

Appendix B: Descriptive statistical parameters for all items of the applied questionnaire after EFA (n = 297).

item	M	(SD)	min.	max.	Skew	Kurt
PK1	4.57	(0.52)	3	5	-0.58	-1.04
PK2	4.47	(0.59)	2	5	-0.69	0.06
PK3	4.44	(0.61)	2	5	-0.80	0.53
PK4	4.52	(0.57)	3	5	-0.68	-0.56
PK5	4.49	(0.61)	3	5	-0.79	-0.37
PK6	4.33	(0.64)	3	5	-0.42	-0.69
PK7	4.54	(0.56)	3	5	-0.68	-0.60
CK1	4.67	(0.51)	3	5	-1.11	0.06
CK3	4.77	(0.42)	4	5	-1.27	-0.39
CK4	4.39	(0.61)	2	5	-0.55	-0.14
CK5	4.56	(0.57)	3	5	-0.86	-0.26
CK6	4.15	(0.70)	2	5	-0.34	-0.54
TK1	4.33	(0.70)	2	5	-0.73	0.00
TK2	3.86	(0.93)	1	5	-0.48	-0.27
TK3	4.25	(0.72)	2	5	-0.58	-0.29
TK4	4.25	(0.70)	2	5	-0.68	0.29
TK5	3.88	(0.85)	1	5	-0.49	0.00
TK6	4.21	(0.79)	2	5	-0.81	0.22
TK7	4.04	(0.81)	2	5	-0.43	-0.51
PCK1	4.41	(0.59)	3	5	-0.39	-0.71
PCK3	4.47	(0.60)	3	5	-0.66	-0.52
PCK4	4.57	(0.55)	2	5	-0.95	0.55
PCK5	4.70	(0.46)	3	5	-0.99	-0.74
PCK6	4.64	(0.51)	3	5	-0.91	-0.49
TPK1	4.29	(0.65)	2	5	-0.43	-0.37
TPK2	4.29	(0.66)	2	5	-0.45	-0.42
TPK3	4.36	(0.66)	2	5	-0.62	-0.35
TCK1	4.13	(0.76)	2	5	-0.45	-0.48
TCK2	3.86	(0.79)	2	5	-0.02	-0.86
TCK3	3.88	(0.77)	2	5	-0.15	-0.57
TCK4	3.85	(0.82)	2	5	-0.08	-0.83
TCK5	4.18	(0.70)	3	5	-0.26	-0.96
TPACK1	4.20	(0.67)	2	5	-0.39	-0.28
TPACK2	4.33	(0.62)	2	5	-0.43	-0.21
TPACK3	4.27	(0.64)	2	5	-0.54	0.45
TPACK4	4.22	(0.67)	2	5	-0.42	-0.22
TPACK5	3.98	(0.82)	1	5	-0.47	-0.12

Legend: M – Mean; SD – Standard Deviation; Min – Minimum score; Max – Maximum score; Skew – Skewness coefficient; Kurt – Kurtosis coefficient

Appendix C: Items in the final five-factor model

PK	
PK1	I can adapt my teaching based upon what students currently understand or do not understand.
PK2	I can adapt my teaching style to different learners.
PK3	I can use a wide range of teaching approaches in a classroom setting.
PK4	I can assess student learning in multiple ways.
PK5	I know how to assess student performance in a classroom.
PK7	I know how to organize and maintain classroom management
PCK1	I know how to select effective teaching approaches to guide student thinking and learning in my teaching subject.
PCK3	I know how to develop exercises with which students can consolidate their knowledge of my teaching subject.
PCK4	I know how to evaluate students' performance in my teaching subject.
PCK6	In my teaching subject I can identify from student errors where there are difficulties in understanding and give appropriate feedback.
CK	
CK3	I know the basic theories and concepts of my teaching subject.
CK4	I know the history and development of important theories in my teaching subject.
CK5	I have various ways and strategies of developing my understanding of my teaching subject.
CK6	I am familiar with recent research in my teaching subject.
TK	
TK1	I keep up with important new technologies.
TK2	I frequently play around with the technology.
TK3	I know about a lot of different technologies.
TK4	I have the technical skills I need to use technology.
TK5	I know how to solve my own technical problems.
TK6	I can learn technology easily.
TK7	I have had sufficient opportunities to work with different technologies.
TCK	
TCK1	I know how technological developments have changed the field of my subject.
TCK2	I can explain which technologies have been used in research in my field.
TCK3	I know which new technologies are currently being developed in the field of my subject.
TCK4	I know how to use technologies to participate in scientific discourse in my field.
TCK5	I know technologies which help me understand my subject.
TPACK	
TPACK1	I can use strategies that combine content, technologies, and teaching approaches.
TPACK2	I can choose technologies that enhance the content for a lesson.
TPACK3	I can select technologies to use in my classroom that enhance what I teach, how I teach, and what students learn.
TPACK4	I can teach lessons that appropriately combine my teaching subject, technologies, and teaching approaches.

TPK1	I can choose technologies that enhance the teaching approaches for a lesson.
TPK2	I can choose technologies that enhance students' learning for a lesson.
TPK3	I can adapt the use of the technologies that I am learning about to different teaching activities.

References

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT–TPCK: Advances in technological pedagogical content knowledge (TPCK). *Computers & Education*, *52*(1), 154–168. https://doi.org/10.1016/j.compedu.2008.07.006

Archambault, L. M., & Barnett, J. H. (2010). Revisiting technological pedagogical content knowledge: Exploring the TPACK framework. *Computers & Education*, *55*(4), 1656–1662. https://doi.org/10.1016/j.compedu.2010.07.009

Baser, D., Kopcha, T. J., & Ozden, M. Y. (2016). Developing a technological pedagogical content knowledge (TPACK) assessment for preservice teachers learning to teach English as a foreign language. *Computer Assisted Language Learning*, *29*(4), 749–764. https://doi.org/10.1080/09588221.2015.1047456

Bos, B. (2011). Professional Development for Elementary Teachers Using TPACK. *Contemporary Issues in Technology and Teacher Education*, *11*(2), 167–183.

Bostancıoğlu, A., & Handley, Z. (2018). Developing and validating a questionnaire for evaluating the EFL 'Total PACKage': Technological Pedagogical Content Knowledge (TPACK) for English as a Foreign Language (EFL). *Computer Assisted Language Learning*, *31*(5–6), 572–598. https://doi.org/10.1080/09588221.2017.1422524

Canbazoğlu Bilici, S., Yamak, H., Selcen Guzey, S., & Bilici, C. (2013). Technological Pedagogical Content Knowledge Self-Efficacy Scale (TPACK-SeS) for Pre-Service Science Teachers: Construction, Validation, and Reliability. *Eurasian Journal of Educational Research*, *52*, 37–60.

CARNET. (2019). *e-Schools: Establishing a System for Developing Digitally Mature Schools (pilot project)* . https://www.carnet.hr/en/projekt/e-skole-establishing-a-system-for-developing-digitally-mature-schools-pilot-project/

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/S15327906MBR0102_10

Chai, C. S., Chin, C. K., Koh, J. H. L., & Tan, C. L. (2013). Exploring Singaporean Chinese Language Teachers' Technological Pedagogical Content Knowledge and its Relationship to the Teachers' Pedagogical Beliefs. *The Asia-Pacific Education Researcher*, *22*(4), 657–666. https://doi.org/10.1007/s40299-013-0071-3

Chai, C. S., Koh, J. H. L., & Tsai, C. C. (2016). A Review of the Quantitative Measures of Technological Pedagogical Content Knowledge (TPACK). In M. C. Herring, M. J. Koehler, & P. Mishra (Eds.), *Handbook of Technological Pedagogical Content Knowledge (TPACK) for Educators* (2nd ed.). Routledge. https://doi.org/10.4324/9781315771328

Chai, C. S., Koh, J. H. L., & Tsai, C.-C. (2010). Facilitating Preservice Teachers' Development of Technological, Pedagogical, and Content Knowledge (TPACK) on JSTOR. *Journal of Educational Technology & Society*, *13*(4), 63–73. https://www.jstor.org/stable/jeductechsoci.13.4.63

Chai, C. S., Koh, J. H. L., & Tsai, C.-C. (2011). Exploring the factor structure of the constructs of technological, pedagogical, content knowledge (TPACK). *Asia-Pacific Education Researcher*, *20*(3), 595–603. https://scholar.lib.ntnu.edu.tw/en/publications/exploring-the-factor-structure-of-the-constructs-of-technological-2

Chai, C. S., Ng, E. M., Li, W., Hong, H.-Y., & Koh, J. H. L. (2013). Validating and modelling technological pedagogical content knowledge framework among Asian preservice teachers. *Australasian Journal of Educational Technology*, *29*(1). https://doi.org/10.14742/ajet.174

Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education*. Routledge. https://doi.org/10.4324/9780203029053

Cox, S., & Graham, C. R. (2009). Diagramming TPACK in Practice: Using an Elaborated Model of the TPACK Framework to Analyze and Depict Teacher Knowledge. *TechTrends*, *53*(5), 60–69. https://doi.org/10.1007/s11528-009-0327-1

Creswell, J. (2015). *Eduactional Research: Planning. Conducting. and Evaluating Quantitative and Qualitative Research* (5th ed.). Pearson.

Deng, F., Chai, C. S., So, H.-J., Qian, Y., & Chen, L. (2017). Examining the validity of the technological pedagogical content knowledge (TPACK) framework for preservice chemistry teachers. *Australasian Journal of Educational Technology*, *33*(3). https://doi.org/10.14742/ajet.3508

Dobi Barišić, K., & Brust Nemet, M. (2023). The impact of the teacher education study program on the development of TPACK. *Revija Za Elementarno Izobraževanje*, *16*(1), 79–92. https://doi.org/10.18690/rei.16.1.2064

Dobi Barišić, K., Divjak, B., & Kirinić, V. (2019). Education Systems as Contextual Factors in the Technological Pedagogical Content Knowledge Framework. *Journal of Information and Organizational Sciences*, *43*(2), 163–183. https://doi.org/10.31341/jios.43.2.3

Dong, Y., Chai, C. S., Sang, G.-Y., Koh, J. H. L., & Tsai, C.-C. (2015). Exploring the Profiles and Interplays of Pre-Service and In-Service Teachers' Technological Pedagogical Content Knowledge (TPACK) in China. *Educational Technology & Society*, *18*(1), 158–169.

Eisenschmidt, Eve., Heidmets, Mati., Kitsing, Maie., Kasesalk, Mikk., & Vanari, Katlin. (2023). Aim high and work hard: Building a world-class learning system in Estonia. In *National Center on Education and the Economy*.

European Commission. (2023). *The Digital Economy and Society Index (DESI) | Shaping Europe's digital future*. https://digital-strategy.ec.europa.eu/en/policies/desi

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). Sage Publications Ltd.

Gess-Newsome, J. (1999). Pedagogical Content Knowledge: An Introduction and Orientation. In J. Newsome & N. G. Lederman (Eds.), *Examining Pedagogical Content Knowledge* (Vol. 6). Springer. https://doi.org/10.1007/0-306-47217-1_1

Gess-Newsome, J., Taylor, J. A., Carlson, J., Gardner, A. L., Wilson, C. D., & Stuhlsatz, M. A. M. (2017). Teacher pedagogical content knowledge, practice, and student achievement. *International Journal of Science Education*, *41*(7), 944–963. https://doi.org/10.1080/09500693.2016.1265158

Graham, C. R. (2011). Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers & Education*, *57*(3), 1953–1960. https://doi.org/10.1016/j.compedu.2011.04.010

Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Upper Saddle River; Prentice Hall.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, *6*(1), pp53-60-pp53-60. https://academic-publishing.org/index.php/ejbrm/article/view/1224

Horn, J. L. (1965). A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Jang, S.-J., & Tsai, M.-F. (2012). Exploring the TPACK of Taiwanese elementary mathematics and science teachers with respect to use of interactive whiteboards. *Computers & Education*, *59*(2), 327–338. https://doi.org/10.1016/j.compedu.2012.02.003

Kabakci Yurdakul, I., Odabasi, H. F., Kilicer, K., Coklar, A. N., Birinci, G., & Kurt, A. A. (2012). The development, validity and reliability of TPACK-deep: A technological pedagogical content knowledge scale. *Computers & Education*, *58*(3), 964–977. https://doi.org/10.1016/j.compedu.2011.10.012

Kansanen, P. (2002). Didactics and its relation to educational psychology: Problems in translating a key concept across research communities. *International Review of Education*, *48*(6), 427–441. https://doi.org/10.1023/A:1021388816547

Karadeniz, Ş., & Vatanartıran, S. (2013). Adaptation of a TPACK survey to Turkish for secondary school teachers. *Journal of New Results in Ccience*, *10*(2), 34–47.

Kimmons, R. (2015). Examining TPACK's Theoretical Future. *Journal of Technology and Teacher Education*, *23*(1), 53–77.

Koehler, M. J., Shin, T. S., & Mishra, P. (2012). How Do We Measure TPACK? Let Me Count the Ways. In *Educational Technology, Teacher Knowledge, and Classroom Impact* (pp. 16–31). IGI Global. https://doi.org/10.4018/978-1-60960-750-0.ch002

Koehler, M. J. & Mishra, P. (2009). What Is Technological Pedagogical Content Knowledge? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70. https://citejournal.org/volume-9/issue-1-09/general/what-is-technological-pedagogicalcontent-knowledge/

Koh, J. H. L., & Chai, C. S. (2011). Modeling pre-service teachers' technological pedagogical content knowledge (TPACK) perceptions. *ASCILITE Publications*, 735–746. https://doi.org/10.14742/apubs.2011.1741

Koh, J. H. L., Chai, C. S., & Tsai, C. C. (2010). Examining the technological pedagogical content knowledge of Singapore pre-service teachers with a large-scale survey. *Journal of Computer Assisted Learning*, *26*(6), 563–573. https://doi.org/10.1111/J.1365-2729.2010.00372.X

Koh, J. H. L., Chai, C. S., & Tsai, C.-C. (2014). Demographic Factors, TPACK Constructs, and Teachers' Perceptions of Constructivist-Oriented TPACK. *Educational Technology and Society*, *17*(1), 185–196. https://doi.org/10.2307/JEDUCTECHSOCI.17.1.185.PDF

Lee, E., & Luft, J. A. (2008). Experienced secondary science teachers' representation of pedagogical content knowledge. *International Journal of Science Education*, *30*(10), 1343–1363. https://doi.org/10.1080/09500690802187058

Lee, M. H., & Tsai, C. (2010). Exploring teachers' perceived self efficacy and technological pedagogical content knowledge with respect to educational use of the World wide Web. *Instructional Science*, *38*(1), 1–21. https://doi.org/10.1007/S11251-008-9075-4

Liang, J.-C., Chai, C. S., Koh, J. H. L., Yang, C.-J., & Tsai, C.-C. (2013). Surveying in-service preschool teachers' technological pedagogical content knowledge. *Australasian Journal of Educational Technology*, *29*(4). https://doi.org/10.14742/ajet.299

Lin, T.-C., Tsai, C.-C., Chai, C. S., & Lee, M.-H. (2013). Identifying Science Teachers' Perceptions of Technological Pedagogical and Content Knowledge (TPACK). *Journal of Science Education and Technology*, *22*(3), 325–336. https://doi.org/10.1007/s10956-012-9396-6

Liu, Q., Zhang, S., & Wang, Q. (2015). Surveying Chinese In-Service K12 Teachers' Technology, Pedagogy, and Content Knowledge. *Journal of Educational Computing Research*, *53*(1), 55–74. https://doi.org/10.1177/0735633115585929

Luik, P., Taimalu, M., & Suviste, R. (2018). Perceptions of technological, pedagogical and content knowledge (TPACK) among pre-service teachers in Estonia. *Education and Information Technologies*, *23*(2), 741–755. https://doi.org/10.1007/s10639-017-9633-y

Mansilla, V. B., & Lenoir, Y. (2010). Interdisciplinarity in UnIted states schools: Past, Present, and Future. *Issues in Integrative Studies*, *28*, 1–27.

Matanović, I. (2017). Kurikularna reforma kao oblik standardizacije obrazovanja – pedagoško-didaktička refleksija [Curriculum reform as a form of education standardization – a pedagogical-didactic reflection]. *Život i Škola: Časopis Za Teoriju i Praksu Odgoja i Obrazovanja*, *63*(1), 13–27.

McDonald, R. P. (1999). Test theory: A unified treatment. In *Test Theory: A Unified Treatment* (1st ed.). Psychology Press. https://doi.org/10.4324/9781410601087/TEST-THEORY-RODERICK-MCDONALD/RIGHTS-AND-PERMISSIONS

Menon, V., & Muraleedharan, A. (2020). Internet-based surveys: relevance, methodological considerations and troubleshooting strategies. *General Psychiatry*, *33*(5), e100264. https://doi.org/10.1136/gpsych-2020-100264

Ministry of Science and Education. (2019a). *Kurikulum međupredmetne teme Uporaba informacijske i komunikacijske tehnologije za osnovne i srednje škole*. https://mzom.gov.hr/istaknute-teme/odgoj-i-obrazovanje/nacionalni-kurikulum/medjupredmetne-teme/kurikulum-medjupredmetne-teme-uporaba-informacijske-i-komunikacijske-tehnologije-za-osnovne-i-srednje-skole/3854

Ministry of Science and Education. (2019b). *Kurikulumi nastavnih predmeta i međupredmetnih tema.* https://skolazazivot.hr/kurikulumi-2/

Mishra, P. (2019). Considering Contextual Knowledge: The TPACK Diagram Gets an Upgrade. *Journal of Digital Learning in Teacher Education*, *35*(2), 76–78. https://doi.org/10.1080/21532974.2019.1588611

Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record: The Voice of Scholarship in Education*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mohammad-Salehi, B., & Vaez-Dalili, M. (2022). Examining EFL Teachers' Perceptions of Technological Pedagogical Content Knowledge and Web 2.0 Technologies Using a Structural Equation Modeling Technique. *Journal of Modern Research in English Language Studies*, *9*(2), 51–76. https://doi.org/10.30479/JMRELS.2021.14550.1779

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers*, *32*(3), 396–402. https://doi.org/10.3758/BF03200807/METRICS

OECD. (2019). *TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners.* (TALIS). OECD. https://doi.org/10.1787/1D0BC92A-EN

Osnovne i srednje škole | Državni zavod za statistiku. Retrieved April 1, 2025, from https://podaci.dzs.hr/hr/podaci/obrazovanje/osnovne-i-srednje-skole/

Paidican Soto, M. A., & Arredondo Herrera, P. A. (2022). Techno-pedagogical and disciplinary knowledge of primary school teachers and demographic factors . *Mendive. Revista de Educación*, *20*(3).

Pamuk, S., Ergun, M., Cakir, R., Yilmaz, H. B., & Ayas, C. (2015). Exploring relationships among TPACK components and development of the TPACK instrument. *Education and Information Technologies*, *20*(2), 241–263. https://doi.org/10.1007/s10639-013-9278-4

Petko, D., Mishra, P., & Koehler, M. J. (2025). TPACK in context: An updated model. *Computers and Education Open*, *8*, 100244. https://doi.org/10.1016/j.caeo.2025.100244

Ritzhaupt, A. D., Huggins-Manley, A. C., Ruggles, K., & Wilson, M. (2016). Validation of the Survey of Pre-service Teachers' Knowledge of Teaching and Technology: A Multi-Institutional Sample. *Journal of Digital Learning in Teacher Education*, *32*(1), 26–37. https://doi.org/10.1080/21532974.2015.1099481

Roig, R., Mengual, S., & Quinto, P. (2015). Primary Teachers' Technological, Pedagogical and Content Knowledge. *Comunicar*, *23*(45), 151–159. https://www.revistacomunicar.com/ojs/index.php/comunicar/article/view/C45-2015-16

Sahin, I. (2011). Development of Survey of Technological Pedagogical and Content Knowledge (TPACK). *Turkish Online Journal of Educational Technology - TOJET*, *10*(1), 97–105.

Sang, G., Tondeur, J., Chai, C. S., & Dong, Y. (2016). Validation and profile of Chinese pre-service teachers' technological pedagogical content knowledge scale. *Asia-Pacific Journal of Teacher Education*, *44*(1), 49–65. https://doi.org/10.1080/1359866X.2014.960800

Schmid, M., Brianza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education*, *157*, 103967. https://doi.org/10.1016/j.compedu.2020.103967

Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological Pedagogical Content Knowledge (TPACK). *Journal of Research on Technology in Education*, *42*(2), 123–149. https://doi.org/10.1080/15391523.2009.10782544

Semiz, K., & Ince, M. L. (2012). Pre-service physical education teachers' technological pedagogical content knowledge, technology integration self-efficacy and instructional technology outcome expectations. *Australasian Journal of Educational Technology*, *28*(7). https://doi.org/10.14742/ajet.800

Shinas, V. H., Yilmaz-Ozden, S., Mouza, C., Karchmer-Klein, R., & Glutting, J. J. (2013). Examining Domains of Technological Pedagogical Content Knowledge Using Factor Analysis. *Journal of Research on Technology in Education*, *45*(4), 339–360. https://doi.org/10.1080/15391523.2013.10782609

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Shulman, L. S. (1987). Knowledge and Teaching:Foundations of the New Reform. *Harvard Educational Review*, *57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Sofyan, S., Habibi, A., Sofwan, M., Yaakob, M. F. M., Alqahtani, T. M., Jamila, A., & Wijaya, T. T. (2023). TPACK–UotI: the validation of an assessment instrument for elementary school teachers. *Palgrave Communications*, *10*(1), 1–7. https://doi.org/10.1057/S41599-023-01533-0

Valtonen, T., Sointu, E., Kukkonen, J., Kontkanen, S., Lambert, M. C., & Mäkitalo-Siegl, K. (2017). TPACK updated to measure pre-service teachers' twenty-first century skills. *Australasian Journal of Educational Technology*, *33*(3). https://doi.org/10.14742/ajet.3518

Zelkowski, J., Gleason, J., Cox, D. C., & Bismarck, S. (2013). Developing and Validating a Reliable TPACK Instrument for Secondary Mathematics Preservice Teachers. *Journal of Research on Technology in Education*, *46*(2), 173–206. https://doi.org/10.1080/15391523.2013.10782618

Journal of Information and Organizational Sciences
Volume 49, Number 2 (2025)
Journal homepage: jios.foi.hr

JIOS

DOI: 10.31341/jios.49.2.5
Open Access

UDC 004.056.55:004.738.5
Original Scientific Paper

Bootstrap Forest based method for Encrypted Network Traffic Analysis

Shobana Durairaju[1] and Aswani Kumar Cherukuri[1*]

[1]School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamilnadu, India
[*]Correspondence: cherukuri@acm.org

PAPER INFO

Paper history:
Received 17 April 2025
Accepted 19 August 2025

Citation:
Durairaju, S. & Cherukuri, A. K. (2025). Bootstrap Forest based method for Encrypted Network Traffic Analysis. In Journal of Information and Organizational Sciences, vol. 49, no. 2, pp. 235-249

ABSTRACT

Encrypting communications and data over the Internet becomes essential in ensuring the privacy of communications and protecting the data from increasing threats. Hence, majority of Internet traffic and networked communications are encrypted now. However, encryption also provides a means for attackers to hide them behind encrypted communications and conduct malicious activities. Analyzing the unencrypted communications is relatively easy. The same task is highly challenging due to the presence of encryption in network communication. Conventional network analysis methods fail to analyze encrypted communications. There are methods like flow monitoring that are available to detect encrypted traffic and analyze traffic flow related features. By using traditional analysis methods, we could not achieve accurate detection and classification of encrypted network packets in various types of network traffic such as VoIP, Text, Audio, Video, VPN traffic. In our work, we have proposed the Bootstrap Forest model to analyze and classify encrypted network traffic. Bootstrap Forest model accurately classifies the encrypted network traffic using statistical and time-based features. The performance of the proposed model is evaluated and compared with the performance of other machine learning models under various performance metrics. The three publicly available datasets such as UNSW-NB15, ISCXTor 2016 and ISCXVPN 2016 datasets were used in our experimentations and evaluations. The experimental results show that our proposed model provides the best performance for classifying encrypted network traffic while comparing the F1 score with other methods.

Keywords: Encrypted Network Traffic, Benign Packets, Machine Learning, Network Packets, Network Traffic, Bootstrap Forest

1. Introduction

Network traffic analysis and classification is an important measure in network management. Classification of network traffic becomes essential in various aspects such as analyzing normal traffic, analyzing encrypted network traffic and further classifying normal and malicious traffic present in the network. Due to the major increase in encrypted network traffic, the availability of resources should be properly shared by the network and maintained by the network management to improve the quality of service. The data communication through encrypted IP packets have become as the standard in modern communication (Velan et al., 2015). Further, the amount of network traffic is rapidly increasing in each and every network. So, secured data communication plays a vital role in all the aspects of network communication. Even though the network is secured, malicious attackers are trying to analyze the network traffic and include their malicious communications in the encrypted network. Identifying the malicious attacker communication involved in encrypted network traffic is still a major challenging task.

Various traditional methods are available for identifying malicious traffic and classifying encrypted network traffic. Some of them are Deep Packet Inspection (DPI), Port-based, Payload-based, Protocol-based etc. But still these methods have many limitations on their own. In general, both plaintext traffic information and encrypted data information are available in encrypted network traffic (Velan et al., 2015). Due to unavailability of data in the encrypted network traffic, some of the traditional methods, for example DPI and Pattern matching, could not provide accurate traffic classification. Classifying encrypted network traffic is still a major problem. Nowadays, various machine learning models like Support Vector Machine (SVM), Neural networks, Random Forest, K-Nearest Neighbor (K-NN), Decision Tree, Naive Bayes, etc., are used to improve the performance on classifying encrypted network traffic.

The main purpose of this work is to perform better analysis and improve the classification performance in encrypted network traffic using our proposal to apply Bootstrap Forest (BF) model. The proposed BF model works on the concept of bagging, which makes prediction based on multiple decision trees. Further, the proposal of using BF model could satisfy the research gap identified in the various machine learning models used in the classification of encrypted network traffic by concentrating the misclassification rate. To show the effectiveness of the proposed model, experiments were carried out on the three publicly available datasets such as UNSW-NB15 (Moustafa & Slay, 2015), ISCXTor 2016 (Xu et al., 2022) and ISCXVPN 2016 datasets (Aouedi et al., 2022), (Lotfollahi et al., 2020). To perform the analysis on the encrypted network traffic, we have done the literature review on the various machine learning models which are already used and compared their strengths and limitations with the proposed model. In the above mentioned datasets, statistical and time-based features were taken for the analysis and classification of encrypted network traffic. The experimental results of BF model and other machine learning models are compared and discussed.

The main objectives of our work can be summarized as follows:

- First, we have analysed the publicly available datasets and identified the statistical and time-based features to classify encrypted network traffic.
- Next, we have done a study on how these selected features can be used for the analysis and classification of encrypted network traffic.
- Further, we have proposed the BF model to improve the performance of encrypted network traffic analysis and classification.
- In this work, we have selected three publicly available datasets to evaluate the performance of some machine learning models and compared them with our proposed BF model.
- Then, the experimental results of the proposed BF model are compared with the results of other machine learning models in classifying encrypted network traffic.

This is the first work to analyse and classify encrypted network traffic using BF model. The rest of the paper is organized as follows. Section 2 defines the main sources of existing Encrypted Network Traffic Analysis (ENTA), various protocols and its structure used in encrypted network traffic, types of network traffic, existing classification techniques used for ENTA and followed by related work. Section 3 explains the proposed model used for the classification of encrypted network traffic. Section 4 presents the features and datasets used for the classification of encrypted network traffic, experimental setup and its analysis and results. Finally, section 5 summarizes our research and conclusions.

2. Background

ENTA: Encrypted network traffic can be analysed from the sources of information obtained from the protocols used in secure communication. They are Hypertext Transfer Protocol Secure (HTTPS), Internet Protocol Security (IPsec), Secure Shell (SSH) protocol and Secure Real-time Transport Protocol (SRTP). HTTPS uses Transport Layer Security (TLS) as well as its predecessor Secure Socket Layer (SSL). HTTPS can be used in all the online applications while transferring sensitive data such as online banking, email service, online purchase, etc. SSL is used to encrypt the user data in order to provide secure communication when data is transferred through the Internet. Through its secure communication, it provides data privacy, authentication and data integrity. In public network, IPsec is used to transfer the data in secured manner by providing its authentication and by encrypting the network packets that are communicated in the Internet Protocol (IP) Network. SRTP is used to control and provide security for real-time transport protocol sessions. For secure data communication, most of the networks use protocols such as HTTPS, TLS, SSH and IPsec (Cherukuri et al., 2024). IPsec provides its security in both transport and tunnel mode. In the transport mode, IP packet gets encrypted and in the tunnel mode, the entire IP packet along with the header details gets encrypted. We can also understand the security from the basic frame format of TLS and SSH protocol packet format. We can gather the payload and header information from the protocol frame format for inspection (Elmaghraby et al., 2024). Here, Payloads can be verified from the information obtained from the payload of the packet. Packet Inspection follows the predefined patterns like regular expression as digital signatures for each protocol.

DPI, Shallow Packet Inspection (SPI), Medium Packet Inspection (MPI) are the three types of packet inspection techniques used to monitor the network packets. SPI is the basic network packet monitoring technique which is used to monitor and inspect the data packets based on source IP address, destination IP address, source port number and destination port number. It can inspect the packets only at Network Layer, Data Link Layer and Physical Layer in OSI and TCP/IP Layers. MPI can act as an intermediate node in the networks. It is also called middle-boxes and they are used to monitor and inspect the data packets only at presentation layer, session layer, transport layer, network layer, data link layer and physical layer. At the same time, MPI cannot be used in Application Layer and it can just act as a gateway between nodes and Internet Service Providers. DPI is the only technique which uses IP header and payload information to monitor and inspect the data packets in the networks. The major advantage of DPI is that we can monitor and inspect the real time data packets. Also, DPI can be used in all layers of OSI and TCP/IP models. Both SPI and MPI did not use the IP header and payload information to inspect the data packets in the networks while the same can be done using DPI. So, DPI is the best technique when compared to SPI and MPI and it can be used in all the TCP/ IP Layers. DPI is the best traditional packet inspection technique because it can check whether every packet is transmitted in the correct format or not.

2.1. Types of network traffic

During the data communication in Internet, many different types of network traffic namely, web browsing, email, chat, live streaming, file transfer, VoIP, VPN, Secure browsing, torrents, etc.. enters into the network at once. They can be classified based on their traffic pattern and their format. Voice traffic is the real time traffic. It is a very sensitive traffic type. The major factors to be considered in Voice traffic for lossless communications are delay time, jitter and packet loss. The protocol used in Voice traffic is UDP. The important application in voice traffic is VoIP in which we can make voice calls using Internet connection instead of traditional phone lines. Video traffic plays an important role in today’s scenario. It can be of one among the two types such as stored video traffic and real time video traffic. Stored Video traffic like watching YouTube is not real time video traffic in which the time delay and packet loss does not affects its traffic. At the same time, Videoconferencing is a real-time video traffic in which time delay and packet loss play an important role. If the time delay and packet loss are high, this type of communications will become impossible. The protocol used in video traffic is also UDP. Next, data traffic is not a real time traffic and it is not a sensitive traffic type. The main applications in data traffic are emails, web pages, file transfer, etc. Time delay and packet loss are not the important factors in data traffic. The protocol used in data traffic is TCP. In data traffic, if delay and packet loss occurs, TCP will help in retransmission of data which does not affect the communication. Among all the other types, encrypted traffic plays an major role in the current environment. Here, network traffic is defined as the total amount of data that is moving across various connected devices (i.e., network) at a particular time. Network data is always encapsulated in network packets, which provide the payload in the network.

Encrypted network traffic plays an important role in secured communication. HTTPS, TLS/SSL plays a vital role in improving security in data transfer. It can be used in all online applications while transferring sensitive data such as online banking, email service, online purchase etc. We can classify the network traffic based on protocol-based, port-based, flow-based and machine learning-based methods. They are as follows:

Protocol-based classification method (Velan et al., 2015): We can classify network traffic based on information collected from the packet header. This is the traditional model used to identify and classify network traffic. The protocol format should be changed every time based on the protocols used in every packet. It is a very fast and simple method but susceptible to attack.

Port-based classification method (Boumhand et al., 2023): In this method, the network traffic can be classified based on the port information of the application type received from TCP/UDP. From that we can get the details of source and destination port numbers of the packets. By comparing the port numbers used by the protocols with the standard port numbers, we can classify the network easily. This method is very simple, but it is vulnerable to port detection attack.

Flow-based classification method (Azab et al., 2024): This method is used to classify the network traffic by analysing the flow-based features. Various flow-based features play a major role in the classification of encrypted network traffic. From this, we can manually select and derive the features to classify the network.

Payload-based classification method (Sheikh & Peng, 2022): This method analyses the network traffic based on the information available in the packet's payload. In addition, this method uses the pattern matching approach to classify the encrypted network traffic. So, we can consider it as DPI.

2.2. Techniques used for ENTA

Statistical based analysis (Velan et al., 2015; Papadogiannaki & Ioannidis, 2021): It is a method of analysing the pattern of encrypted network traffic data. Statistical based analysis helps us to identify the normal behaviour of the network traffic and the same is used to detect the anomalies present in the network traffic. It is a simple method to classify the encrypted network traffic.

Machine learning based analysis (Ahmad et al., 2021; Bagui et al., 2017): It trains the machine learning model based on the known input and predicts the future output. In this method, Random Forest, Decision Tree, Logistic Regression, SVM, Naive Bayes and K-NN classification algorithms can be used to classify the encrypted network traffic. By applying these methods, we can analyse and classify the network traffic to achieve accurate classification results without the knowledge of human experts.

Deep learning-based analysis: (Seydali et al., 2023; Lotfollahi et al., 2020) Deep Learning is a type of machine learning techniques which uses artificial neural networks to perform computations for complicated data. This method also works well in classification problems.

2.3. Related Work

Various researchers have used traditional methods like DPI and port-based methods for analysing and classifying encrypted network traffic. However, these methods produce less performance in classifying encrypted network traffic. Nowadays, protocol-based, payload-based, flow-based and machine learning based classification methods are widely used to identify the important features for classifying encrypted network traffic (Zhao et al., 2021). (Chen Y et al., 2024) used hierarchical features to avoid the issue of inaccurate feature generation and improved the efficiency. (Seydali et al., 2023) introduced a hybrid learning approach called as CBS which uses spatio-temporal and statistical features to achieve higher accuracy when compared to other models. (Velan et al., 2015) have done a deep survey on the methods of ENTA and classification. In their survey, they have used protocol structure and identified that the protocol structure gives the lot of information from the initial connection phase for ENTA and classification. Then, they have done survey on payload and feature-based classification methods and encryption protocols which are used for encrypted network traffic. Next, they have done the survey on feature-based classification methods and identified strength and weakness of feature-based classification for encrypted network traffic.

(Moustafa & Slay, 2015) proposed a method to generate the features of UNSW-NB15 dataset. In this work, they have explained the different types of features used in other three different datasets and analysed those features to perform Network Intrusion Detection Systems. Further, they have compared those features with UNSW datasets features. They have identified and stated that UNSW datset can be used for analysing NIDS in the research community. (Lotfollahi et al., 2020) proposed a framework that can analyse both traffic characterization and application identification in encrypted network traffic using publicly available VPN-non-VPN dataset. They have done experiments using various machine learning models and compared the performance of the models using performance metrics. (Bagui et al., 2017) used time related features to classify encrypted network traffic in VPN and non-VPN datasets. The authors proposed a new framework using machine learning models, comparing the classification performance and identified the suitable model for encrypted network traffic classification using time related features. (Xu et al., 2022) proposed an encrypted traffic classification with Path signature using session packet length and done traffic path transformations to expose additional features. The authors applied a few machine learning models on six different datasets and compared the performance of the models to identify the suitable model. (Dener et al., 2023) proposed a model which combines feature selection, gradient recurrent unit's algorithms to achieve data balancing. The authors selected the features by using Random Forest algorithm and used some oversampling techniques to reduce the negative effect in imbalanced datasets. The authors have implemented various machine learning models and compared the performance with their proposed model.

(Hu et al., 2023) used spatial and temporal features using LSTM and CNN and further they have included automatic mapping and labelling of feature mechanisms. This mechanism helps to adopt network traffic datasets quickly and can handle both encrypted and unencrypted traffic. By using various machine learning models, they have achieved higher performance in classifying encrypted network traffic. (Elmaghraby et al., 2024) proposed a framework using features selected from packet length, time stamps or TLS information, encrypted payload information as input. The authors applied some ensemble method and achieved higher performance in classifying encrypted network traffic. (Ahmad et al., 2021) proposed feature clusters method and used machine learning models. The authors achieved better performance by their proposed method using feature clusters and achieved better classification performance. (Aouedi et al., 2022) proposed an ensemble-based approach for ENTA and achieved better performance.

(Zhao et al., 2021) presents a survey on the various common features used in the classification of encrypted network traffic. (Azab et al., 2024) have done a review on network traffic classification based on various classification techniques, datasets and features used. They have also reviewed malicious behavior identification using various datasets, features and challenges in network traffic classification. (Chen Z et al., 2023) proposed a multi-flow encrypted network traffic classification method with various features. (Peng et al., 2024) have used both plaintext and encrypted text for classifying encrypted network traffic and they have done experiments on two publicly available datasets. (Lashkari et al., 2017) used time-based features to analyse Tor traffic flows and application type traffic. (Boumhand et al., 2023) proposed a framework which is used to detect multi-activity traces in network traffic classification. (Sheikh & Peng, 2022) have done a survey on encrypted network traffic classification techniques, procedures to perform analysis, datasets, various machine learning models and the challenges. (Papadogiannaki & Ioannidis, 2021) have done a survey on various traditional ENTA techniques, applications and challenges.

3. Proposed Methodology

In our work, we propose applying ensemble enabled BF model for ENTA and its classification. Considering the literature, on achieving some more additional benefits, we have proposed the BF model in place of Random Forest in our ENTA and its classification. JMP has introduced this BF model by implementing Random Forest algorithm as part of it with some additional features. Among the various additional features, column contribution is the major one. It helps to identify the set of most contributing important predictors or features in the dataset which could simplify the model by reducing the dimensionality and focus on the most relevant data. Further, BF provides another advantage in terms of model interpretability. Here, it provides a way to explicitly identify the most valuable features with which the model works better while the Random Forest model could not provide it explicitly. Further, column contribution provides the visual representation of the most contributing predictors or features along with its quantifiable measure to reduce the impurity. These are the key benefits we could achieve, while applying the BF model Instead of Random Forest model. Overall, the main difference between Random Forest model and BF model is in configuring and controlling, further it is not in algorithm.

BF model makes predictions by combining multiple decision trees. Each tree in Bootstrap samples is built with random samples of training data with replacements. It also trains the model by building multiple decision trees to improve the performance of the model with less error rate. In Random Forest, we need not mention how many predictors required to be used for implementation. In JMP (SAS Institute Inc., 2023), we can manually enter the number of predictors. Further, all of these predictors are sampled at each split in the decision tree. In addition, we can identify and check whether the selected predictors are more important for classification or not. It is decided based on the noticed misclassification rate. BF provides measures for calculating the error rate for out-of-bag observations to estimate predictive performance. This out-of-bag misclassification rate serves as a cross-validation as like the estimate of model performance. If, this value is higher in validation than training, we can choose the other predictors. The predictors can also be identified by column contributions in JMP. So, error rate can be easily identified and rectified using BF model. It makes final predictions by averaging the predictions of multiple decision trees. BF model uses bagging to reduce variance and improve classification accuracy. By using the BF model, we can easily identify and reduce the misclassification rate of the models. So, we can train the model and create a model to obtain the better performance in analyzing and classifying the encrypted network traffic. In this way, BF model helps us to make interactive analysis and reporting.

Figure 1. shows the proposed architecture of BF model for ENTA and classification. First, raw encrypted network traffic datasets are collected. Those data sets are then preprocessed to get extracted features. From those extracted features, further important features are selected. For experimental analysis, 70% of samples are taken for training and 30% of samples are taken for testing from the dataset. Then, the BF model is applied by entering BF specifications. The BF model performs bagging to improve the performance of the model by reducing bias. From the raw datasets, the BF model makes a final prediction by combining multiple decision trees obtained from every random subset of samples with replacements taken for training. It also trains the model by building multiple decision trees and makes predition by aggregating multiple decision trees. It also diplays misclassification rate and makes final predictions with lower misclassification rate to improve the performance of the model. Thus, the BF model improves performance.

Figure 2. shows the workflow of ENTA and classification using machine learning models. Workflow consists of following steps such as dataset collection, data representation and preprocessing with labeling. Then by applying Principal Component Analysis, important features can be extracted for analysis and classification. Here, we have used three datasets like UNSW-NB15, ISCXTor 2016 and ISCXVPN 2016 and their statistical and time-based features can be analysed to classify the encrypted network traffic. We have mainly focused on flow-based statistical features, session-based and time-based features in the preprocessed data. We have split each dataset into training and testing sets to implement various algorithms like SVM, BF and Neural networks. Here, R-square value and confusion matrix can be obtained from the experiments. The R-square value denotes the performance of the model and the other performance metrics like Precision, Recall and F1 score can be calculated from confusion matrix. While comparing the R-square value and F1 score, BF model could show the better performance with less misclassification rate.

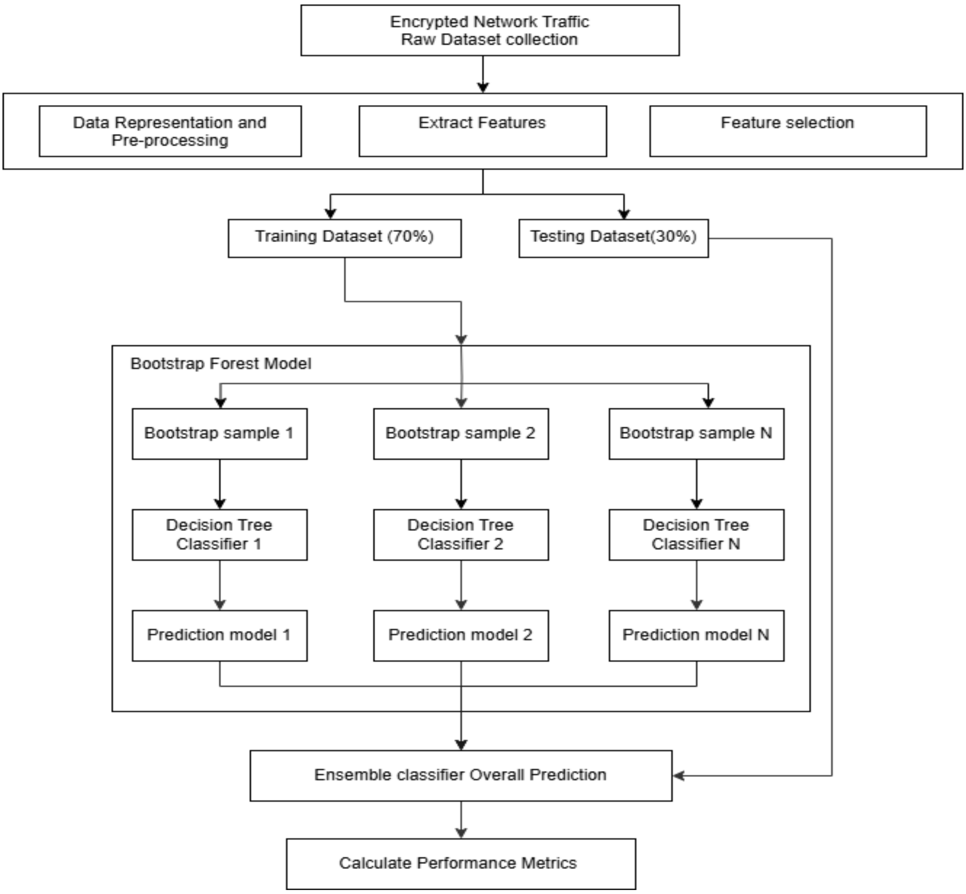

Figure 1. Proposed Architecture of Bootstrap Forest (BF) Model

4. Experimental Setup and its Analysis

4.1. Features used for ENTA and classification

Encrypted network traffic uses different types of features for its analysis and classification. They are as follows. Statistical features (Seydali et al., 2023; Azab et al., 2024): We can define the statistical features from our datasets. We can obtain the packet length by calculating the mean, average, std length of the packets. From these features we can analyze the original length which can be used for classification.

Temporal features (Hu et al., 2023): We can identify the temporal features by analyzing packet size, packet frequency and session duration of the packets. These time dependent features can be used to identify and classify encrypted network traffic.

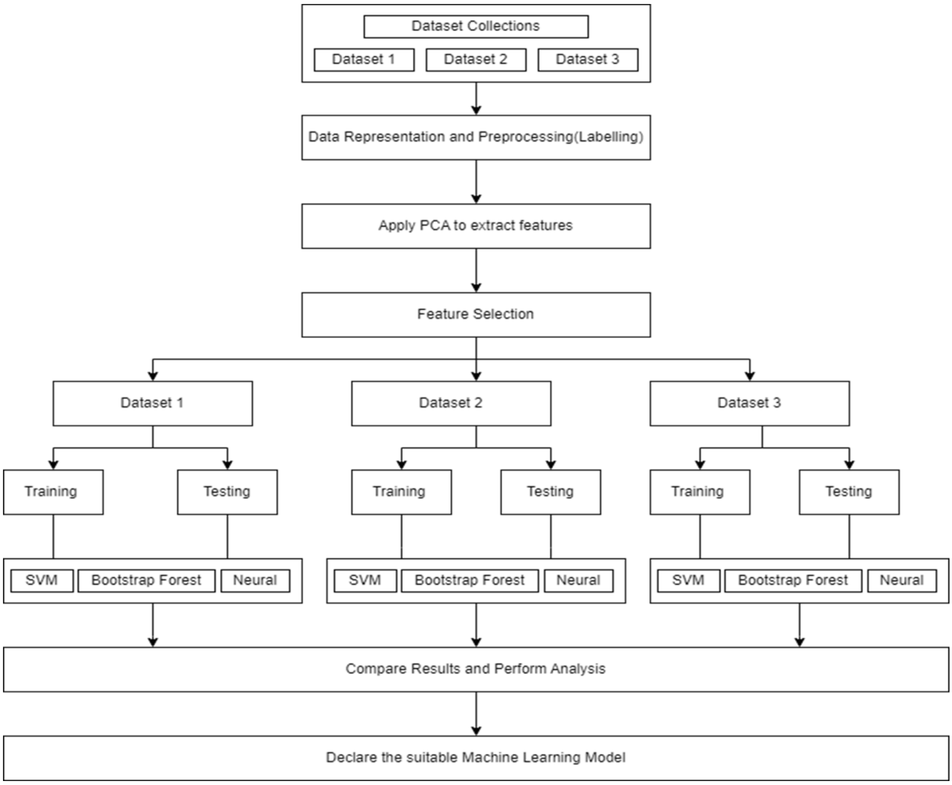

Figure 2. Workflow of ENTA using machine learning models

Residual Plain text features (Peng et al., 2024): The plain text features obtained from the payload of the packets can be used for the analysis of encrypted network traffic. The features like cipher suite, compression method and TLS information can be extracted from the packet payload and used for analysis of packets.
Spatial features (Hu et al., 2023): We can identify the spatial features from the particular session of network packets in bytes. It helps to capture the spatial relationship between network packets in bytes and that can be used to analyze and classify the encrypted network traffic.

Inter-Arrival Time sequence features (Seydali et al., 2023; Bagui et al., 2017; Lashkari et al., 2017): We can calculate the Inter-Arrival Time (IAT) from the flow of packets like IAT min, IAT max, IAT mean, IAT std. From the calculated IAT, we can analyze the packets by matching with the time sequence from the normal packets.

Length sequence features (Seydali et al., 2023; Chen Z et al., 2023): We can find the length sequence features such as min, max, mean and std length of the packets. By calculating these features, we can match with the normal packet values which can be used for analysis and classification.

Feature selection plays very important role to analyse the encrypted network traffic. When, we analyse the flow of data from encrypted network traffic, we can see temporal features, spatial features and statistical features. From the literature, to get high accuracy in classifying encrypted network traffic, we have taken statistical and time-based features to improve the prediction accuracy in ENTA and classification. The list of flow-based statistical features is shown in the Table 1 (Seydali et al., 2023). Out of 79 features, we first consider the features related to packet length. The most important features related to packet length considering encrypted network traffic classification are min packet length, packet length mean, backward packet length min, backward packet length mean, packet length std, average packet size and average forward segment size. From these features, we can analyse the length of the packet for classification. In general, the packet size for Maximum Transmission Unit (MTU) can be 1500 bytes consisting of IP packet with IP header. So, we can classify the traffic by considering the size of packet. If the packet size is lesser or minimum than 1500 bytes, we can consider as benign packet. If the packet size is greater than or equal to 1500 bytes, we can consider as normal packet. If a packet size is lesser than 1500 bytes then there might be the possibility of

Sl. No	Feature Name	Feature Description	Minimum and Maximum value	How it is used for analysis
Statistical Features				
1	Min Packet Length	Lower limit on length of a packet	21-1500 bytes	If the size of the packet is less than MTU it is considered a malicious otherwise normal packet.
2	Packet Length Mean	Mean length of a packet		
3	Packet Length Std	Standard deviation length of a packet		
4	Backward Packet Length Min	Minimum size of the packet from receiver to sender		
5	Backward Packet Length Mean	Mean size of the packet from receiver to sender		
6	Average Packet Size	Average size of packet	1460 bytes	
7	Average Forward segment size	Average size of the segment from source to destination		
8	Flow IAT Mean	Mean time between two packets sent in the flow	Bot communication generates uniform sized small TCP/UDP packets	If the IAT exceeds the threshold value, then it is considered a malicious otherwise normal packet
9	Flow IAT Std	Standard deviation time between two packets sent in the flow		
10	Flow IAT Max	Maximum time between two packets sent in the flow		
11	Flow IAT Min	Minimum time between two packets sent from sender to receiver		
12	Min Packet Length	Lower limit on length of a packet	21 bytes	If the packet or header length exceeds the threshold value, then it is considered a malicious otherwise normal packet
13	Max Packet Length	Upper limit on length of a packet	65535 bytes	
14	Fwd Header Length	Total bytes used for packet header from source to destination	20-60 bytes	
15	Idle Max Time	Maximum time a flow was idle before becoming active	600 seconds or 10 minutes for TCP/UDP flows	If the idle time of the flow exceeds 600 sec, then it is considered a malicious otherwise normal packet

Table 1. The list of flow-based features used for analysis

Sl. No	Feature Name	Feature Description	Minimum and Maximum value	How it is used for analysis
Residual Plaintext Features				
16	Cipher suite	Cryptographic algorithm used	TLS AES 128 GCM SHA256 TLS AES 256 GCM SHA384 TLS AES 128 CCM SHA256 TLS AES 128 CCM 8 SHA256 TLS CHACHA20 POLY1305 SHA256 TLS AES 256 GCM SHA384 TLS CHACHA20 POLY1305 SHA256 TLS AES 128 CCM SHA256 TLS AES 128 CCM 8 SHA256	If the cipher suite information gathered matches the following standard five combinations, then it is considered as normal otherwise, malicious packet
17	Compression method	Compression method length	One byte length always and method set to null (0)	If the compression method value equals to null then it is considered as normal otherwise malicious packet
18	TLS extension information	Extension information	Every field has standard value	If the length of the extension field matches the standard length, then it is normal otherwise malicious packet

Table 1. The list of flow-based features used for analysis - continued

DDoS attack and if the attack has happened, the size of packets is always very small. Further, attacker always generates smaller packets or empty packets. So, from the features related to Packet Length, we can classify malicious or benign traffic.

Considering encrypted network traffic classification, other important features related to IAT (Bagui et al., 2017; Lashkari et al., 2017) are calculated as follows: The time between the two packets sent from the same source helps to identify the features such as Flow IAT Mean, Flow IAT Std, Flow IAT Max and Flow IAT Min. From these features, we can analyse the IAT for encrypted Network Traffic classification. By using the features related to IAT, we can classify Botnet traffic from normal traffic, because normal bot communication will indicate more uniform sized, small TCP/UDP packets. If the IAT exceeds the threshold value, then it is considered as malicious, otherwise normal packet. The features like Min and Max Packet Length, Fwd Header Length (Chen Z et al., 2023) are used to identify and classify encrypted network traffic. If the packet or header length exceeds the threshold value, then it is considered as malicious otherwise normal packet. The feature Idle Max Time is used to classify the normal or malicious TCP/UDP flow packets. If the Max Idle Time of the TCP/UDP flow exceeds 600 seconds or 10 minutes, then it is considered as malicious otherwise normal packet. Feature selection is one of the most important parts in encrypted network traffic classification because it plays a very important role in its performance. We have done a major study on the features used for different machine learning models and found that various features are extracted and used for classification namely, flow-based, payload-based, packet-based, statistical based features, etc. (Zhao et al., 2021; Seydali et al., 2023; Bagui et al., 2017; Hu et al., 2023). Similar studies are already available in the literature for identifying the features. However, our work is different from others in terms of how these features are useful for analysing and classifying the encrypted network traffic. In our work, we have identified the statistical and time-based features which are used to identify and classify the malicious and benign packets in the encrypted network traffic. In addition, with the support of column contributions in JMP we can identify the features which has strong impact on our model performance. To analyse each feature, we have identified packet length and IAT. Further, to classify malicious and benign packet, we have compared the captured packet length and IAT with the standard values. The list of flow-based features are already shown in the Table 1. and they are helpful in analysing and classifying encrypted network traffic. By this work, we could provide a better understanding on feature selection to be used for machine learning based ENTA and classification. The first step in the analysis is to detect encrypted network traffic then to classify the network into various categories and subsequently identify any malicious traffic present in the network. Our main aim is to perform accurate classification of encrypted network traffic using machine learning algorithms.

4.2. Datasets used for encrypted network traffic classification

There are several datasets available in the literature for the purpose of ENTA and classification, such as UNSW-NB15 (Bhandari et al., 2023), NSL KDD, ASNM-CDX-2009, ISCXVPN 2016 (Dener et al., 2023), ISCXTor 2016 (Xu et al., 2022) etc. From the survey, we can understand that they have collected, analysed and used different traditional methods for classification. Since, they have faced different challenges while implementing traditional methods for real time ENTA and classification, we have focused towards achieving accurate classification of encrypted network traffic using machine learning models with minimum number of features and without involving experts. From the available datasets, we have taken the following datasets for analysis.

UNSW-NB15 Dataset (Moustafa & Slay, 2015): In this research, first we have taken UNSW-NB15 dataset (UNB, 2020). It contains more than 68,000 flows in total. From these flows, the network contains seven groups of traffic classification. Out of 68,000 flows, we have taken more than 3,300 samples for analysis and classification. In this dataset, there are seven different types of traffic flow like fuzzers, exploits, shellcode, reconnaissance, DoS, generic and normal were analysed and classified.

ISCXTor 2016 dataset (Xu et al., 2022): The second dataset which we have taken for analysis is ISCXTor 2016 dataset (UNB, 2023). The dataset contains more than 14,000 flows in total. Out of 14,000 flows, we have taken more than 8,000 samples for our analysis. In this dataset, there are six different types of network traffic flows like Audio, Browsing, File Transfer, Video, VOIP and P2P are analysed and classified.

ISCXVPN 2016 dataset (Dener et al., 2023): The third dataset which we have taken for analysis is ISCXVPN 2016 dataset (UNB, 2024). The dataset contains more than 14,600 flows in total. Out of 14,600 flows, we have taken more than 14,000 samples for our analysis. In this dataset, there are seven different types of network traffic flows such as chat, file transfer, mail, streaming, VOIP, P2P, browsing etc. are analysed and classified. From the samples, we have taken flow-based features for accurate classification of encrypted network traffic.

4.3. Results and Analysis

The proposed work was implemented and evaluated using JMP Pro, Version 18.0 (SAS Institute Inc., 2023) SAS Institute Inc., Cary, NC, USA a statistical analysis and machine learning tool.

Evaluation: To perform analysis and classification of encrypted network traffic, we have chosen the UNSW dataset (Ahmad et al., 2021) with 3350 samples for multiclass classification. For analysis and classification, statistical features such as dur, sbytes, dbytes, sttl, dttl, sload, dload, spkts, dpkts, smeansz and dmeansz were used. In order to perform the classification, various machine learning models, SVM, BF and Neural networks were used. Among these, BF model provides the best R-square and F1 score value which is fitted for the UNSW-NB15 dataset. The BF specification for implementing the datasets was given as follows: Number of terms sampled per split was 4, Maximum splits per tree was 30, Maximum number of terms was 10 and Random seed control for reproducibility was 123. Next, we have taken 14508 samples from ISCXTor 2016 dataset. For the analysis and classification of encrypted network traffic, the features (Baldini, 2020) such as flow duration, flow bytes, flow pckts, fwd IAT min, bwd IAT min, active mean, active max and active min were used. The same machine learning models as SVM, BF and Neural networks were used for classifying the encrypted network traffic. Among these, BF model provides the better R-square and F1 score value that is fitted for the ISCXTor 2016 dataset.

Then, we have taken 14651 samples from the ISCXVPN 2016 dataset (Dener et al., 2023). For analysis and classification of encrypted network traffic, the features like duration, total biat, mean biat, flow bytes, min flow IAT, max flow IAT and std active were used. The same machine learning models like SVM, BF and Neural networks were used for classifying the encrypted network traffic. Among these, BF model provides the best R-square and F1 score value that is fitted for the ISCXVPN 2016 dataset.

4.4. Model comparison

The SVM, BF, Neural networks models were compared using performance metrics such as Precision, Recall and F1 score. Table 2. provides the evaluation metrics obtained by applying various machine learning models. We could not compare the models by calculating accuracy for an imbalanced datasets. So, we have used the performance metrics such as Precision, Recall and F1-score to analyze and classify the encrypted network traffic. Here, the above mentioned performance metrics are calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1 score} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}}$$

where TP,FP,TN and FN denotes true positive, false positive, true negative and false negative respectively.

The R-square value helps us to identify the suitable model for the given datasets. From Table 2, while comparing the R-square value of all the models for the three datasets, BF model provides better performance with the lowest misclassification rate. Further, we have calculated the performance metrics to compare the F1 score of all the models are given in the Table 2. This F1 Score is obtained from the confusion matrix for our three datasets which suggests the performance of the model. Here, the F1 score of the BF model is higher. It shows that, this model provides better performance compared to other models in classifying encrypted network traffic.

Algorithm	R-square	Precision	Recall	F1 score	Misclassification Rate	
					Training	Validation
UNSW NB15 dataset						
Bootstrap Forest (BF)	0.9421	69.930	66.457	68.150	0.0690	0.0418
Neural	0.8895	49.000	46.468	47.701	0.7085	0.7403
SVM	0.8890	57.018	48.320	52.316	0.1321	0.0776
ISCXTor 2016 dataset						
Bootstrap Forest (BF)	0.9332	78.732	79.329	79.029	0.2119	0.1950
Neural	0.8765	58.019	61.521	59.719	0.3000	0.3078
SVM	0.8189	69.618	60.897	64.948	0.3340	0.3349
ISCXVPN 2016 dataset						
Bootstrap Forest (BF)	0.9003	70.035	66.480	68.211	0.2549	0.2567
Neural	0.8035	50.958	36.384	42.455	0.3612	0.3686
SVM	0.8060	63.939	49.950	56.085	0.3278	0.3457

Table 2. Evaluation Metrics obtained by applying various Machine Learning Algorithms

The Receiver Operating Characteristic (ROC) curve and the Precision Recall curve help us to compare the performance of the machine learning models for an imbalanced datasets. The metric Area Under the Curve (AUC) is also useful to compare the performance of the models (i.e) AUC-ROC and AUC Precision Recall (AUC-PR). ROC curve is a graphical representation which shows the performance of the classification models. Precision is defined as the ratio of true positives to the total number of true positives and true negatives. It shows how many predicted responses are correctly classified as true positives. Recall is defined as the ratio of true positives to total number of true positives and false negatives. It shows that how many actual positive responses are correctly classified as positives. Particularly, it helps to identify the performance of the model which has minority classes of samples present in the datasets taken for experiments. The performance of the machine learning models obtained from the experiments is shown in Figures 3, 4 and 5. Figures 3(A), 4(A) and 5(A) represents the ROC curves obtained for UNSW-NB15, ISCXTor 2016 and ISCXVPN 2016 datasets using machine learning models respectively. Here, the ROC curve lies above the diagonal and it displays the efficiency of the model.

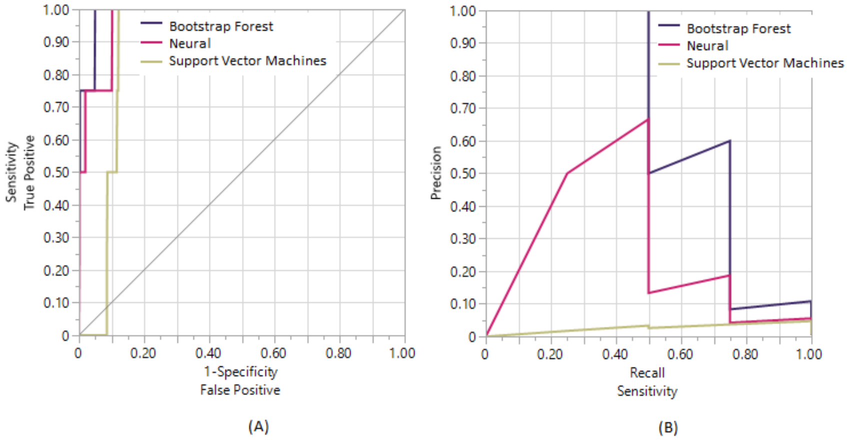

Figure 3. (A) ROC Curve for UNSW-NB15 dataset using machine learning models and (B) Precision Recall Curve for UNSW-NB15 dataset using machine learning models

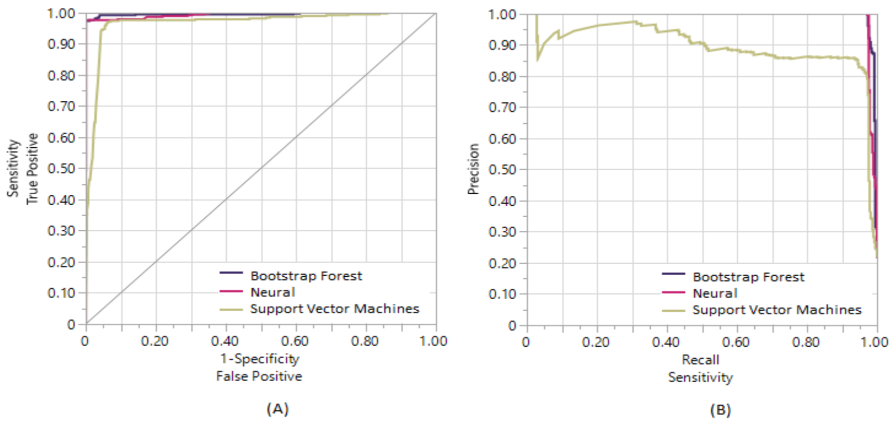

Figure 4. (A) ROC Curve for ISCXTor 2016 dataset using machine learning models and (B) Precision Recall Curve for ISCXTor 2016 dataset using machine learning models

The curve passes through the top left corner of the graph indicates the perfect model suitable for the given datasets. The goodness of the fit for the model are analyzed through the area under the curve.The ROC curve produced by the proposed model is very closer to the top left corner, when compared to ROC curve of all other models. It indicates the proposed model produces better performance compared to all other models. Here, for each dataset, ROC curve of BF model performs better when compared to all other models to classify the encrypted network traffic.

Figures 3(B), 4(B) and 5(B) represent the performance of Precision Recall curves obtained for UNSW-NB15, ISCXTor 2016 and ISCXVPN 2016 datasets using machine learning models respectively. It also helps us to predict the minority classes present in the samples taken for experimental analysis. The Precision Recall curve produced by the proposed model is more closer to the top right corner, when compared to Precision Recall curve of all other models. Further, Table 3. provides the metrics to compare the various machine learning models. From the Table 3. it is observed that the metrics AUC-ROC and AUC-PR of the proposed model is higher when compared to all other models. It indicates that the proposed model produces better performance when compared to all other models. Here in each dataset, Precision Recall curve of the BF performs better compared to all other models to classify encrypted network traffic.

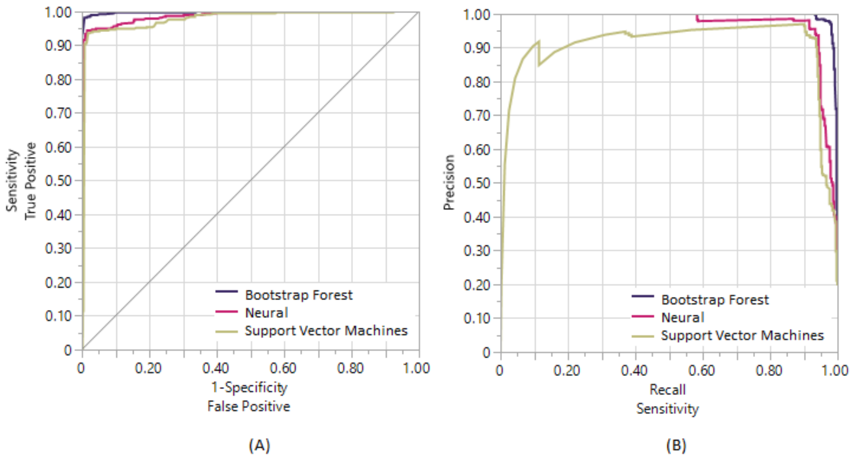

Figure 5. (A) ROC Curve for ISCXVPN 2016 dataset using machine learning models and (B) Precision Recall Curve for ISCXVPN 2016 dataset using machine learning models

Machine learning models	Dataset 1 UNSW-NB15		Dataset 2 ISCXTor 2016		Dataset 3 ISCXVPN 2016	
	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR
Bootstrap Forest (BF)	0.9869	0.6614	0.9960	0.9938	0.9982	0.9953
Neural	0.9688	0.2607	0.9946	0.9896	0.9879	0.9700
SVM	0.8971	0.0265	0.9682	0.8948	0.9776	0.8972

Table 3. Model Comparison using AUC-ROC and AUC-PR

We have done ENTA and classification by using an ensemble method called BF which is the main contribution of our work. The experimental results show that BF model provides better performance and lower misclassification rate when compared to other machine learning models. While focusing the limitations of our BF model, though it is robust ensemble method, they are not immune to the effects of imbalanced datasets, model interpretability and computational cost. In future works, we try to address these limitations.

5. Conclusion

The encrypted network traffic is growing tremendously every day in internet. This traffic needs to be analyzed to classify malicious traffic from the regular benign network traffic present in the network. Our work is to analyze and classify the encrypted network traffic by applying the various statistical and time-based features in different machine learning models. Similar studies are already available in the literature. Our proposed BF model is different from them in terms of analyzing and classifying encrypted network traffic. Moreover, this is the first analysis of encrypted network traffic classification using the BF model in the literature. By introducing bagging methodology, the proposed model is able to provide better performance with the lowest misclassification rate. In addition, we have analyzed features related to packet length and IAT and provided the description about how these features are useful for analyzing and classifying encrypted network traffic.

References

Ahmad, M., Riaz, Q., Zeeshan, M., Tahir, H., Haider, S.A., Khan, M.S. (2021). Intrusion detection in internet of things using supervised machine learning based on application and transport layer features using unsw-nb15 data-set. EURASIP Journal on Wireless Communications and Networking, 2021, 1–23, https://doi.org/10.1186/s13638-021-01893-8

Aouedi, O., Piamrat, K., Parrein, B. (2022). Ensemble-based deep learning model for network traffic classification. IEEE Transactions on Network and Service Management, 19 (4), 4124–4135, https://doi.org/10.1109/TNSM.2022.3193748

Azab, A., Khasawneh, M., Alrabaee, S., Choo, K.-K.R., Sarsour, M. (2024). Network traffic classification: Techniques, datasets, and challenges. Digital Communications and Networks, 10 (3), 676–692, https://doi.org/https://doi.org/10.1016/ j.dcan.2022.09.009

Bagui, S., Fang, X., Kalaimannan, E., Bagui, S.C., Sheehan, J. (2017). Comparison of machine-learning algorithms for classification of vpn network traffic flow using time-related features. Journal of Cyber Security Technology, 1 (2), 108–126, https://doi.org/10.1080/23742917.2017.1321891

Baldini, G. (2020). Analysis of encrypted traffic with time-based features and time frequency analysis. 2020 global internet of things summit (giots) (pp. 1–5).

Bhandari, A., Cherukuri, A.K., Ikram, S.T. (2023). Analysis of encrypted network traffic using machine learning models. Big data analytics and intelligent systems for cyber threat intelligence (Vol. 1, pp. 71–86). River Publishers, Denmark.

Boumhand, A., Singh, K., Hadjadj-Aoul, Y., Liewig, M., Viho, C. (2023). Network traffic classification for detecting multi-activity situations. 2023 ieee symposium on computers and communications (iscc) (pp. 681–687).

Chen, Y., Yang, J., Cui, S., Dong, C., Jiang, B., Liu, Y., Lu, Z. (2024). Unveiling encrypted traffic types through hierarchical network characteristics. Computers & Security, 138, 103645, https://doi.org/10.1016/j.cose.2023.103645 Retrieved from https://doi.org/10.1016/j.cose.2023.103645

Chen, Z., Cheng, G., Wei, Z., Niu, D., Fu, N. (2023). Classify traffic rather than flow: Versatile multi-flow encrypted traffic classification with flow clustering. IEEE Transactions on Network and Service Management, 21 (2), 1446–1466, https://doi.org/10.1109/TNSM.2023.3322861

Cherukuri, A.K., Ikram, S.T., Li, G., Liu, X. (2024). Classification of encrypted network traffic. Encrypted network traffic analysis (pp. 47–59). Springer,Cham, Switzerland.

Dener, M., Al, S., Ok, G. (2023). Rfse-gru: Data balanced classification model for mobile encrypted traffic in big data environment. IEEE Access, 11, 21831– 21847, https://doi.org/10.1109/ACCESS.2023.3251745

Elmaghraby, R.T., Aziem, N.M.A., Sobh, M.A., Bahaa-Eldin, A.M. (2024). Encrypted network traffic classification based on machine learning. Ain Shams Engineering Journal, 15 (2), 102361, https://doi.org/https://doi.org/10.1016/j.asej.2023. .102361

Hu, F., Zhang, S., Lin, X., Wu, L., Liao, N., Song, Y. (2023). Network traffic classification model based on attention mechanism and spatiotemporal features. *EURASIP Journal on Information Security*, *2023* (1), 6, https://doi.org/10.1186/s13635-023-00141-4

Lashkari, A.H., Gil, G.D., Mamun, M.S.I., Ghorbani, A.A. (2017). Characterization of tor traffic using time-based features. *International conference on information systems security and privacy* (Vol. 2, pp. 253–262).

Lotfollahi, M., Jafari Siavoshani, M., Shirali Hossein Zade, R., Saberian, M. (2020). Deep packet: A novel approach for encrypted traffic classification using deep learning. *Soft Computing*, *24* (3), 1999–2012, https://doi.org/10.1007/s00500-019-04030-2 Retrieved from https://doi.org/10.1007/s00500-019-04030-2

Moustafa, N., & Slay, J. (2015). Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). *2015 military communications and information systems conference (milcis)* (pp. 1–6).

Papadogiannaki, E., & Ioannidis, S. (2021). A survey on encrypted network traffic analysis applications, techniques, and countermeasures. *ACM Computing Surveys (CSUR)*, *54* (6), 1–35, https://doi.org/https://doi.org/10.1145/ 345790

Peng, W., Cui, L., Cai, W., Ding, Z., Hao, Z., Yun, X. (2024). Efficiently and effectively: A two-stage approach to balance plaintext and encrypted text for traffic classification. *arXiv preprint arXiv:2407.19687,*, https:// doi.org/10.48550/arXiv.2407.19687

Seydali, M., Khunjush, F., Akbari, B., Dogani, J. (2023). Cbs: A deep learning approach for encrypted traffic classification with mixed spatio-temporal and statistical features. *IEEE Access*, *11,* 141674-141702, https://doi.org/10.1109/ACCESS.2023.3343189 Retrieved from https://doi.org/10.1109/ACCESS.2023.3343189

Sheikh, M.S., & Peng, Y. (2022). Procedures, criteria, and machine learning techniques for network traffic classification: a survey. *IEEE Access*, *10,* 61135–61158, https://doi.org/10.1109/ACCESS.2022.3181135

UNB (2020). UNSW-NB15 dataset:available online at https://www. unsw. adfa. edu. au/unswcanberra-cyber/cybersecurity/adfa-nb15-datasets. Accessed on, June 10,

UNB (2023). ISCXTor2016 dataset:available online at https://www.unb.ca/cic/datasets/tor.html. Accessed on, September 9,

UNB (2024). ISCXVPN2016 dataset:available online at https://www.unb.ca/cic/datasets/vpn.html. Accessed on, January 20,

Velan, P., ˇCerm´ak, M., ˇCeleda, P., Draˇsar, M. (2015). A survey of methods for encrypted traffic classification and analysis. International Journal of Network Management, 25 (5), 355–374, https://doi.org/10.1002/nem.1901 Retrieved from https://doi.org/10.1002/nem.1901

Xu, S.-J., Geng, G.-G., Jin, X.-B., Liu, D.-J., Weng, J. (2022). Seeing traffic paths: Encrypted traffic classification with path signature features. IEEE Transactions on Information Forensics and Security, 17, 2166–2181, https://doi.org/ 10.1109/TIFS.2022.3179955

Zhao, J., Jing, X., Yan, Z., Pedrycz, W. (2021). Network traffic classification for data fusion: A survey. Information Fusion, 72, 22–47, https://doi.org/10.1016/j.inffus.2021.02.009

SAS Institute Inc. (2023). *JMP® Pro (Version 18)*. Cary, NC: SAS Institute Inc.

Journal of Information and Organizational Sciences
Volume 49, Number 2 (2025)
Journal homepage: jios.foi.hr

JIOS

DOI: 10.31341/jios.49.2.6
Open Access

UDC 004.94:37.091.3:159.922.27
Original Scientific Paper

Factors Affecting Students' Acceptance of Learning Simulation Tools in Computing Education Courses from Social, Technology, and Personal Trait Perspectives

Yonathan Dri Handarkho[1], Theresia Devi Indriasari[1*], Yohanes Sigit Purnomo W.P.[1], Findra Kartika Sari Dewi[1], Citra Yayu' Palangan[1], Vinindita Citrayasa[1], Aloysius Gonzaga Pradnya[1] and Herlina[1]

[1]Informatics Department, Industrial Technology Faculty, Universitas Atma Jaya Yogyakarta, Jalan Babarsari 43, Yogyakarta 55281, Indonesia
[*]Correspondence: devi.indriasari@uajy.ac.id

PAPER INFO

Paper history:
Received 19 May 2025
Accepted 12 August 2025

Citation:
Handarkho, Y. D., Indriasari, T. D., Purnomo Y. S., Dewi, F. K. S., Palangan, C. Y., Citrayasa, V., Pradnya, A. G. and Herlina (2025). Factors Affecting Students' Acceptance of Learning Simulation Tools in Computing Education Courses from Social, Technology, and Personal Trait Perspectives. In Journal of Information and Organizational Sciences, vol. 49, no. 2, pp. 251-270

ABSTRACT

This study presents a theoretical model to explore the factors influencing students' acceptance of simulation tools in computing education. These factors include social influences, technology-related aspects, and personal characteristics. The term "simulation tools" refers to systems that can replicate complex processes and situations, providing students with realistic, hands-on experiences without the risks or costs associated with physical setups. To validate the proposed model, 312 responses from university students were collected. A cross-sectional online survey was conducted, and the participants were selected through purposive sampling. The findings indicated that subjective norms have the most significant direct effect on students' perceptions of usefulness, influencing their views on learning outcomes from using simulation tools in computing education courses. Additionally, social support and self-efficacy were also found to have significant effects. However, the impacts of fidelity and innovativeness were not supported. This study sets itself apart from previous research by using a comprehensive approach to explore the factors influencing student acceptance of simulation tools in computing education. Specifically, this research develops a theory based on the Technology Acceptance Model (TAM) and expands it by incorporating environmental factors and personal characteristics of students.

Keywords: Simulation tools, Learning, SEM, Social influence, Personal characteristics, TAM

1. Introduction

In today's digital era, technological advancements have reached almost every part of life, including education. The use of technology in education has opened new paths for innovative learning methods (Haleem et al., 2022a), especially in fields that require both theoretical understanding and practical skills, such as computing, medicine, and engineering. In these areas, traditional learning often relies on expensive equipment, complex setups, and various supplies that may not always be available or affordable (Haleem et al., 2022b; Soliman et al., 2021). Therefore, simulation tools have become a practical and popular solution (Lavrentieva et al., 2020; Tang et al., 2022; Abdulrahaman et al., 2020). Simulation tools create an immersive learning environment where students can experiment, make decisions, and see realistic results without the risks and costs of real-life situations. For example, in medicine, simulations allow students to practice surgical procedures in a controlled and repeatable setting, helping them build essential skills before working with actual patients. Similarly, in engineering and computer science, simulations let students design and test complex systems, like digital circuits or network configurations, enhancing their technical skills and problem-solving abilities. By providing interactive, hands-on learning experiences, simulations bridge the gap between theoretical knowledge and practical application, making them an essential part of modern education and expanding opportunities for hands-on learning within educational institutions.

In addition to their many benefits, simulation tools address several challenges present in traditional lab-based learning environments. Physical labs and real-world exercises often demand expensive equipment, take significant time to set up and supervise, and are typically limited to small groups due to the need for specialized tools and instructor guidance (Brinson et al., 2015; Wei et al., 2019). These limitations can restrict students' opportunities for hands-on practice, which is crucial for mastering practical skills in fields like engineering, medicine, and computer science. In contrast, virtual simulations allow students to practice repeatedly and at their own pace, promoting deeper learning and reducing the need for constant supervision, which helps institutions manage their resources more efficiently (Brinson et al., 2015; Wei et al., 2019).

As simulation tools have evolved, they now offer more than simple task-based exercises; many features sophisticated interactive models and virtual scenarios that closely mimic real-world systems (Zheng et al., 2022). These tools can simulate complex processes and situations, giving students realistic, hands-on experiences without the risks or costs associated with physical setups. For instance, students can conduct virtual experiments, solve problems, and test various scenarios in a safe, controlled digital environment. This advancement has spurred researchers to investigate the factors that may impact the effectiveness of simulation tools, including ease of use, user satisfaction, perceived usefulness, and other user-centered aspects (Adams et al., 1992). Understanding these factors is crucial, as they are believed to positively influence students' engagement, motivation, and learning outcomes.

The growing integration of simulation-based learning tools in higher education is reshaping the learning landscape across diverse fields such as engineering, nursing, and business. Studies consistently highlight the role of simulation in bridging theoretical knowledge with practical skills, offering students immersive experiences that can boost confidence, self-efficacy, and skill acquisition. For example, Mwansa et al. (2024) and Campos et al. (2020) explore the use of simulation in resource-constrained environments and online STEM (Science, Technology, Engineering, and Mathematics) education, respectively, noting that simulation can effectively supplement traditional labs by enhancing practical skills and motivation. Similarly, in nursing education, Hung et al. (2020) demonstrate that repeated simulation exposures can significantly increase perceived competence and learning satisfaction, suggesting that experiential learning principles enable students to move from "knowing" to "doing" through hands-on practice.

The increased use of simulation tools in education has prompted many scholars to investigate the adoption and acceptance of this technology in educational practice, particularly from the student perspective. Research related to how Studies applying theoretical frameworks like the Technology Acceptance Model (TAM) reveal that students’ acceptance of simulation tools is heavily influenced by factors such as perceived usefulness, ease of use, and enjoyment. This is particularly evident in the work of Yu (2017) in a merchandising context and Lisana and Suciadi (2021) with high school physics students, where enjoyment emerged as a crucial predictor for sustained engagement. Altalbe (2019) adds to this discourse by integrating TAM with ABET objectives to assess simulation-based virtual laboratories for engineering students, finding that perceived usefulness and instrumentation directly impacted performance outcomes. These findings align with Chernikova et al. (2020), who, through a meta-analysis, showed that well-designed simulations, when paired with proper scaffolding and instructional support, could maximize learning gains across cognitive, procedural, and affective domains. The literature also underscores the importance of contextual and pedagogical factors in leveraging simulation for educational efficacy. Wong et al. (2022) and Hamilton et al. (2021) illustrate that realism and usability in simulation design greatly impact learning outcomes by making learning tools more relevant and effective. Furthermore, studies by Yang et al. (2022) and Hung et al. (2020) emphasize the role of self-efficacy and social aspects in driving students' engagement and performance. Overall, the current literature highlights simulation as a versatile and impactful educational tool that, when implemented thoughtfully with attention to usability, realism, and targeted learning objectives, can significantly enhance student engagement, learning satisfaction, and academic performance across fields.

While many studies demonstrate that the quality of simulation tools can enhance learning, few investigate how social, technological, and personal influences interact to shape students' willingness to use these tools in computing courses. Previous research, summarized in Table 1, shows that factors such as realism, intuitive interfaces, and well-designed simulations contribute to increased learning gains (Chernikova et al., 2020; Hamilton et al., 2021; Wong et al., 2022). However, student acceptance is often treated as a secondary concern. Meanwhile, Yu (2017) and Altalbe (2019) focus on factors like perceived usefulness, ease of use, and enjoyment in predicting the intention to use simulations, yet they rarely explore how classroom dynamics affect these perceptions. Additionally, research by Yang et al. (2022) and Hung et al. (2020) emphasizes the role of self-efficacy in driving student engagement and performance but overlooks other aspects that influence engagement, such as the role of the instructor.

Taken together, these summaries highlight that no single line of inquiry captures the interplay of social support, tool qualities, and learner dispositions. This is especially critical in introductory computer courses, where limited physical resources make effective simulations essential. To address this gap, this study expands the Technology Acceptance Model (TAM) by distinguishing three overarching aspects and detailing their components. The social aspect assesses peer and instructor support, ranging from informal encouragement to shared classroom norms. The technology aspect evaluates simulator features such as realism, reliability, and accessibility, which shape students' judgments of value and effort. And last, but not least, the personal aspect reflects learners' confidence in using new software and their curiosity about innovation, influencing how they perceive benefits and usability. By analyzing how these aspects jointly influence perceived usefulness, ease of use, and ultimately, learning outcomes, our model provides a theory-driven yet practical understanding of simulation acceptance. This approach offers educators and designers actionable guidance based on a more comprehensive view of student behavior.

This study, therefore, aims to fill the research gap by examining the factors influencing students' acceptance of simulation tools in computing education from three main perspectives: social factors, technology-related factors, and personal traits. To anchor the investigation in an educational context, data were collected from students enrolled in introductory computer networking courses that utilize Cisco's Packet Tracer simulator and EC-Council's iLabs as part of the Certified Secure Computer User (CSCU) Version 3 curriculum. Cisco's Packet Tracer helps students learn core networking concepts, including routing, subnetting, and network configuration. In contrast, iLabs offers an interactive, hands-on environment where students can practice security skills through real-world simulated scenarios that involve security threats and countermeasures. By providing a comprehensive analysis of these diverse factors, this study seeks to deepen the understanding of what drives students to adopt and effectively use simulation tools in computing education. The findings are anticipated to offer valuable insights for educators, instructional designers, and policymakers, helping them create and promote simulation-based learning tools that are not only accessible and functional but also tailored to the varied needs and expectations of students in today's digital learning environments. This study, therefore, formulates the following research question: RQ. What factors significantly influence students' acceptance of learning simulation tools in computing education courses?

2. Literature Study and Hypotheses Development

Table 1 summarizes the current understanding of students' acceptance of simulation-based learning tools and identifies areas where knowledge is still lacking. The first group of studies, grounded in the Technology Acceptance Model (TAM), demonstrates that perceived usefulness, ease of use, and, increasingly, enjoyment are reliable predictors of behavioral intention across various disciplines, ranging from fashion merchandising (Yu, 2017) to engineering (Altalbe, 2019) and immersive virtual reality (Hamilton et al., 2020). The second group employs learning and cognitive theories to show that well-designed scaffolding and realistic interfaces promote competence, self-efficacy, and satisfaction (Hung et al., 2021; Chernikova et al., 2020; Wong et al., 2022). The third group emphasizes contextual constraints, such as equipment shortages or the adequacy of infrastructure, to illustrate that simulations can mitigate limited physical resources (Campos et al., 2020; Mwansa et al., 2024). Lastly, a smaller but growing number of studies connect personal traits (e.g., self-efficacy, innovativeness) to learning outcomes (Yang et al., 2020).

Collectively, these strands confirm that each aspect, technology, pedagogy, context, and individual characteristics, plays a significant role. However, no prior research integrates all these elements within a single explanatory model, nor do any studies specifically focus on computing courses where simulations replace expensive hardware. Our work addresses this gap by expanding TAM to include social-support variables that operationalize peer and instructor influence (technology qualities that capture realism, reliability, and accessibility, and personal dispositions like motivation and self-confidence. By analyzing the combined effects of these constructs, we aim to provide a comprehensive, theory-driven, and practically useful account of simulation acceptance in computing education.

The focus of the study	Basic Theory	Variable	Result	Reference
To examine students' acceptance and perceptions of simulation software technology in a fashion merchandising course and its effect on critical thinking skills.	Technology Acceptance Model (TAM)	The model includes perceived usefulness, ease of use, enjoyment, attitude toward technology use, and improvements in critical thinking skills.	Findings show that students who perceived the simulation software as useful, easy to use, and enjoyable had more positive attitudes and reported improvements in critical thinking skills; enjoyment had the strongest influence on attitudes.	(Yu, 2017)
The study focuses on examining students' acceptance of simulation-based learning (SBL) using an extended Technology Acceptance Model (TAM) within nursing and prehospital emergency care education	Technology Acceptance Model (TAM)	Attitude toward use, behavioral intention, perceived ease of use, perceived usefulness, subjective norm, facilitating conditions, self-efficacy, and fidelity	Attitude, self-efficacy, fidelity, and subjective norm significantly influenced students' acceptance of SBL	(Lemay, et al., 2018)
This study examines the impact of simulation-based virtual laboratories on the performance of engineering students, specifically in the context of Electrical Engineering education in Australia	Technology Acceptance Model (TAM)	Perceived ease of use (PEOU), perceived usefulness (PU), instrumentation (INSTR), creativity and innovation (CI), and performance impact (IMPT)	Perceived usefulness and instrumentation are the most significant factors impacting students' performance	(Altalbe., 2019)
The study investigates how student self-efficacy affects learning outcomes in a business simulation mobile game course, comparing hierarchical teaching and general teaching methods	Self-System Model of Motivational Development (SSMMD)	Self-efficacy, student engagement, and learning outcomes	Self-efficacy directly and indirectly affected learning outcomes through engagement in the hierarchical method	(Yang, et al., 2020)
The study examines simulation-based education (SE) and its application in STEM fields across different European universities, including both online and on-campus models	-	Student motivation, engagement, skill acquisition, as well as practical knowledge of complex systems like logistics and engineering processes	The study finds that SE significantly benefits student engagement, problem-solving skills, and practical knowledge, though it requires careful implementation to avoid distraction and ensure effective learning	(Campos, et al., 2020)
The study focuses on the effectiveness of simulation-based learning in higher education, specifically its role in developing complex skills across	A Meta-Analysis	The type of simulation, technology use, duration, authenticity, and scaffolding methods, along with the	The study finds that simulation-based learning is highly effective in fostering complex skills, especially when using high authenticity simulations and targeted scaffolding	(Chernikova, et al., 2020)

diverse academic domains		learners' prior knowledge and familiarity with the subject matter	approaches, though the effects vary based on learners' prior knowledge and the specific instructional supports used	
To evaluate the effectiveness of immersive virtual reality (I-VR) as an educational tool through a systematic review of quantitative learning outcomes and experimental design in various academic fields.	Technology Acceptance Model and Bloom's taxonomy of educational objectives	Key variables include cognitive learning outcomes, procedural skills, affective learning outcomes, intervention characteristics, assessment measures, and methodological quality scores of the studies reviewed.	The study finds that I-VR generally improves cognitive learning outcomes, particularly in fields requiring spatial understanding, although its benefits vary by subject and study design.	(Hamilton. et al., 2020)
The study investigates the effects of simulation-based learning (SBL) on nursing students' perceived competence, self-efficacy, and learning satisfaction across repeated exposures	Kolb's Experiential Learning Theory	Nursing competence, self-efficacy, and learning satisfaction	Repeated simulation exposures significantly improved nursing students' competence, self-efficacy, and learning satisfaction	(Hung et al., 2021)
This study examines the factors influencing high school students' acceptance of a 3D simulation Android app for learning physics as a form of mobile learning	Technology Acceptance Model (TAM)	Perceived Usefulness, Perceived Ease of Use, Perceived Enjoyment, and Behavioral Intention	Perceived Enjoyment and Perceived Usefulness positively influenced students' intention to use the app	(Lisana, L., & Suciadi, 2021).
To investigate students' perceptions and acceptance of simulation games as a learning tool in higher education, particularly in STEM learning contexts	Human cognition and information processing theories	Students' perceptions of simulation games as valid representations of reality, the application of theoretical knowledge, and ease of use of the game interface	The study found that students generally have positive perceptions of simulation games as learning tools, with high satisfaction levels and perceived learning benefits.	(Wong, et al., 2022)
The study evaluates the impact of simulation tools, especially Cisco Packet Tracer, on enhancing practical computer networking skills in a resource-constrained higher education context in South Africa.	The CIPP (Context, Input, Process, Product) model.	Practical skill acquisition, theoretical understanding, tool effectiveness, infrastructure adequacy, and learning outcome sustainability.	The study concludes that simulation tools significantly improve practical skills, theoretical comprehension, and preparedness for professional networking work, despite occasional software and compatibility challenges.	(Mwansa, et al., 2024)

Table 1. Overview of prior studies

2.1. Literature Study and Hypotheses Development

This study uses the Technology Acceptance Model (TAM) as the foundation for developing a proposed theoretical model to explain student acceptance of Learning Simulation Tools. This framework comprises two main constructs: Perceived Usefulness (PU) and Perceived Ease of Use (PEOU). These constructs address two critical questions in a required lab setting: whether the simulator effectively aids students in learning the subject and whether students can use it with minimal effort. In settings where simulation tools are mandatory components of the curriculum (e.g., virtual labs, training simulators), PEOU and PU remain the primary levers of acceptance, while constructs such as Facilitating Conditions (from UTAUT) become relatively constant, as all students have access to the same equipment and support.

Previous studies have also confirmed that TAM is an effective framework for exploring the acceptance of simulation tools in educational settings through its two key components, PU and PEOU (Yu, 2017; Lemay et al., 2018; Altalbe, 2019; Hamilton et al., 2020; Lisana & Suciadi, 2021). These two variables have been shown to have a strong influence on users' attitudes and intentions regarding the use of specific technologies (Fussell & Truong, 2022). In the context of learning, Bagdi et al. (2023) highlighted that TAM effectively explains how students adopt specific technologies in education. Based on this, we propose that both perceived usefulness and ease of use significantly influence students' positive perceptions of learning outcomes when using simulation tools in computing education courses. Perceived usefulness refers to the belief that using simulation tools will help students understand the learning material better (Lisana & Suciadi, 2021). This indicates that learning simulation tools can significantly benefit student learning activities. Meanwhile, ease of use refers to the degree to which an individual believes that using a particular system will require minimal effort (Estriegana et al., 2019). This implies that users can expect that utilizing simulation tools in their learning activities will not require excessive effort to become familiar with them (Fussell & Truong, 2022). Thus, the following hypotheses are proposed:

H1. Perceived usefulness positively and significantly affects learning outcomes from learning simulation tools in computing education courses

H2. Perceived ease of use positively and significantly affects learning outcomes from learning simulation tools in computing education courses

In educational environments, social pressure from influential individuals—particularly those students consider important and significant- influences their perceptions of technology use (Nofita et al., 2024). Binyamin et al. (2018) note that in high school settings, students' attitudes toward certain technologies are shaped by the opinions of individuals they respect, such as teachers and close friends. This concept is referred to as Subjective Norms. In university contexts, Ermilinda et al. (2024) demonstrate that the behavior of lecturers has a significant impact on students' technology usage. Consequently, this study adopts the definition of Subjective Norms from Kim et al. (2021) and adjusts it for the context of this research, defining it as the extent to which students believe their lecturers think they should use simulation tools. The study, therefore, posits that subjective norms have a substantial effect on users' perceptions of the usefulness and ease of use of these tools, consistent with previous research (Aji et al., 2020; Al Kurdi et al., 2021; Kim et al., 2021), lead to following hypotheses:

H3: Subjective Norms positively and significantly affect Perceived usefulness

H4: Subjective Norms positively and significantly affect Perceived ease of use

A recent study in higher education emphasizes the importance of positive support in student learning activities, which fosters effective teaching methods (Wilson et al., 2024). Specifically, support from individuals in the academic environment, such as peers, teaching assistants, and professors, can have a positive effect on students, enhancing their confidence in conducting academic activities (Khan et al., 2024). This support, known as social support, refers to an individual's perception that they are cared for, valued, and part of a mutually supportive community (Wang et al., 2023). In the context of this study, when students believe they will receive help or support from their peers, teaching assistants, or lecturers when facing challenges in their academic activities, it enhances their positive perception of using learning tools. This assertion is supported by Shen et al. (2006), who indicate that the influence of users in different roles within higher education, such as lecturers, peers, or teaching assistants, affects students' perceptions of the perceived usefulness and perceived ease of use of learning tools. According to Huang and Zhang (2022), this construct also explains the perceived available assistance, or the actual support received, which can increase positive feelings, especially when students encounter challenges in their studies. Thus, this study believes that social support from individuals in different roles, such as lecturers or teaching assistants, enhances students' perceptions of system usage, including its usefulness and ease of use. Therefore, this study proposes the following hypothesis:

H5: Social Support positively and significantly affects Perceived usefulness

H6: Social Support positively and significantly affects Perceived ease of use

One significant factor in determining the quality of a simulation program is fidelity. This concept refers to the extent to which a virtual environment resembles the real world (Jiang et al., 2024; Wen & Wang, 2020). Specifically, it explains how closely simulation tools can replicate the original experience found in reality (McMahan et al., 2012). According to Mahalil et al. (2020), this aspect influences user acceptance of the tools, including their perceived usefulness and ease of use. Jiang et al. (2024) emphasize that the level of fidelity will affect the tools' ability to deliver better outcomes, ultimately leading to improved performance expectations for simulator tools. Therefore, we propose the following hypothesis:

H7: Fidelity positively and significantly affects Perceived usefulness

H8: Fidelity positively and significantly affects Perceived ease of use

Prior research highlights the impact of personal traits on students' acceptance of technology usage in higher education, particularly their self-belief in using specific systems. Ermilinda et al. (2024) argue that students with confidence in their ability to utilize a particular system are more likely to maximize the platform's benefits for their learning activities. This concept is known as self-efficacy, which refers to "people's judgment of their capabilities to organize and execute courses of action required to attain designated types of performance" (Bandura, 1977). In the context of simulation education tools, Xie et al. (2022) emphasize the importance of self-efficacy in system usage. Specifically, self-efficacy helps students mitigate negative perceptions regarding the complexity of successfully using a system (Lisana & Handarkho, 2024). Individuals with high self-efficacy also tend to believe in their likelihood of succeeding while using a particular system (Cardullo et al., 2021). Additionally, Ali et al. (2021) note that self-efficacy influences students' perceptions of the effort required to use the system for their learning activities effectively. Fathema et al. (2015) also state that if individuals doubt their ability to use a system, they are likely to view it as less useful and more difficult to use. This study, therefore, proposes the following hypothesis:

H9: Self Efficacy positively and significantly affects Perceived usefulness

H10: Self Efficacy positively and significantly affects Perceived ease of use

Other factors that influence technology adoption include innovativeness. In an educational context, innovativeness refers to students' willingness to take on challenges, explore new ideas, and seek out additional learning opportunities (Wang & Lin, 2021). Maki et al. (2016) note that individuals who possess this trait often adopt innovative technology on a daily basis. According to Kim et al. (2021), people with this characteristic are more likely to embrace systems that offer unique or novel approaches. In this study, we believe that students who exhibit higher levels of innovativeness are more likely to embrace learning technologies such as simulation tools. This embrace leads to positive perceptions of these technologies, including their usefulness and ease of use (Wu & Liu, 2023; Akour et al., 2022). Based on this, we formulate the following hypothesis:

H11: Innovativeness positively and significantly affects Perceived usefulness

H12: Innovativeness positively and significantly affects Perceived ease of use

3. Research Design and Methodology

This study conducted an online, cross-sectional survey of university students who had used a simulation tool as part of their learning. The participants were students in their first, second, and third years of study in informatics or computer science. They had completed courses that utilized simulation tools, including computer networking courses that used Cisco's Packet Tracer simulator and iLabs for the Certified Secure Computer User (CSCU) certification. We selected participants using the purposive sampling method based on the criteria outlined in Neuman's (2014) sampling frame. This approach was necessary to ensure that participants had experience with simulation tools in their university computing courses. The focus constructs of the proposed model (perceived usefulness, ease of use, and learning outcomes) can only be assessed by students who have actually worked with a simulation tool. If novices were asked to respond hypothetically, it would weaken construct validity. Without hands-on experience, respondents could only guess how useful or easy the tool might be. Thus, hypothetical questions may introduce systematic response bias, known as hypothetical bias, leading to measurement error (Kaderabek & Sinibaldi, 2022). However, we recognize that excluding inexperienced students could result in self-selection bias, which might inflate favorable ratings and reduce the spread of scores. Therefore, we invited every eligible student to participate in order to capture the widest possible range of experiences, and we acknowledge this limitation for future research that could include first-time users or pre- or post-exposure designs.

The sample size for this study was determined using the formula provided by Kline (2016), which specifies a minimum of 20 respondents for each factor in the model. Since our model includes eight latent constructs, this necessitates at least $8 \times 20 = 160$ respondents. However, Kline (2016) further recommends a minimum of 200 cases for stable estimation in structural equation modeling (SEM), the method employed in this study. The confirmatory factor analysis (CFA) involved calculating the average variance extracted (AVE) and composite reliability (CR) to ensure the convergent validity of the data, following the criteria proposed by Fornell and Larcker (1981). In addition, George and Mallery's (2003) guidelines were used to assess data reliability through the coefficients of Cronbach's alpha. Meanwhile, discriminant validity was evaluated by ensuring that the square roots of the AVE values were greater than the correlations among the latent variables (Barclay et al., 1995). Lastly, this study used AMOS software to analyze the data and validate the proposed effects in the theoretical model through SEM based on guidance from Kline (2016).

In line with recent SEM guidelines (Kline, 2016), we estimated our model using covariance-based SEM in AMOS with a latent structural-regression (LSR) specification. This approach was chosen over alternatives such as path analysis, partially latent structured regression (PLSR), or variance-based PLS-SEM. LSR treats all seven theoretical constructs as fully latent variables, each measured by a complete set of reflective indicators. This enables us to test the entire Technology Acceptance Model extension, covering both measurement and structural relationships, within a single likelihood-based framework. This choice aligns with our goal of theory confirmation. CB-SEM/LSR provides global fit indices (e.g., χ^2, CFI, RMSEA) that are essential for evaluating how well the hypothesized model reproduces the observed covariance matrix. In contrast, PLS-SEM primarily focuses on prediction and is mainly recommended for formative or exploratory research (Schumaker & Lomax, 2016). Furthermore, the complexity of our model, including eight latent constructs and twenty-seven indicators, fits well within the analytical scope of LSR. Utilizing path analysis or PLSR would oversimplify the measurement component by collapsing some latent variables into single composite scores. Additionally, we set threshold values for skewness and kurtosis at less than three and seven, respectively, to meet the distributional assumptions for covariance-based SEM (Kline, 2016). All these factors confirm that covariance-based SEM with an LSR measurement model is the most rigorous and appropriate approach for testing our theoretically grounded hypotheses. Finally, a consent form has been incorporated into the survey to guarantee that all participants have given informed consent to engage in the study. This research utilized a voluntary and anonymous survey that did not capture sensitive or personal data. We ensured adherence to ethical norms by obtaining explicit consent from individuals.

4. Theoretical Model and Measurement

Figure 1 shows the theoretical model. Meanwhile, all the measuring instruments used to validate the model can be seen in Table 2. The questionnaire was adopted from several prior studies and developed with input from a focus group of student representatives who have varying levels of experience with simulation tools, ranging from novice to expert. To ensure the accuracy and contextual relevance of the Indonesian version of the questionnaire, bilingual experts with experience in this type of research were involved in the translation process. Additionally, a pilot study was conducted to gather feedback from selected respondents, which helped to refine the questionnaire to better align with the specific objectives of the study.

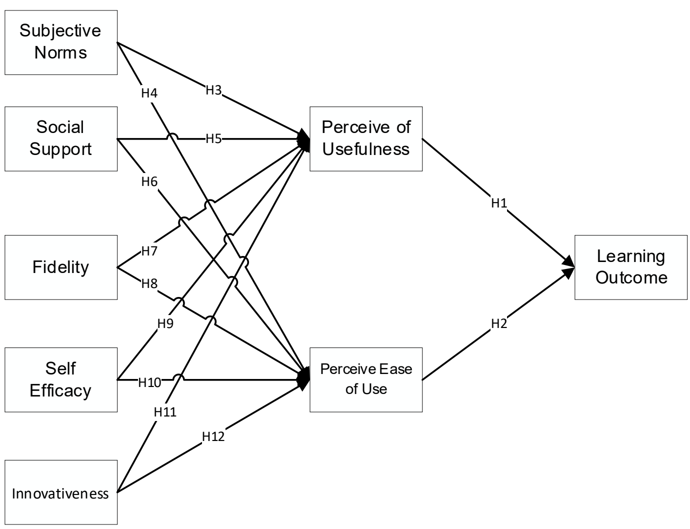

Figure 1. The theoretical model

Variable (Symbol)	Indicator	Measuring Instrument	Adopted from
Learning outcomes	LO1	The simulation tool made me learn to apply theory to practice	(Yang et al., 2022)
	LO2	This simulation tool helped me understand the course material	
	LO3	The simulation tool gave me insight into the course material	
Ease of Use	EOU1	Learning how to use the simulation tool is easy.	(Hong et al., 2006)
	EOU2	The simulation tool is clear and understandable to use.	
	EOU3	I find the simulation tool easy to use.	
Perceived usefulness	PU1	The use of the simulation tool to fulfill my learning activities will improve my performance.	(Handarkho, 2020)
	PU2	The use of the simulation tool to fulfill my learning activities will improve my effectiveness.	
	PU3	The use of the simulation tool helps me to carry out my learning activities.	
	PU4	In general, the use of the simulation tool is beneficial for fulfilling my learning activities.	
Subjective Norm	SN1	The simulation tool is important for my learning in class	(Binyamin et al., 2018)
	SN2	The lecturer thought that I needed the simulation tool to help me study.	
	SN3	I want to do what my lecturer thinks I should do.	
	SN4	The lecturer thinks that with the simulation tool, my learning will increase.	
Social support	SS1	When I am faced with difficulties, some people (peers, lecturer, lecturer assistant) in class comfort and encourage me.	(Shanmugam et al., 2016)
	SS2	When I am faced with difficulties, some people (peers, lecturer, lecturer assistant) in class show interest and concern for my well-being.	
	SS3	In the class, some people (peers, lecturer, lecturer assistant) offer suggestions when I need help.	
	SS4	When I encounter a problem, some people in class (peers, lecturer, lecturer assistant) give me information to help me overcome the problem.	
Fidelity	FD1	The scenario used in the simulator resembled a real-life situation	(Lemay et al., 2018)
	FD2	Real-life factors, situations, and variables were built into the simulation scenario.	
Self-efficacy	SE1	I believe that I can use the simulation tool to get the learning information I need.	(Chen & Tseng, 2012)
	SE2	I believe that I can use the simulation tool to unlock lecturer-given assignments.	
	SE3	I believe that the experience when I use the simulation tool help me to take quizzes given by the lecturer.	
Innovativeness	IN1	If I find out about a new technology, I seek ways to experience it	(Handarkho, & Harjoseputro, 2020)
	IN2	I can usually figure out new technology without help from others	
	IN3	I enjoy the challenge of figuring out a new technology	
	IN4	In general, I am among the first in my circle of friends to acquire new technology when it appears	

Table 2. Indicators and measuring instrument

5. Data Preparation & Descriptive Analyses

A total of 312 responses were collected to validate the proposed model. We conducted a Confirmatory Factor Analysis (CFA) using AMOS software to determine the loading factors, which were then used to calculate the Average Variance Extracted (AVE) and Composite Reliability (CR) to assess convergent validity. The results indicate that all values of AVE and CR meet the minimum thresholds established by Fornell and Larcker (1981), as shown in Table 3. Additionally, Cronbach's alpha coefficient also meets the standard values proposed by George and Mallery (2003), demonstrating that the collected responses exhibit data reliability.

However, the results for discriminant validity (Table 4) show a cross-correlation between the constructs of Ease of Use and Learning Outcome, suggesting that these two constructs are highly correlated and may overlap significantly. This implies that they may not be distinct from each other as originally intended (Fornell & Larcker, 1981). To thoroughly investigate these issues, we conduct the HTMT (Heterotrait–Monotrait Matrix Test), which has a higher power to detect problematic overlaps by analyzing the factor loadings associated with each construct (Henseler et al., 2015). After running every combination of indicators for Ease of Use (EOU1 to EOU3) and Learning Outcome (LO1 to LO3), while ensuring that we maintain at least two indicators per construct (the minimum typically recommended for Structural Equation Modeling), the results indicate that all HTMT values surpass the 0.90 threshold. This is above the commonly accepted cut-off for discriminant validity, where HTMT values should be below 0.85 for conceptually distinct constructs or below 0.90 for constructs that are more closely related, as suggested by Henseler et. al. (2015), even after excluding indicators with loadings below 0.70.

Indicator	Factor loadings	AVE	CR	CA	Indicator	Factor loadings	AVE	CR
LO1	0.521	0.576	0.796	0.831	SS1	0.784	0.687	0.898
LO2	0.835				SS2	0.85		
LO3	0.871				SS3	0.843		
EOU1	0.859	0.760	0.907	0.813	SS4	0.837		
EOU2	0.884				FD1	0.757	0.652	0.789
EOU3	0.426				FD2	0.855		
PU1	0.794	0.671	0.891	0.889	SE1	0.856	0.704	0.877
PU2	0.839				SE2	0.842		
PU3	0.833				SE3	0.818		
PU4	0.809				IN1	0.74	0.517	0.810
SN1	0.79	0.597	0.856	0.854	IN2	0.647		
SN2	0.749				IN3	0.765		
SN3	0.752				IN4	0.718		
SN4	0.799							

Note: CA refers to Cronbach's alpha

Table 3. Factor analysis and Cronbach's alpha coefficient.

	LO	EOU	PU	SN	SS	FD	SE	IN
Learning outcomes	0.759							
Ease of Use	.911**	0.872						
Perceived usefulness	.626**	.702**	0.819					
Subjective Norm	.578**	.633**	.711**	0.773				
Social support	.404**	.431**	.547**	.541**	0.829			
Fidelity	.487**	.521**	.544**	.614**	.538**	0.807		
Self-efficacy	.570**	.620**	.658**	.692**	.563**	.621**	0.839	
Innovativeness	.452**	.502**	.507**	.507**	.443**	.576**	.563**	0.719

Notes: **. Correlation is significant at the 0.01 level (2-tailed); Highlighted columns indicate a cross-correlation between the constructs of Ease of Use and Learning Outcome

Table 4. Discriminant validity

This finding confirms that the overlap between Learning Outcome (LO) and Ease of Use (EOU) is structural rather than being driven by one or two anomalous items. Table 5 shows all the HTMT ratios for each subset that retains at least 2 items per construct. Consequently, this study has decided to drop the construct of Ease of Use since the Learning Outcome is the dependent variable we aim to explain. As a result, we have adjusted the proposed theoretical model, as illustrated in Figure 2, which details the final indicator and factor loadings values presented in Table 6. The final indicator still maintains a factor loading value under 0.7 (LO1). Loadings as low as 0.50 or 0.60 are acceptable if the Average Variance Extracted (AVE) exceeds 0.50 and the Composite Reliability (CR) is above 0.70. These criteria indicate sufficient convergent validity and internal consistency reliability (Hair et al., 2010).

LO items kept	EOU items kept	HTMT for LO–EOU
LO1, LO2	EOU2, EOU3	0.98
LO1, LO3	EOU1, EOU3	1.04
LO1, LO2	EOU1, EOU2, EOU3	1.06
LO1, LO2	EOU1, EOU2	1.08
LO1, LO2, LO3	EOU1, EOU3	1.09
LO1, LO2, LO3	EOU2, EOU3	1.1
LO1, LO2	EOU1, EOU3	1.1
LO1, LO2, LO3	EOU1, EOU2, EOU3	1.11
LO2, LO3	EOU1, EOU3	1.12
LO1, LO3	EOU1, EOU2, EOU3	1.12
LO2, LO3	EOU2, EOU3	1.15
LO1, LO2, LO3	EOU1, EOU2	1.15
LO1, LO3	EOU1, EOU2	1.15
LO2, LO3	EOU1, EOU2, EOU3	1.16
LO1, LO3	EOU2, EOU3	1.17
LO2, LO3	EOU1, EOU2	1.22

Table 5. Sub HTMT ratios for LO and EOU

Indicator	Factor loadings	AVE	CR	CA	Indicator	Factor loadings	AVE	CR
LO1	0.521	0.576	0.796	0.831	SS1	0.784	0.687	0.898
LO2	0.835				SS2	0.85		
LO3	0.871				SS3	0.843		
PU1	0.794	0.671	0.891	0.889	SS4	0.837		
PU2	0.839				FD1	0.757	0.652	0.789
PU3	0.833				FD2	0.855		
PU4	0.809				SE1	0.856	0.704	0.877
SN1	0.79	0.597	0.856	0.854	SE2	0.842		
SN2	0.749				SE3	0.818		
SN3	0.752				IN1	0.74	0.517	0.810
SN4	0.799				IN2	0.647		
					IN3	0.765		
					IN4	0.718		

Table 6. Final factor loading, with AVE, CR, and Cronbach's alpha value.

Table 7 presents the detailed characteristics of the respondents. The respondents consist of first-year students (32.1%), second-year students (35.95%), and third-year students (32.15%) who are currently attending classes. Additionally, male respondents dominate the group, making up 80% of the total. All the students have experience with simulation tools in their computing course at the university, indicating they are qualified to be part of the research.

Further, based on the profiles of the respondents, we conducted T-tests and ANOVA to analyze significant differences in mean (M) scores based on gender and student semester, respectively. The results from the T-test revealed a significant difference between males and females, particularly in terms of innovativeness.

Males had a significantly higher mean score (M = 3.87) compared to females (M = 3.57) in this area. While both genders showed a positive response, male innovativeness was notably more pronounced, especially regarding their intention to explore new technology, as indicated by items IN1, IN3, and IN4, which reflect their enjoyment in discovering and acquiring new technological tools. Meanwhile, the ANOVA results indicated significant differences in the mean values for the Fidelity factor. First-year students (M = 4.09) considered this aspect more important than second-year (M = 3.83) and third-year students (M = 3.84), particularly regarding how well the simulator reflects real-life situations (FD1).

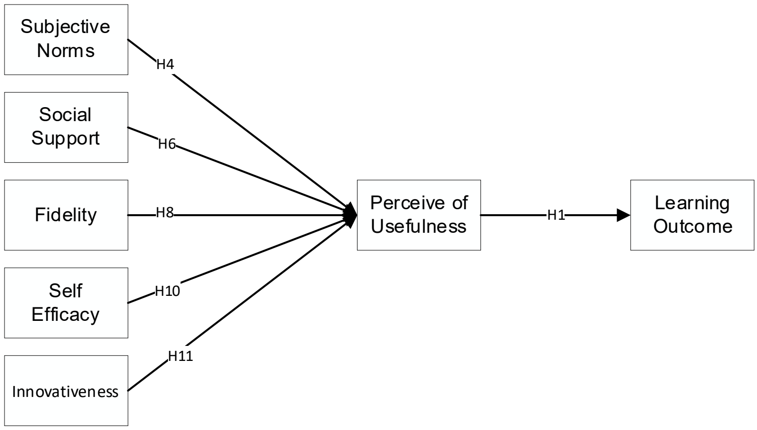

Figure 2. The adjusted theoretical model

Lastly, Skewness and kurtosis values were calculated to ensure the data were suitable for SEM. The recorded values were less than three and seven, respectively, thus satisfying the requirement (Kline, 2016).

		Frequency	Percent	Valid Percent	Cumulative Percent
Age	< = 20	236	75.6	75.6	75.6
	> 20	76	24.4	24.4	100.0
	Total	312	100.0	100.0	
Semester	1-2	100	32	32	32.
	3-4	112	36	36	68
	5-6	100	32	32	100.0
	Total	312	100.0	100.0	
Gender	Male	252	80.8	80.8	80.8
	Female	60	19.2	19.2	100.0
	Total	312	100.0	100.0	

Table 7. Respondents' characteristics

6. Result of Direct Effects

The results of the Structural Equation Modeling (SEM) analysis are presented in Table 8 and Figure 3. The analysis revealed that the direct effect of Subjective Norms on Perceived Usefulness is the strongest in the model, followed by Self-Efficacy and Social Support. Additionally, Perceived Usefulness was found to significantly predict students' perceptions of learning outcomes resulting from the use of simulation tools. However, two of the proposed hypotheses, specifically those related to Fidelity and Innovativeness, were found to be insignificant. Furthermore, since we removed the Ease-of-Use construct from the model, all hypotheses associated with this factor (H2, H4, H6, H8, H10, and H12) were also excluded from this study. Meanwhile, Table 7 indicates that the data fit the model statistically, based on Kline's (2016) model fit criterion.

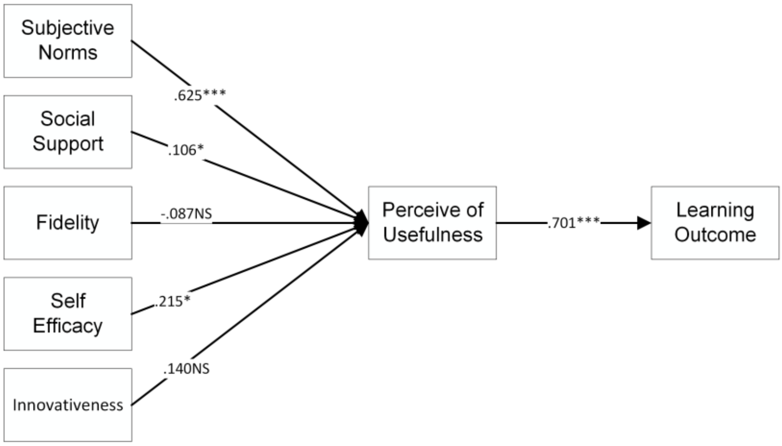

Figure 3. The result of the direct effect in the Theoretical Model

Direct effect	Total Effect	Status
Usefulness ➔ Learning Outcome (H1)	.701***	Accepted
Subjective Norms➔ Usefulness (H4)	.625***	Accepted
Social Support ➔ Usefulness (H6)	.106*	Accepted
Fidelity ➔ Usefulness (H8)	-.087NS	Rejected
Self-efficacy ➔ Usefulness (H10)	.215*	Accepted
Innovativeness ➔ Usefulness (H12)	.140NS	Rejected

Table 8: Hypothesis testing results

Sample Size	Normed chi-square (NC) = $\chi 2/df$	RMR	SRMR	GFI	AGFI	NFI	IFI	CFI	TLI	RMSEA
312	424.219/236 = 1.798	.018	.037	.901	.874	.914	.960	.960	.953	.051
	R^2: LO: 0.567; PU:0.727									

Note(s): R^2 is the proportion of the variance explained by the variables affecting it

Table 9. Fit statistic for the proposed model

Table 9 demonstrates that the structural model fits the data well according to all recommended indices. The normed chi-square is 1.80 (χ^2 = 424.219, df = 236), which is comfortably below the conservative ceiling of 3.0, indicating an acceptable fit (Kline, 2016). Both absolute and incremental indices show strong results: the Goodness-of-Fit Index (GFI = .901) exceeds the guideline of .90 (Byrne, 2010), while the Adjusted GFI (AGFI = .874) surpasses the benchmark of .85 for complex models (Hooper, Coughlan, & Mullen, 2008). Incremental measures, including the Normed Fit Index (NFI = .914) and the Incremental Fit Index (IFI = .960), both exceed the threshold of .90, indicating a significant improvement over the null model (Hu & Bentler, 1999). Additionally, the Comparative Fit Index (CFI = .960) and the Tucker-Lewis Index (TLI = .953) both surpass the .95 criterion for a close fit (Hu & Bentler, 1999). Residual-based indices confirm minimal misfit, with a raw root mean square residual (RMR) of .018, where values less than or equal to .05 are considered good. Moreover, the root mean square error of approximation (RMSEA) is .051, with values below .06 indicating a good fit, and the standardized RMR (SRMR) is .038, which is well below the ceiling of .08 (Hu & Bentler, 1999). The model's explanatory power is also significant, accounting for 72.7% of the variance in Perceived Usefulness (R^2 = .727) and 56.7% in Learning Outcome (R^2 = .567). Both proportions exceed Cohen's (1988) "substantial" benchmark of .26.

7. Discussion

The results indicate that social factors play a crucial role in the successful adoption of simulation tools in computing education courses. The findings suggest that lecturers' opinions significantly influence students' perceptions of the usefulness of these tools (H4). In the context of Indonesian higher education, characterized by high power distance and a strong respect for authority, this influence is even more pronounced. When lecturers incorporate tools like Cisco Packet Tracer and iLabs into their lectures, they not only demonstrate technical workflows but also convey institutional legitimacy. This, in turn, enhances student motivation and engagement (Ermilinda et al., 2024; Kim et al., 2021). For instance, one lecturer's enthusiastic demonstration of a virtual network configuration inspired even the most skeptical students to explore advanced features. This suggests that lecturer enthusiasm acts as both an informational and normative influence. These findings support the work of Binyamin et al. (2018), who identified respect for instructors as a key factor in technology acceptance. Additionally, this research expands upon theirs by highlighting the cultural aspects of authority in collectivist contexts.

The findings further show that support from peers, teaching assistants, and professors can positively impact students when they encounter difficulties using simulation tools in their learning activities (H6), which confirms Khan et al. (2024). This indicates that when students believe they will receive support from their peers, teaching assistants, or lecturers while facing academic challenges, it positively affects their perception of learning tools. This notion is backed by Shen et al. (2006), who argue that the involvement of various individuals in higher education influences students' perceptions of the usefulness of these tools. Furthermore, the availability of assistance and the actual support received can enhance students' positive feelings, especially when they encounter difficulties in their studies (Huang and Zhang, 2022).

From a personal perspective, although dispositional innovativeness (H11) was not a significant predictor, self-efficacy (H10) had a strong influence on perceived usefulness. This finding suggests that students' confidence in their ability to use the system impacts their perception of its usefulness (Lisana & Handarkho, 2024; Fathema et al., 2015). It indicates that students who are confident in their ability to navigate a particular system are more likely to fully leverage the platform's benefits for their learning activities (Ermilinda et al., 2024). This result confirms that students with high self-efficacy are more likely to believe in their chances of success when using a specific system (Cardullo et al., 2021).

Meanwhile, the results for H11 indicate that the effect of innovativeness was rejected. This result contradicts prior studies that suggest this construct encourages individuals to use innovative systems (Kim et al., 2021). This difference may be attributed to the tendency for innovation to be linked with technology used outside of education, such as for financial and entertainment purposes. Our respondent, who uses simulator tools as part of the course, may influence the results. When a tool is required rather than selected freely, students' dispositional innovativeness might play a smaller role (Brown et. al., 2002). This could lead them to focus more on adapting to the tool instead of actively exploring its novel features. Furthermore, the effect of fidelity of simulation tools for technology quality is also not supported (H7), which contradicts previous studies (Mahalil et al., 2020; Jiang et al., 2024). This outcome may be attributed to the context of computing education courses, where the realism of simulations is seen as a supplementary feature. In these cases, the primary focus of students is to understand the course material rather than to experience the realism of the simulation. From a statistical perspective, since all students used the same simulator (Cisco Packet Tracer and Ilabs simulator) and had similar exposure to the course, perceived realism ("Fidelity") was uniformly high, creating a ceiling effect that weakened the correlations (Cohen, 1988).

Overall, our findings indicate that the acceptance and effective use of simulation tools primarily depend on supportive social aspects. When lecturers integrate simulations into course objectives and demonstrate their use, they legitimize the technology and set clear expectations for students. Moreover, encouraging collaboration among peers and ensuring that teaching assistants are readily available creates a positive learning environment. In such an environment, students can troubleshoot, exchange strategies, and celebrate their achievements together. This social support enhances their learning experiences and boosts their confidence, making simulation tools vital components of the learning process. Consequently, students feel empowered to experiment, refine their approaches, and confidently grasp complex computing concepts.

The theoretical contribution of this study lies in its comprehensive approach to explaining the factors that influence student acceptance of simulation tools in computing education. Specifically, this research develops a theory based on the Technology Acceptance Model (TAM) and extends it by incorporating environmental aspects and personal characteristics of students. While several previous studies have focused on theoretical frameworks that emphasize the capabilities of simulation software in enhancing the student learning experience from a technological perspective, others have discussed student cognitive processes to explain their perceptions and acceptance of learning tools in educational settings. In contrast, this study offers an alternative approach that also considers the influence of environmental factors, such as lecturers, peers, and teaching assistants, which prior research has yet to explore in depth. This study addresses this gap in the literature. The results indicate that social factors are crucial in students' acceptance of alternative teaching and learning methods in computing education courses that use simulation tools. Consequently, this finding contributes to the theoretical framework of this research area.

Several practical actions can be proposed to enhance students' perception of the usefulness of simulation tools in supporting their learning activities in class. Based on the results regarding subjective norms, lecturers play a crucial role in encouraging students to maximize the use of these tools. Therefore, lecturers need to demonstrate the value of the tools by integrating them into course objectives. By aligning the tools with syllabus guidelines, students will recognize their importance in the course and appreciate their usefulness in the learning process. The results also indicate the importance of support from peers, teaching assistants, and lecturers when students face challenges in their academic activities, particularly in using simulation tools. Faculty or departments should consider creating study groups or incorporating the adoption of these tools as a topic in study sessions. This way, upper-year students can share their experiences with newcomers regarding the use of simulation tools. Ideally, through these groups, students can assist one another in maximizing the effectiveness of the tools to enhance their learning. Additionally, lecturers should ensure that their teaching assistants are properly trained to provide hands-on guidance in using simulation tools during class sessions. This will allow them to assist students who face challenges in operating the simulation tools. To enhance self-efficacy among students, lecturers can provide training and tutorials outside of regular class sessions, particularly for newcomers. This support can be facilitated through student study groups or additional classes, helping students become more familiar with using various tools. The goal is for students to gain confidence in utilizing simulation tools. Lecturers should also design their courses to incorporate these tools gradually. They can start by introducing simpler cases or assignments to help students become familiar with them. By beginning with basic course material that is simple enough, students can build their confidence and reduce any intimidation they may feel when using these tools.

However, even though the effect of fidelity is rejected, the ANOVA result indicates that 1st years students give more attention to this construct. This pattern suggests that newcomers place more weight on surface realism, perhaps because they lack hands-on experience. Therefore, the department needs to design first-year simulation exercises with relatable real-world scenarios, such as simulating a home or small office network setup, along with visual cues and step-by-step tasks that mirror actual field conditions, helping students better grasp how course concepts apply in practical environments.

In order to convert our findings into improvements that benefit all stakeholders, we recommend several interconnected measures. Instructors should incorporate simulation exercises into course objectives and assessment criteria, highlighting their importance and encouraging student engagement. By aligning each simulation with specific syllabus goals, students can view these tools as essential rather than supplementary. Further, given the strong influence of subjective norms on acceptance, departments can bolster peer support by establishing informal study groups or mentoring clinics where senior students offer practical advice to newcomers. Additionally, teaching assistants should receive specialized training to provide practical help during laboratory sessions. Instructors, meanwhile, can boost students' self-efficacy by offering brief optional lessons outside regular class hours and by structuring assignments that progress from low-stakes, basic tasks to more challenging scenarios. This approach allows novices to build their confidence gradually.

Next, developers of educational tools can support pedagogical efforts by optimizing user interfaces with wizard-style workflows, in-line error notifications, and context-sensitive micro-tutorials that facilitate initial use. They should also incorporate adaptive-fidelity controls, enabling instructors to switch between low- and high-fidelity modes to manage cognitive load effectively. Integrating accessibility features, such as keyboard-only navigation, text-to-speech capabilities, and bandwidth-adaptive media, will enhance participation, particularly in resource-limited environments often found in Indonesian higher education institutions. Instructional designers, meanwhile, can further improve these initiatives by implementing dynamic analytics dashboards that showcase competence heat maps. This allows educators to focus feedback on areas where simulation data indicate recurring mistakes. Furthermore, they can package simulation segments as Learning Management System (LMS) content blocks through Learning Tools Interoperability (LTI) or Application Programming Interface (API), supporting spaced-practice schedules that reinforce knowledge retention.

Together, these pedagogical, technical, and design strategies should enhance the perceived utility and user-friendliness of simulation tools while optimizing their effectiveness for learning.

Moreover, broader institutional support, such as establishing dedicated simulation laboratories with standardized quality criteria, regular system updates, and comprehensive training programs, could enhance students' engagement by providing a consistent and reliable learning environment. For instance, universities could implement formal guidelines for the periodic evaluation of simulation software quality, usability assessments involving students and lecturers, and systematic integration of industry-based scenarios. These measures would ensure that simulation tools not only meet educational objectives but also reflect real-world technological advancements.

Universities could further motivate active student adoption of simulation technologies by introducing gamified elements, such as digital badges, leaderboards, or achievement certificates linked directly to course grades or extra credits. Additionally, recognizing outstanding student performance through academic awards or showcasing their simulation projects publicly, for instance, in university exhibitions or open-day events, could reinforce positive perceptions and encourage greater student enthusiasm toward utilizing these educational technologies.

8. Limitation

Several limitations should be considered when interpreting and generalizing our findings. First, although we ensured anonymity to minimize bias, the use of self-reported data introduces the potential for common-method bias. Additionally, since participation was voluntary, students who are more confident with or interested in technology-based learning may be over-represented. This may have resulted in more favorable responses regarding perceived usefulness and learning outcomes than would be observed in a broader, more diverse student population. Second, the study was conducted within a single computing program at an Indonesian university using purposive sampling. This limits the external validity of the findings, particularly when applied to other academic disciplines or institutional contexts. For instance, fields such as medicine or aviation often use more complex simulation environments that emphasize high-fidelity interaction and realism. Moreover, our participants engaged specifically with Cisco's Packet Tracer and EC-Council's iLabs (CSCU v3), so the findings primarily reflect experiences with those platforms. We did not disaggregate responses by tool, which limits our ability to identify platform-specific effects. Third, our assessment of simulation tool quality focused primarily on fidelity, that is, how well the tools replicate real-world tasks. Other important aspects, such as usability, accessibility, or interoperability, were not examined but may also influence student acceptance and learning outcomes.

Finally, we acknowledge the theoretical significance of Perceived Ease of Use as a core construct in the Technology Acceptance Model (TAM). Due to persistent discriminant validity issues with the Learning Outcome, we made the difficult decision to remove Ease of Use from the final model. While this choice ensured construct clarity and statistical robustness, it may reduce comparability with prior TAM-based studies. Future research should consider alternative modeling strategies, such as bifactor or higher-order models, to retain this construct without compromising validity. Recognizing these methodological, contextual, and theoretical limitations is important for accurately interpreting our results. Future studies may benefit from including novice users, comparing different simulation tools, involving more diverse academic programs or institutions, and adopting broader frameworks for evaluating technology acceptance.

9. Conclusion and Future Research

This study shows how various social factors, including the roles and influences of lecturers, peers, and teaching assistants, significantly shape students' perceptions of the effectiveness of simulation tools in enhancing learning outcomes in computing education courses. It also emphasizes the importance of students' self-efficacy, which deserves particular attention. However, the impact of innovativeness and fidelity was found to be insignificant. Overall, the findings suggest that supportive social factors play a crucial role in determining the acceptance and effective use of simulation tools in computing education settings. From a theoretical standpoint, this research enhances the understanding of the factors affecting student acceptance of simulation tools by moving beyond a purely technological perspective. It takes into account environmental influences, such as those from lecturers, peers, and teaching assistants, an aspect that has not been thoroughly explored in previous research. This positions our study as a valuable contribution to the existing body of knowledge.

Future research should focus on longitudinal designs that track perceptions and learning outcomes over multiple semesters. This approach will help clarify how social influence and self-efficacy develop with continued exposure to simulation tools. Additionally, using more representative respondent profiles, considering factors such as gender, age, major, and GPA, will allow for moderating factor analysis, enhancing the findings. Furthermore, researchers could replicate and extend the proposed framework to diverse fields like nursing, finance, and language learning to determine whether the same influencing factors exist or if new ones emerge. Lastly, broadening the concept of technology quality to include aspects like system usability, cognitive load, and interaction richness will provide clearer insights into which technical attributes most significantly affect user acceptance and performance.

References

Abdulrahaman, M. D., Faruk, N., Oloyede, A. A., Surajudeen-Bakinde, N. T., Olawoyin, L. A., Mejabi, O. V., Imam-Fulani, Y. O., Fahm, A. O., & Azeez, A. L. (2020). Multimedia tools in the teaching and learning processes: A systematic review. Heliyon, 6(11), Article e05312. https://doi.org/10.1016/j.heliyon.2020.e05312

Adams, D. A., Nelson, R. R., & Todd, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. MIS Quarterly, 16(2), 227–247. https://doi.org/10.2307/249577

Aji, H. M., Berakon, I., & Riza, A. F. (2020). The effects of subjective norm and knowledge about riba on intention to use e-money in Indonesia. Journal of Islamic Marketing, 12(6), 1180–1196. https://doi.org/10.1108/JIMA-10-2019-0203

Akour, I. A., Al-Maroof, R. S., Alfaisal, R., & Salloum, S. A. (2022). A conceptual framework for determining metaverse adoption in higher institutions of Gulf area: An empirical study using hybrid SEM-ANN approach. Computers and Education: Artificial Intelligence, 3, Article 100052. https://doi.org/10.1016/j.caeai.2022.100052

Al Kurdi, B., Alshurideh, M., Nuseir, M., Aburayya, A., & Salloum, S. A. (2021). The effects of subjective norm on the intention to use social media networks: An exploratory study using PLS-SEM and machine learning approach. Advances in Intelligent Systems and Computing (Vol. 1339, pp. 581–592). Springer. https://doi.org/10.1007/978-3-030-69717-4_55

Ali, J., Marri, M. Y. K., & Khurram, S. (2021). University social responsibility and self-efficacy as antecedents of intention to use e-learning: Examining mediating role of student satisfaction. Psychology and Education, 58(2), 4219–4230.

Altalbe, A. (2019). Performance impact of simulation-based virtual laboratory on engineering students: A case study of Australia virtual system. IEEE Access, 7, 177387–177396.

Bagdi, H., Bulsara, H. P., Sankar, D., & Sharma, L. (2023). The transition from traditional to digital: Factors that propel Generation Z's adoption of online learning. International Journal of Educational Management, 37(3), 695–717.

Brinson, J. R. (2015). Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research. Computers & Education, 87, 218–237. https://doi.org/10.1016/j.compedu.2015.07.003

Brown, S. A., Massey, A. P., Montoya-Weiss, M. M., & Burkman, J. R. (2002). *Do I really have to? User acceptance of mandated technology*. European Journal of Information Systems, 11, 283–295. https://doi.org/10.1057/palgrave.ejis.3000438

Binyamin, S. S., Rutter, M. J., & Smith, S. (2018). The influence of computer self-efficacy and subjective norms on the students' use of learning management systems at King Abdulaziz University. International Journal of Information and Education Technology, 8(10), 693–699. https://doi.org/10.18178/ijiet.2018.8.10.1124

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge.

Campos, N., Nogal, M., Caliz, C., & Juan, A. A. (2020). Simulation-based education involving online and on-campus models in different European universities. International Journal of Educational Technology in Higher Education, 17, Article 8. https://doi.org/10.1186/s41239-020-0181-y

Cardullo, V., Wang, C.-H., Burton, M., & Dong, J. (2021). K-12 teachers' remote teaching self-efficacy during the pandemic. Journal of Research in Innovative Teaching & Learning, 14(1), 32–45. https://doi.org/10.1108/JRIT-10-2020-0055

Chen, H. R., & Tseng, H. F. (2012). Factors that influence acceptance of web-based e-learning systems for the in-service education of junior high school teachers in Taiwan. Evaluation and Program Planning, 35(3), 398–406. https://doi.org/10.1016/j.evalprogplan.2011.11.007

Chernikova, O., Heitzmann, N., & Stadler, M. (2020). Simulation-based learning in higher education: A meta-analysis. Review of Educational Research, 90(4), 499–541. https://doi.org/10.3102/0034654320933544

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

Ermilinda, L., Handarkho, Y. D., & Emanuel, A. W. R. (2024). Factors influencing student intention to continue using e-learning platform post COVID-19 pandemic: Case study of University of Nusa Nipa Indonesia. Procedia Computer Science, 234, 1043–1052.

Estriegana, R., Medina-Merodio, J. A., & Barchino, R. (2019). Student acceptance of virtual laboratory and practical work: An extension of the technology acceptance model. Computers & Education, 135, 1–14. https://doi.org/10.1016/j.compedu.2019.02.010

Fathema, N., Shannon, D., & Ross, M. (2015). Expanding the technology acceptance model (TAM) to examine faculty use of learning management systems (LMSs) in higher education institutions. Journal of Online Learning and Teaching, 11(2), 210–233.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50. https://doi.org/10.1177/002224378101800104

Fussell, S. G., & Truong, D. (2022). Using virtual reality for dynamic learning: An extended technology acceptance model. Virtual Reality, 26(1), 249–267. https://doi.org/10.1007/s10055-021-00554-x

George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference, 11.0 update. Allyn and Bacon.

Handarkho, Y. D., & Harjoseputro, Y. (2020). Intention to adopt mobile payment in physical stores: Individual switching behavior perspective based on push–pull–mooring (PPM) theory. Journal of Enterprise Information Management, 33(2), 285–308. https://doi.org/10.1108/JEIM-06-2019-0179

Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022a). Understanding the role of digital technologies in education: A review. Sustainable Operations and Computers, 3, 275–285. https://doi.org/10.1016/j.susoc.2022.05.004

Haleem, A., Javaid, M., Singh, R. P., & Suman, R. (2022b). Medical 4.0 technologies for healthcare: Features, capabilities, and applications. Internet of Things and Cyber-Physical Systems, 2, 12–30.

Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. Journal of Computer Education, 8(1), 1–32. https://doi.org/10.1007/s40692-020-00169-2

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, *43*, 115-135.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53–60.

Huang, L., & Zhang, T. (2022). Perceived social support, psychological capital, and subjective well-being among college students in the context of online learning during the COVID-19 pandemic. Asia-Pacific Education Researcher, 31(5), 563–574. https://doi.org/10.1007/s40299-021-00608-3

Hung, C. C., Kao, H. F. S., Liu, H. C., Liang, H. F., Chu, T. P., & Lee, B. O. (2020). Effects of simulation-based learning on nursing students' perceived competence, self-efficacy, and learning satisfaction: A repeat measurement method. Nurse Education Today, 97, Article 104725. https://doi.org/10.1016/j.nedt.2020.104725

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jiang, T., Fang, Y., Goh, J., & Hu, S. (2024). Impact of simulation fidelity on identifying swing-over hazards in virtual environments for novice crane operators. Automation in Construction, 165, Article 105580. https://doi.org/10.1016/j.autcon.2024.105580

Kaderabek, A., & Sinibaldi, J. (2022). Assessing measurement error in hypothetical questions. *Survey Practice*.

Khan, A., Zeb, I., Zhang, Y., Fazal, S., & Ding, J. (2024). Relationship between psychological capital and mental health at higher education: Role of perceived social support as a mediator. Heliyon, 10(8), Article e29472. https://doi.org/10.1016/j.heliyon.2024.e29472

Kim, E. J., Kim, J. J., & Han, S. H. (2021). Understanding student acceptance of online learning systems in higher education: Application of social psychology theories with consideration of user innovativeness. Sustainability, 13(2), Article 896. https://doi.org/10.3390/su13020896

Lemay, D. J., Morin, M. M., Bazelais, P., & Doleck, T. (2018). Modeling students' perceptions of simulation-based learning using the technology acceptance model. Clinical Simulation in Nursing, 20, 28–37.

Lisana, L., & Handarkho, Y. D. (2024). The effects of environmental factors on user's personal traits related to mobile payment adoption: A case study of Indonesia. Global Knowledge, Memory and Communication. Advance online publication. https://doi.org/10.1108/GKMC-01-2024-0036

Lisana, L., & Suciadi, M. F. (2021). The acceptance of mobile learning: A case study of 3D simulation Android app for learning physics. International Journal of Interactive Mobile Technologies, 15(17), 4–17.

Mahalil, I., Yusof, A. M., & Ibrahim, N. (2020). A literature review on the usage of technology acceptance model for analysing a virtual reality's cycling sport applications with enhanced realism fidelity. In 2020 8th International Conference on Information Technology and Multimedia (ICIMU) (pp. 237–242). IEEE.

Makki, A. M., Ozturk, A. B., & Singh, D. (2016). Role of risk, self-efficacy, and innovativeness on behavioral intentions for mobile payment systems in the restaurant industry. Journal of Foodservice Business Research, 19(5), 454–473.

McMahan, R. P., Bowman, D. A., Zielinski, D. J., & Brady, R. B. (2012). Evaluating display fidelity and interaction fidelity in a virtual reality game. IEEE Transactions on Visualization and Computer Graphics, 18(4), 626–633. https://doi.org/10.1109/TVCG.2012.43

Mwansa, G., Ngandu, M. R., & Dasi, Z. S. (2024). Enhancing practical skills in computer networking: Evaluating the unique impact of simulation tools, particularly Cisco Packet Tracer, in resource-constrained higher education settings. Education Sciences, 14(10), Article 1099. https://doi.org/10.3390/educsci14101099

Neuman, W. L. (2014). Social research methods: Qualitative and quantitative approaches (6th ed.). Allyn and Bacon.

Nofita, M., Handarkho, Y. D., & Mudjihartono, P. (2024). Student acceptance of LMS in Indonesian high schools: The SOR and extended GETAMEL frameworks. Interdisciplinary Journal of Information, Knowledge, and Management, 19(25), 48–53. https://doi.org/10.28945/5251

Shanmugam, M., Sun, S., Amidi, A., & Khani, F. (2016). The applications of social commerce construct. International Journal of Information Management, 36(3), 425–432. https://doi.org/10.1016/j.ijinfomgt.2016.01.007

Shen, D., Laffey, J., Lin, Y., & Huang, X. (2006). Social influence for perceived usefulness and ease-of-use of course delivery systems. Journal of Interactive Online Learning, 5(3), 270–282.

Schumaker, R.E., Lomax, R.G., (2016) A Beginner's Guide to Structural Equation Modeling (4th Edition), Routledge, New York.

Soliman, M., Pesyridis, A., Dalaymani-Zad, D., Gronfula, M., & Kourmpetis, M. (2021). The application of virtual reality in engineering education. Applied Sciences, 11(6), Article 2879. https://doi.org/10.3390/app11062879

Tang, Y. M., Chau, K. Y., Kwok, A. P. K., Zhu, T., & Ma, X. (2022). A systematic review of immersive technology applications for medical practice and education—Trends, application areas, recipients, teaching contents, evaluation methods, and performance. Educational Research Review, 35, Article 100429. https://doi.org/10.1016/j.edurev.2021.100429

Wang, W. T., & Lin, Y. L. (2021). An integrated theoretical investigation of healthcare students' perceived compatibility using online learning systems on their learning performance. Journal of Internet Technology, 22(1), 143–155.

Wen, F., & Wang, H. (2020). Research on fidelity technology evaluation index of virtual simulation experiment. In Proceedings of the 18th IEEE International Conference on Emerging eLearning Technologies and Applications (ICETA 2020) (pp. 774–779). IEEE.

Wilson, S. L., Riva, E., & Lister, K. (2024). Positive pedagogies: Co-creation partnerships to support social and emotional learning in higher education. Social and Emotional Learning: Research, Practice, and Policy, 3, Article 100035.

Wong, S. Y., Hussain, A., & Saare, M. A. (2022). A survey analysis: Students' perceptions of using simulation game as learning tool. ASEAN Engineering Journal, 12(1), 105–110.

Wu, Z., & Liu, Y. (2023). Exploring country differences in the adoption of mobile payment service: The surprising robustness of the UTAUT2 model. International Journal of Bank Marketing, 41(2), 237–268.

Yang, Q., Chen, Q., Wang, J., & Ou, R. (2022). The effect of student self-efficacy on learning outcomes in a business simulation mobile game: A quasi-experimental study. Library Hi Tech. Advance online publication.

Xie, T., Zheng, L., Liu, G., & Liu, L. (2022). Exploring structural relations among computer self-efficacy, perceived immersion, and intention to use virtual reality training systems. Virtual Reality, 26(4), 1725–1744.

Yu, U. J. (2017). Examination of students' acceptance and usage of simulation software technology in an advanced-level merchandising class. International Journal of Fashion Design, Technology and Education, 10(3), 295–304.

Zheng, X. (2022). Real-time simulation support for real-time systems. In H. Kopetz & S. Chakraborty (Eds.), Handbook of real-time computing (pp. 591–604). Springer Nature Singapore.

Journal of Information and Organizational Sciences
Volume 49, Number 2 (2025)
Journal homepage: jios.foi.hr

JIOS

DOI: 10.31341/jios.49.2.7
Open Access

UDC 378.014.242:001.891.3(4-6EU)
Original Scientific Paper

Towards the Institutional Student Mobility Ecosystem (ISME): Model Based on Systematic Literature Review

Katarina Pažur Aničić[1*], Susan Loubet[2], Izabela Oletić Tušek[1], Anja Žnidaršič Mohorič[3], Martina Đuras Sekovanić[1], Jean-Pierre Aubin[2], Manuel Clergue[2], Lucie Lendelova[4] and Bojana Ivanović Bovan[5]

[1]Faculty of Organization and Informatics, University of Zagreb, Varaždin, Croatia
[2]Groupe ESIEA, Paris, France
[3]Faculty of Organizational Sciences, University of Maribor, Kranj, Slovenia
[4]Faculty of Management Science and Informatics, University of Žilina, Žilina, Slovakia
[5]Faculty of Organizational Sciences, University of Belgrade, Belgrade, Serbia
[*]Correspondence: kpazur@foi.unizg.hr

PAPER INFO

Paper history:
Received 30 April 2025
Accepted 01 October 2025

Citation:
Pažur Aničić, K., Loubet, S., Oletić Tušek, I., Žnidaršič Mohorič, A., Đuras Sekovanić, M., Aubin, J., Clergue, M., Lendelova, L. & Ivanović Bovan, B. (2025). Towards the Institutional Student Mobility Ecosystem (ISME): model based on systematic literature review. In Journal of Information and Organizational Sciences, vol. 49, no. 2, pp. 271-289

ABSTRACT

The internationalization of higher education, especially in terms of student mobility, is a key indicator of the quality of higher education. Additionally, increasing the number of students participating in credit mobility is one of the strategic goals of the European higher education area. To help more students take advantage of student mobility programs, it is essential to understand the factors that influence student mobility at both the institutional and individual levels. This paper proposes the Institutional Student Mobility Ecosystem (ISME) for credit mobility. It is based on a systematic literature review of 321 initially retrieved sources, with 22 analyzed in detail. The results are supplemented by the analysis of 11 policy and professional documents. The proposed ISME identifies student decisional factors, supporting mechanisms and stakeholders as key enablers of student mobility. Additionally, it outlines the outcomes for students, HEIs and society that result from student mobility. This model provides valuable groundwork for researchers in the field of student mobility, facilitating further in-depth analysis of specific elements within the student mobility ecosystem.

Keywords: International student mobility, Ecosystem, Systematic Literature Review

1. Introduction

Since its establishment by the EU in 1987, the "Erasmus" program has provided nearly 14 million individuals with opportunities for mobility and skills development. It has significantly impacted the internationalization of higher education institutions (HEIs), with the largest portion of its budget dedicated to Learning Mobility, which accounted for 53% of the total budget in 2022 (European Commission, 2023). The internationalization of higher education and specifically student mobility have gained even greater importance with the Bologna declaration in 1999. Nowadays, student mobility is one of the important criteria in the processes of (re)accreditation of HEIs and student mobility within and across higher education systems takes its place in the Standards and Guidelines for Quality Assurance in the European Higher Education Area (2015). As such, student mobility occupies an important place in strategic documents at the EU level. The Council Recommendation "Europe on the Move" set an important EU-level 2030 target - at least 23% of tertiary graduates should have a learning mobility experience abroad by that year (Council of the European Union, 2024) According to the Education and Training Monitor 2024, an estimated 10.9% of graduates have experienced learning mobility, with degree mobility accounting for 4.2% and credit mobility for 6.7% (European Commission, 2024).

In order to achieve the ambitious goals set at EU level, Member States need to make certain efforts at all levels, including ministries responsible for education, universities and at the level of individual faculties. However, despite the undeniable advantages that mobility programs offer to students, particularly in enhancing their employability and lifelong learning skills (European Commission, 2018), the potential benefits of HE students' participation in mobility programs are still not fully exploited, either on an individual or organizational level. According to the Erasmus Student Network (ESN) Survey XV, the main challenges include trust issues between partner universities, decision-making power for recognition depending on individual professors, lack of understanding of how the ECTS works, access to information on available courses and pre-departure support in preparations for the learning agreement (Erasmus Student Network, 2023).

In addition to policy papers, reports, and extensive surveys at the European level, international student mobility is also a subject of scientific research. Papers addressing student perspectives on mobility often focus on three main areas: motivational factors for student mobility (Cuzzocrea and Krzaklewska, 2022; Sousa et al., 2024), barriers and obstacles to participating in student mobility (Heirweg et al., 2020; Kmiotek-Meier et al., 2019; Souto-Otero et al., 2013), and the outcomes of student mobility (Granato et al., 2024; Kratz and Netz, 2018; Roy et al., 2019). These studies frequently focus on specific groups of respondents, such as students with disabilities (Heirweg et al., 2020) or students from lower socioeconomic backgrounds (Granato and Schnepf, 2024). Additionally, they may examine specific aspects of student mobility, such as the financial returns (Kratz and Netz, 2018) or the development of a European identity (Mocanu and Llurda, 2024) as outcomes of the experience. Not surprisingly, the majority of these papers emphasize the student perspective, exploring a wide range of topics. Conversely, the perspectives of other stakeholders, such as academic or study abroad coordinators (Konakli et al., 2024) and international relations officers (Bulut-Sahin, 2023), are infrequently addressed in the literature.

Several theoretical models of student mobility can be found in the literature. Sousa et al. (2024) presented a theoretical model outlining two factors influencing students' decisions to participate in mobility programs. Push factors are the internal motivations that drive students to seek activities fulfilling their needs, explaining why they pursue a mobility program. On the other hand, pull factors are the external motivations associated with the destination country that influence their choice of where to study or work abroad. These factors clarify the reasons behind selecting a specific country or institution for the mobility program. Kratz and Netz (2018) developed a conceptual framework to explain the financial returns of international student mobility (ISM). Their findings indicate that graduates with ISM experience enjoy faster wage growth after graduation and earn higher wages in the medium term than their peers. Roy et al. (2019) conducted a systematic literature review of international student mobility programs, analyzing cultural, personal, and employment/career outcomes and benefits derived from participation in such programs.

While the benefits of student mobility—academic, professional, and personal—are widely recognized, as are the barriers and obstacles to participation, various other factors related to the study program and the HEI can also influence students' decisions regarding participation in mobility programs and their ability to successfully take advantage of these opportunities. For example, support services provided by the host university can play a significant role in this regard (Perez-Encinas et al., 2017).

As the existing model falls short in recognizing all the important factors related to student mobility, the present study aims to contribute to the field of research on student mobility in higher education by proposing the Institutional Student Mobility Ecosystem (ISME). The inspiration for this model lies in the University-Business Cooperation (UBC) ecosystem proposed by Davey et al. (2011), which includes benefits, drivers, barriers, and situational factors on the factor level; strategies, structure & approaches, activities and framework conditions at the action level; and outcomes that are materialized at the results level. Therefore, based on the systematic literature review (SLR), this study answers the following main research question: "What are the main elements of the Institutional Student Mobility Ecosystem (ISME)?"

2. Methodology

This study utilized SLR which was performed in several steps, as shown in Figure 1, following the steps as described in Kitchenham (2004), described below. SLR is a common method in different fields, the main aim being to provide deep and comprehensive insight into a certain field. In this research, SLR was applied in order to recognize and design the Institutional Student Mobility Ecosystem (ISME).

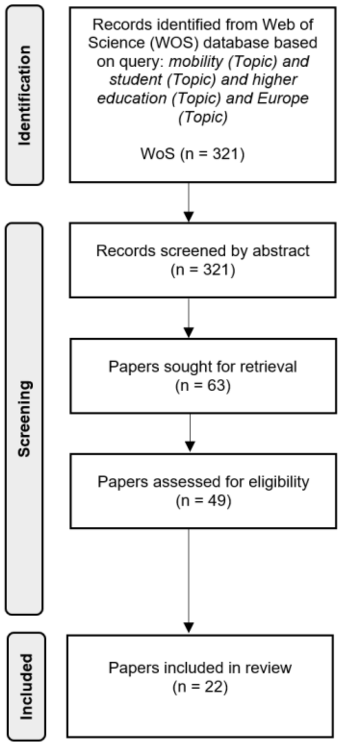

Figure 1. Literature review steps (based on the PRISMA 2020 flow diagram template[1])

2.1. Identification of a need for review

This study is a part of the Erasmus+ co-funded project "SuMoS - Strengthening the ecosystem for sustainable student mobility", aimed at supporting awareness-raising about the green transition, environmental and climate-change challenges in practices and processes related to student mobility, at institutional and individual levels. Tertiary education ecosystems are recognized as dynamic structures comprising several subsystems (Hazelkorn and Locke, 2023). One of these subsystems, introduced in this paper, is the Institutional Student Mobility Ecosystem, or ISME. Student mobility can be observed at systemic, organizational and individual levels (European Commission, 2018); the ISME therefore needs to be coherent, collaborative, coordinated and co-produced with other higher education sub-ecosystems. As highlighted in the introductory section, while many papers discuss different aspects of student mobility, the concept of the ISME has not yet been introduced. Therefore, to enhance the ISME further, it is essential to identify all of its components.

2.2. Specifying the research question

The aim of this SLR was to identify elements of the student mobility ecosystem, whether these elements were mentioned in primary or secondary sources. Additionally, the frequency with which certain elements appear is not significant, as the final results (the proposed ISME) will only comprise the identified elements.

[1] PRISMA flow diagram is available at: https://www.prisma-statement.org/prisma-2020-flow-diagram

2.3. Development of a review protocol and source selection

In order to detect all relevant studies, a broad query was conducted in the Web of Science (WoS) database "mobility (Topic) and student (Topic) and higher education (Topic) and Europe (Topic)", which resulted in 321 papers, in December 2024. There were no additional limits to the query, so it called up journal papers, conference papers and books.

2.4. Quality assessment

After identifying 321 sources as potentially relevant based on the initial query, two researchers independently reviewed abstracts in the first round of quality assessment to determine eligibility for further analysis, categorizing them as Yes, No, or Maybe. The inclusion criteria were that, from the abstract, it was indicative that the paper was focused on a credit student mobility (non-degree seeking) and provided a broad perspective on student mobility in European countries. The principal researcher conducted the second round of title and abstract analysis. In the third round of quality assessment, the principal researcher and one of the two researchers analyzed the articles that lacked consensus to establish their eligibility for further analysis. This process concluded with 53 Yes and 10 Maybe articles included in further analysis. The research team had access to 49 full articles. While analyzing the 49 articles with full access, the readers offered insights on each article, including details about participants, sample size, country, research methods, research goals or questions, key terms, and relevance. Based on their feedback and comments, the principal researcher reviewed all articles and made the final decision on which ones to consider for final analysis. The main inclusion criterion was that the article provided valuable information on student credit mobility in the European context, identifying more elements related to the ISME. Articles were excluded if they focused on degree mobility, postgraduate mobility, mobility outside Europe, broader internationalization activities with only marginal reference to student credit mobility, or on particular aspects offering limited relevance for the development of ISME (e.g., sustainability in student mobility or mobility's contribution to tourism). In addition, articles providing only a narrow description of mobility in specific countries or older studies lacking current relevance for ISME development were also excluded.

In addition to reviewing 22 research papers, the authors thoughtfully analyzed 11 policy and professional papers to refine and enhance the categories identified. Unlike research articles, policy and professional documents on student mobility are not systematically indexed in databases, so there is no single repository that could be used for a structured search. For this reason, recent and influential documents related to student credit mobility that are produced by key stakeholders such as the European Commission, the European Parliament, the Erasmus Student Network (ESN), or the European Association for International Education (EAIE) were included. A full list of the research papers, policy and professional documents included in the analysis is available in the Appendix.

2.5. Data extraction

Both inductive and deductive approaches were used to determine ISME elements. First, the principal researcher read several articles to propose the initial ISME categories and elements, making it easier for other researchers to engage with the study. The articles were distributed among nine team members who conducted the reading and analysis. Prior to involvement of other researchers in reading, a meeting was held with all researchers to explain the methodology. When reading the papers, the researchers marked the recognized elements of ISME and categorize them in an Excel table. As researchers reviewed the complete articles, new categories began to emerge. If a specific element was identified within an article, whether as a primary or secondary reference, it was marked in the Excel table.

The inductive approach was employed to analyze qualitative data using coding to determine ISME elements. Manual coding followed Saldaña's (2013) recommendations for organizing similarly coded data into categories. In inductive research aimed at developing a new theory, recoding and re-categorizing is repeated iteratively until the final set of codes is developed (Saldaña, 2013, pp. 9-10). Therefore, this study used a two-cycle coding process: 1) assigning preliminary codes to the elements recognized from the literature and 2) categorizing the detected elements into smaller sets of categories for clarity. The credibility of the research is established by the involvement of researchers with diverse and relevant backgrounds in student mobility within higher education, including professionals in international relations, a vice-dean in international relations, international relations officers, as well as academic advisors and professors. Moreover, the coding process was systematically documented in Excel sheets for future reference, ensuring the dependability and confirmability of the results.

3. Results

Based on the general structure of the University-Business Ecosystem (UBC) as outlined by Davey et al. (2011)[2], the Institutional Student Mobility Ecosystem (ISME), which focuses on student credit mobility, reflects a similar framework, as illustrated in Figure 2. This proposal draws upon elements identified from an analysis of 22 papers and 11 policy papers, following SLR methodology previously described.

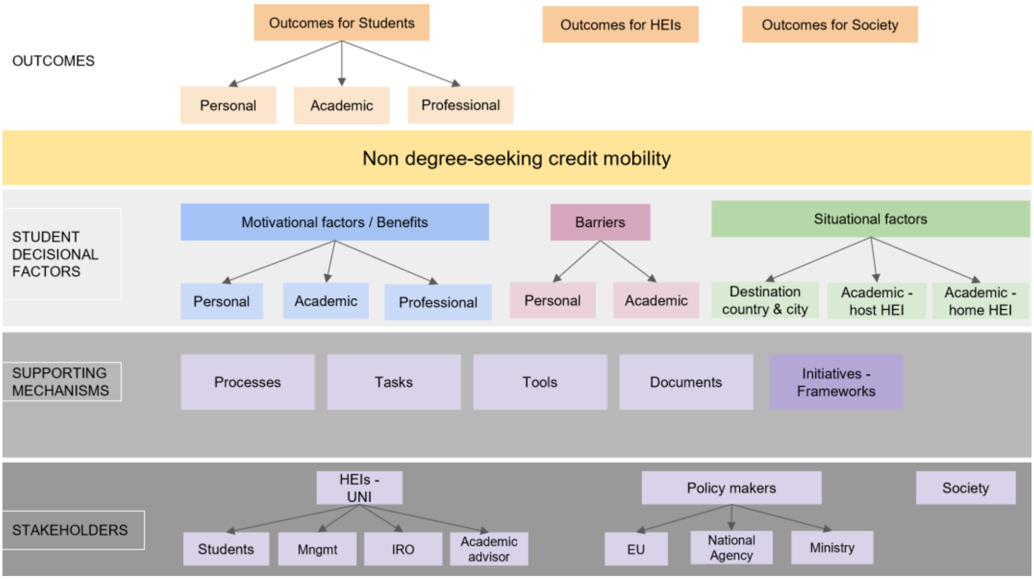

Figure 2. Proposed structure of the Institutional Student Mobility Ecosystem (ISME)

Students are at the center of the proposed model as they seize or decline opportunities for mobility, because ultimately, students' decisions to carry out mobility periods or not, will "make or break" mobility programs, wider internationalization policy and the positive societal change sought after. This chapter further describes the main recognized ISME pillars (student decisional factors, supporting mechanisms, stakeholders and outcomes), focusing on student decisional factors, which are more extensively recognized in the literature, while other elements have been researched less and require further investigation.

3.1. Student Decisional factors

Three main categories of student decisional factors are recognized: motivational factors/perceived benefits, barriers and situational factors.

3.1.1. Motivational factors/benefits

Motivational factors for students can be further classified into three main categories: personal, academic, and professional (Figure 3). Depending on the student's perspective, some may view these motivational or push factors as reasons to decide to participate in an exchange program. In contrast, other students might not consider these factors before their mobility, but they do recognize them as the perceived benefits (outcomes) gained from their study abroad experience.

[2] A more detailed representation of the UBC Ecosystem model, created by Davey, T., Galan-Muros, V., and Meerman, A., can be found in the 2010/2011 country reports available on the website: https://www.ub-cooperation.eu/index/reports

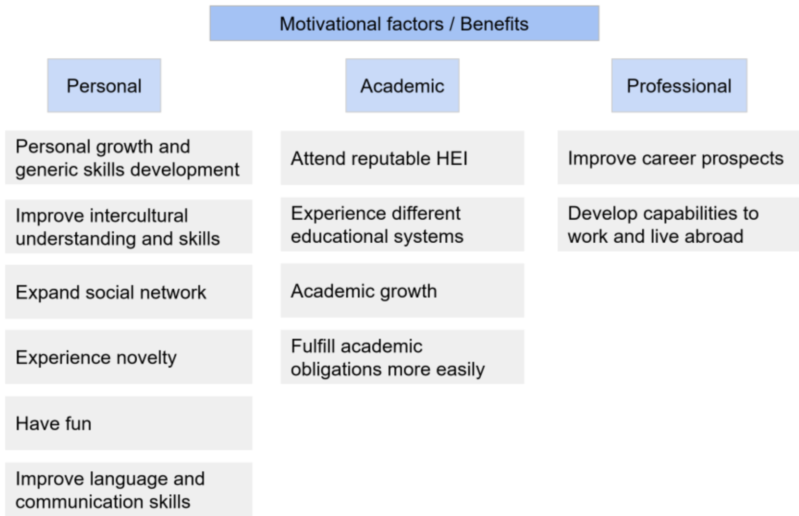

Figure 3. Motivational factors / benefits for student mobility

Personal motivational factors

Within the literature analyzed here, personal factors are recognized as important motivational factors from the students' point of view. Student mobility results in the development of a wide range of student generic/soft skills and personal growth in a broader sense, including an opportunity to mature and improve the student's degree of autonomy (López-Duarte et al., 2023), as well as develop their independence and resiliency (Van Maele et al., 2016; Pineda et al., 2008); increased confidence (Silva et al., 2017; Martins et al., 2016) and open-mindedness (Granato et al., 2024; Van Maele et al., 2016). Furthermore, some students view mobility as a reflexive project of the self (Cuzzocrea and Krzaklewska, 2023). Improving intercultural understanding and skills is another important motivating factor for traditional student mobility (Konakli et al., 2024; Sousa et al., 2024). Students are keen to learn more about the local culture and tradition, enlarge their cultural knowledge, and enhance their cross-cultural proficiency and sensitivity. This enhances their intercultural awareness/understanding. Moreover, students are striving, not only to understand different cultures, but also to acquire or improve their foreign language knowledge and skills (Sousa et al., 2024; Krzaklewska and Cuzzocrea, 2024; Kosmas et al., 2020). Therefore, student mobility fosters cross-cultural competencies, which enhance students' chances of accessing the international labor market and working in multicultural environments (López-Duarte et al., 2021). Mobility provides students with essential skills to build global networks as well. Engaging with new people, forming social connections, making friends, and socializing with other international students are recognized as vital activities during their mobility semester (Mocanu and Llurda, 2024; Krzaklewska and Cuzzocrea, 2024; Kosmas et al., 2020; Sousa et al., 2024). The desire for new experiences while living in another country, as well as the quest to discover if one is "missing something" (Krzaklewska and Cuzzocrea, 2024) serve as strong motivations for student mobility. The "fun" factor, including travelling, organizing meetings, trips and engaging with Erasmus+ students, is also among the significant objectives for student mobility (Konakli et al., 2024; Lopez-Duarte et al., 2023; Sousa et al., 2024). The findings of Sousa et al. (2024) research show that the three main motivations for students to engage in mobility are "For leisure/fun/travel", "Make new friends, create an international social network" and "Learn about another culture and traditions". While students today often experience the pressure to excel and progress, the Erasmus experience is an opportunity for their "oasis of youth", which provides them "a social (or even institutional) space for both becoming (an adult) and being (a young person)", enabling them to 'pause' and enjoy the youthful experience (Krzaklewska and Cuzzocrea, 2024).

Academic motivational factors

Among different motivational factors and powerful incentives, academic ones are the most prominent, as credit mobility occurs during the educational period. For most students, gaining academic or study experience is among the most important reasons for participation in the Erasmus+ programme (Kosmas et al., 2020).

The study abroad period is an opportunity for students' academic growth and enhancement. Indeed, some students consider Erasmus as part of their educational plan even before enrolling in university (Cuzzocrea and Krzaklewska, 2022). Some students see the study abroad period as an opportunity to attend a reputable HEI or prestigious university (Lopez-Duarte et al., 2023; Rodríguez González et al., 2011; Sin et al., 2017). Experiencing an international learning environment was rated the highest among choices concerning the city and the HEI in the research from Sousa et al. (2024), which indicated that students value the opportunity to experience a different educational system, requiring adaptation to a new learning environment and dealing with the resulting "learning shock". In some cases, improved academic performance (Granato et al., 2024) and deeper engagement with coursework (Konakli et al., 2024) motivate students to carry out a study mobility. A study by Granato et al. (2024) found that student mobility does not delay graduation. Instead, it positively affects the final graduation marks of undergraduate students. Improving academic knowledge and acquiring more knowledge and skills is an important part of academic motivation for study abroad (Martins et al., 2016; Silva et al., 2017; Sousa et al., 2024). Sometimes students aspire to learn something new during the study abroad period, i.e. study subjects or specialities that are not offered at their home institution (Pineda et al., 2008). Teaching methodologies of certain universities can motivate students to study at that institution as well (Pineda et al., 2008). For example, that is the case among students choosing the European Project Semester Program for their study abroad period (Sousa et al., 2024). Studying abroad may also benefit students' future educational pathways and increase their motivation for continuing education abroad (Granato et al., 2024).

In contrast, a less admirable motive is that students view the study abroad period as an opportunity to fulfil their academic obligations more easily, i.e. there is less work involved in obtaining the ECTS credits at the host institution (Sousa et al., 2024). Some students find motivation in manageable academic responsibilities (Krzaklewska and Cuzzocrea, 2024) or favourable grading policies at the host institution and/or advantageous grade conversions at their home institution; or even the opportunity to earn passing grades abroad in subjects with high failure rates at home (Sin et al., 2017). According to a study by Granato et al. (2024), evidence shows that Erasmus students achieve higher exam marks only while studying abroad, particularly at host institutions with lower relative quality, and not after returning home.

Professional motivational factors

Erasmus students are motivated to study abroad for both personal and academic reasons (Pineda et al., 2008), but professional reasons play an important role as well. Increasing chances for future employability, developing employability skills and improving career prospects were recognized as significant factors for many students concerning their professional development (Granato et al., 2024, Lopez-Duarte et al., 2021; Sin et al., 2017). Students believe that participating in an exchange program and studying abroad for one semester or two will make them stand out to employers, compared to students who have not participated in mobility programs and who are their competitors in the job market. According to Martins et al. (2016), Erasmus+ students view their experience as essential for integrating the job market, as employers value it highly. They believe this experience makes it easier to find jobs abroad and that employers appreciate the coursework from other universities (Martins et al., 2016). According to research from Sousa et al. (2024), pursuing international mobility to facilitate inclusion in the labor market is a motivational factor rated with the importance $M = 3.27$ ($sd = 1.04$), on a five-point Likert scale (Sousa et al., 2024). Improving the CV and making it more attractive to future employers is another professional motivational factor for students, as they see study period abroad in their CV as an added value (Lopez-Duarte et al., 2023; Martins et al., 2016) and the factor that "makes it stronger" (Silva et al., 2017), especially if it includes an exchange program at a prestigious university (López-Duarte et al., 2023). Moreover, the development of a range of generic skills during student mobility significantly enhances students' professional growth. This process not only helps them acquire cutting-edge skills but also affects their professional development as it provides a competitive advantage in the labor market and widens the possibilities of aspiring to more qualified jobs (Lopez-Duarte, 2023; Sousa et al., 2024). Besides improving their career prospects in different ways, in traditional mobility, students have the opportunity to immerse themselves in other cultures while also seeking educational and job opportunities in different countries (Konakli et al., 2024; Silva et al., 2017). For some students, the desire for long-term immigration may motivate their decision to study abroad, as it allows them to access the international labor market and work towards an international career (Lopez-Duarte et al., 2023; Sousa et al., 2024; Sin et al., 2017), and potentially positively influences their likelihood of living or working abroad. However, career objectives as mobility goals seem to have limited impact, as highlighted in the study by López-Duarte et al. (2023). The authors stress that mobility goals should be based on realistic benefits, particularly in increasing students' chances of securing paid employment.

3.1.2. Barriers

Although there are many motivational factors and benefits associated with student mobility, students' decisions to pursue a period of mobility are significantly influenced by various barriers. These barriers can generally be categorized into two main groups: academic issues and personal factors (Figure 4.).

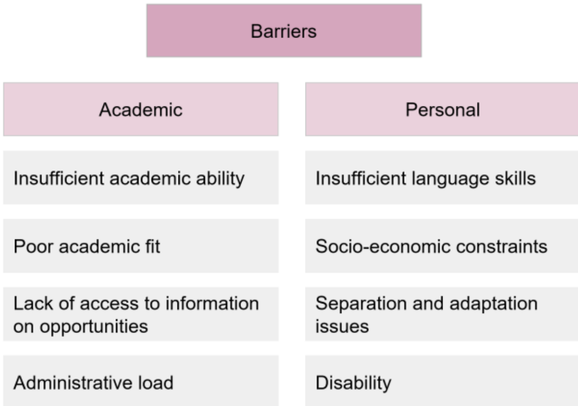

Figure 4. Barriers to student mobility

Different aspects of academic-related barriers are recognized from the literature. Konakli et al. (2024) discuss how study-abroad coordinators view minimum academic standing as a key eligibility factor in Erasmus+ placements, alongside language proficiency. In contrast, results from Granato & Schnepf (2024) indicate that students with lower academic abilities may hesitate to apply for mobility due to fears of failing the selection process or viewing application requirements as too high. In addition to being a barrier for student applications, curricular barriers related to course offerings at host institutions, as well as limitations imposed by compulsory courses at home institutions and inflexibility in taking courses abroad (Sin et al., 2017), have significant and potentially negative impact on the process of creating a mobility study program plan. Additionally, poor fit between academic programs at the home and host institution could be due to the study field or lack of ECTS information packages (course catalogue) in English (Shanaida et al., 2018). The barrier of different academic cultures arises when students encounter unfamiliar teaching methods (Sin et al., 2017), assessment styles, and institutional expectations in their host country, which can hinder their academic performance and overall mobility experience. Granato et al. (2024) emphasized that studying abroad involves time-intensive organizational tasks, as acquiring the necessary organizational and adaptation skills to live in a new environment requires significant time and effort. This may, in turn, lead to poorer academic performance and delays in one's study career. Finally, some students also expected difficulties in recognition and validation of the selected courses and ECTS received by host institutions (Nada et al., 2023), which may delay completing the degree (Sin et al., 2017).

Learning a new language or developing foreign language skills (e.g. Sousa et al. 2024, Silva et al., 2017, Kosmas et al., 2020, Martins et al., 2016) could be one of the main motivators for mobility, while on the other hand language proficiency, often together with minimum academic standing, could be used as selection criteria to meet eligibility for mobility (Konakli et al., 2024, Granato & Schnepf, 2024) and therefore one of the main obstacles (Kmiotek-Meier et al., 2019, Fleaca et al, 2015). Kosmas et al. (2020) describe detailed language obstacles or important challenges such as difficulty in understanding the regional accents/dialects, in understanding the academic language, and in using jargon/slang of fellow students.

While parents, relatives or friends can play a motivating and supportive role in the choice of a mobility (Sousa et al., 2024), a long separation from the family can also be a major barrier to participation in student mobility (Fleaca et al., 2015). Kmiotek-Meier et al. (2019) found that anticipated negative impact on psychological well-being during the mobility period, defined as fear of suffering from stress/loneliness/sadness, was forth barrier to higher education credit mobility participation. In addition to psychological support, family support also has a financial dimension. Granato & Schnepf (2024) emphasize that students whose parents have completed fewer years of study are less likely to apply for mobility, while Lopez-Duarte et al. (2021) emphasize that international student mobility is (still) elitist and often inaccessible to disadvantaged students with lower socio-economic backgrounds. When comparing the advantages and disadvantages of traditional and virtual mobility, Lopez-Duarte et al. (2023) emphasize that in the future, more students may prefer virtual mobility and turn away from traditional mobility programs because families' financial resources are shrinking, and they do not want to burden their families. Kmiotek-Meier et al. (2019) confirm these findings that wealthier families have a higher chance of undertaking international mobility/experiences and add that this could be generalized at country level, as wealthier countries support students with generous government grants.

3.1.3. Situational factors

Situational factors for students can be classified into three main categories: destination country & city, academic factors related to the home institution and academic factors related to the host institution (Figure 5).

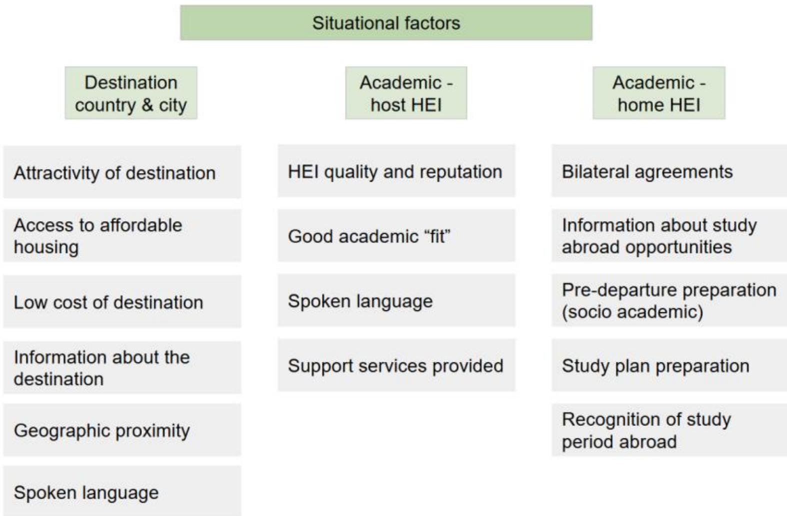

Figure 5. Situational factors for mobility from student perspective

Destination city & country

The push-pull model is one of the most commonly used frameworks for studying motivation for mobility (Sousa et al., 2024). It suggests that pull factors play a key role in selecting a destination country. Sousa et al. (2024) observe that students use the same factors to choose their study-abroad destinations as they do for planning holidays. In fact, many students prioritize cultural and tourism experiences over academic benefits, leading to discussions about "academic tourism" (Silva et al., 2017; Sin et al., 2017; Sousa et al., 2024). As a result, promoting low living costs, assistance in finding accommodation, and leisure opportunities may be more effective than emphasizing academic quality and career prospects when seeking to motivate students to participate in mobility programs. In Cirkvenčić and Lončar (2021), the authors note that limited accommodation, among other factors, is an obstacle that may decrease the quality of life of exchange students. Additionally, security and freedom of the destination country are important. Students may even select their destination based on political and cultural freedom (Sousa et al., 2024). For Cirkvenčić and Lončar (2021), the perception of security is a key factor of satisfaction during mobility in Zagreb. Besides the destination country's attractivity, the overall level of knowledge about the country and the availability of information appears to be important factors in choosing a mobility destination. These factors are ranked higher than personal recommendations from friends or relatives who have previously visited the host country (Sousa et al., 2024). Moreover, the authors highlight that geographical location and proximity of tourist destinations is an important factor in students' choice of mobility destination (Sousa et al., 2024, Cirkvenčić and Lončar, 2021; Pineda et al., 2008). Additionally, different studies indicated that the language spoken at the destination is another important factor for students (Sousa et al., 2024; Mocanu and Llurda, 2024), with differences in assigning importance to learning the local language or English. Finally, insufficient funding was already recognized as a clear barrier to mobility (Pineda et al., 2008; Rodríguez González et al., 2011). For some students, Erasmus scholarships are insufficient to support their mobility, so families' contributions and home-country financial support serve as the second major funding sources. Therefore, the perceived low cost of a destination might be the decisional factor for student mobility (Sousa et al., 2024, Cirkvenčić and Lončar, 2021).

Academic host institution

Choosing a university for an exchange semester can be a challenging process for students. There are numerous factors to consider, and it is often necessary to find a balance between the desired destination—both the country and the specific institution—and several key elements such as compatibility of study programs, overall support for students, and the community of exchange students. Situational factors related to the host institution are to some extent connected to both motivational factors and barriers. A significant incentive for students is the perceived quality or prestige of the institution, which they often see as more challenging than their home institution (Sin et al., 2017). However, some research indicates that the reputation of the institution is not highly regarded when choosing host institutions, with a mean score of 2.82 ($sd = 1.30$) on 5-point scale (Sousa et al., 2024). For some students, the quality of the HEI is evident in international recognition of the HEIs' qualifications (Sousa et al., 2024) or the institution's position in global rankings (Rodríguez González et al., 2011), while for other students, quality is defined by campus attributes, including the quality of professors, peers, teaching and assessment methods, along with the availability and quality of resources offered to students at the host institution (Granato et al., 2024). However, for some students, personal recommendations from friends who studied at the institution may be the important quality indicator in choosing a HEI for study abroad (Sousa et al., 2024).

Another important factor is the availability of courses that match the program at the home institution. Finding courses at the host institution that can substitute for mandatory courses at the home institution is quite challenging. Not only do the courses need to be available in English, but the semester—whether winter or summer—must also align appropriately. Well-prepared and structured course catalogues can facilitate students' choice of courses at the host institution (Shanaida et al., 2018). For courses to align with the study program and the semester at the home institution, it is essential for most students that these courses be offered in English (Sousa et al., 2024). Furthermore, it is important that all staff at the host institution can communicate in English (Silva et al., 2017).

In addition to academic challenges, moving to another country and culture can cause culture shock for incoming students and, therefore, HEIs should put effort into actively supporting socio-academic integration of international students and support for different aspects of culture shock (i.e. through international relations officers' support and buddy programs). Support for incoming students starts several months pre departure, immediately after students are nominated for a semester at the host institution. Besides developing their study plan, students often face the very challenging task of finding accommodation in the host country. Assistance from the host institution or associated accommodation services is much appreciated in that process (Nada et al., 2023; Perez-Encinas et al., 2017). Upon arrival, an orientation or welcome week aimed at integrating students into the new environment figures among the most important student services for incoming students (Nada et al., 2023; Perez-Encinas et al., 2017). However, Nada et al. (2023) found that support for socio-academic integration varies significantly among Erasmus students at different institutions. To address the lack of institutional support, HEIs often rely on organizations like the Erasmus Student Network (Nada et al., 2023; Perez-Encinas et al., 2017).

Academic home institution

Although the support from home institutions is not prominently evident in the analyzed papers, the procedures, practices, rules, and supporting mechanisms at these institutions play a crucial role in the student credit mobility cycle. This includes informing students about mobility opportunities, assisting them with the application process, developing their study plans for the mobility period, and providing support for the recognition of courses upon their return. Additionally, it helps with their reintegration after an intensive academic, social, and cultural experience during the mobility period. To facilitate student mobility, there are two essential requirements for the home institution. First, the institution must establish bilateral agreements with partner HEIs. Second, they need to inform students about the available mobility opportunities. Mobile students depend on bilateral agreements between institutions when selecting an international destination, which can limit their choices (Kmiotek-Meier et al, 2019). Therefore, HEIs play a crucial role in initiating these agreements with institutions that offer compatible study programs. This collaboration helps students adapt more easily to their study programs and ensures proper recognition of their time spent studying abroad. The information about opportunities for students to study abroad is usually available on the institutional websites, but Silva et al. (2017) highlighted the significance of social networks in facilitating student mobility. According to Nada et al. (2023), some previous studies have shown that pre-departure preparation at students' home institutions significantly impacts their positive experiences during their study abroad period. Preparing study plans for the period abroad is another important pre-departure activity resulting in signing a trilateral learning agreement for students, signed by the student, home and host institution (Shanaida et al., 2018). The differences in institutions and curricula around the world are closely linked to the validation and recognition of study abroad, as a crucial aspect of the Erasmus student experience (Nada et al., 2023). Even though a well-prepared and structured host institution course catalogue can help students to be as independent as possible in preparing their study plan (Shanaida et al., 2018), the support of academic advisors from the home institution can greatly facilitate that process for students and ensure smoother recognition of courses upon students' return from the mobility period.

3.1.4. Supporting mechanisms

Several different supporting mechanisms are recognized in the literature: student-oriented processes, partner-oriented tasks, tools, documents and framework conditions. However, it should be noted that supporting mechanisms are not covered in the existing literature to the extent to which student decisional factors are.

Student-oriented processes encompass the tasks performed by international relations officers, academic coordinators, and other relevant staff at HEIs to support students throughout their entire mobility experience. The literature reviewed acknowledges these processes to varying degrees, as noted by several authors (Cirkvenčić and Lončar, 2021; Granato et al., 2024; Fleaca et al., 2015; Konakli et al., 2024; Shanaida et al., 2018), but further research is necessary to explore them more thoroughly. The primary student-oriented processes are: 1) Promotion of the programs, 2) Recruitment of the mobility students, 3) Student application, 4) Eligibility screening and grant selection process / student ranking (criteria), 5) Student decision for participation; 6) Preparing students for the study abroad period, 7) Nominating students, 8) Student application to selected partner university, 9) ECTS recognition and 10) Post-departure reintegration.

Both the reviewed literature and practical experience emphasize the significance of partner-oriented tasks, evident in establishing and sustaining bilateral partnerships, particularly for student exchange initiatives. This is an ongoing process that runs alongside various student-related activities, such as recruitment, application, and selection. Several activities are involved in this process, as noted by different authors (Fleaca et al., 2015; Granato et al., 2024; Konakli et al., 2024) and corroborated by the present authors' practical experience: 1) Signed ECHE (Erasmus Charter for Higher Education), 2) Development of partnerships with the formally signed inter-institutional bilateral agreement (IIA) as an outcome; 3) Management of inter-institutional agreements, i.e. creating the IIA on the Erasmus Without Paper (EWP) dashboard or other mobility management platforms, changes to the agreement if needed, etc. and 4) Renewal of inter-institutional agreements.

While the primary mission of international relations officers (IROs) in HEIs remains the same over time, e.g. creating quality opportunities for student mobility, promoting those opportunities and then managing each student mobility project from start to finish, enormous changes have taken place over the last twenty years concerning the digital tools used to promote and manage international student mobility. The information about opportunities for students to study abroad is usually available on their institutional websites, but Silva et al. (2017) highlighted the significance of social networks in facilitating student mobility. The pervasive use of mobility management platforms such as Beneficiary Module (required for Erasmus+) or platforms of third-party providers such as MoveOn, Mobility-Online, SoleMove and others have not been examined in the literature reviewed here. These are tools which have revolutionized the day-to-day operations of IROs across the world. Contrastingly, the EWP initiative and its attendant tools (digital IIA, Online Digital Learning Agreements (OLA), the European Student Identity Card and the Erasmus+ app) are one of the focal points addressed by Konakli et al. (2024) and in policy documents (i.e. Erasmus Student Network, 2023).

In addition to digital tools, it must be stressed that different initiatives and funding mechanisms (European or otherwise) may be considered another type of "tool" supporting student mobility, i.e., Blended Intensive Programs, the European Universities Initiative, even the Strategic Internationalization Plan (Rumbley and Hoekstra-Selten, 2024).

An international student mobility project is structured and formalized by a series of documents that either facilitate the cooperation between institutions (i.e. IIA), facilitate the integration of a student into the program of a host institution (i.e. OLA), and/or attest to the reality of the mobility and the academic results obtained for credit transfer (i.e. Transcript of records) and in case of audit. The critical importance of structured and reliable course catalogues, meticulous course mapping and preparation of the learning agreement, to the success of an international study mobility, are the focus of Shanaida et al. (2018).

3.1.5. Stakeholders

Four types of actors are recognized within the proposed ISME: Students, Higher education institutions (HEIs), Policy makers, and Society. Internationalization in HEIs is driven by actors at national and supranational levels, from policy makers (ministries, national agencies, the European Commission) and different "societal" agencies such as the Erasmus Student Network or the European Association of International Education, to top management at universities, and deployed by academic faculty and administrative staff. The ultimate stakeholders, the students carrying out a mobility period, and their lived experiences are understandably the focus of most of the literature reviewed here (i.e. Sousa et al. 2024, Granato et al., 2024). There are, however, many other stakeholders in the internationalization of HEIs, of which student mobility is seen as one of the key vectors. Depending on the HEI, roles like international relations professionals/officers, Erasmus coordinators and study abroad coordinators (Konakli et al., 2024) are essential for managing the study abroad process from an organizational perspective. When it comes to credit recognition and creating a study plan, positions such as academic advisors, academic coordinators / ECTS coordinators, or even a committee for recognition in some cases, play a key role. Names differ from institution to institution, but the roles included in the student mobility process are similar throughout Europe (and beyond). Within the reviewed literature, there are only a few papers directly referring to stakeholders. While Cirkvenčić and Lončar (2021) carry out a brief review of relevant stakeholders and then focus their attention on the student mobility experience, the papers of Konakli et al. (2024) focus on the work being done by study abroad coordinators.

3.1.6. Outcomes

Outcomes of student exchange are recognized for the students, HEIs and society as stakeholders in internationalization. Students benefit greatly as individuals from participating in a student mobility program. As mentioned earlier, the personal, academic, and professional motivations for engaging in mobility have been identified. These motivations can also be viewed as outcomes that students recognize after completing their studies abroad. However, outcomes on the level of HEI and society are rarely recognized within the literature analyzed here, despite their being recognized in policy documents (European Commission, 2018). This area needs to be researched further.

The government has the responsibility of implementing programs that promote peace, prosperity, and well-being for citizens. The European Union's Erasmus mobility framework is a key example of an initiative for positive societal change. Achieving societal goals through higher education internationalization allows governments to reap significant benefits, as student mobility leads to impactful outcomes. Mobility programs contribute to a stronger European identity by promoting European values, culture, and a sense of belonging to the European community, reinforcing European citizenship (Mocanu and Llurda, 2024; Krzaklewska and Cuzzocrea, 2024, Lopez-Duarte et al., 2021; Rodríguez González et al., 2011), which is one of the EU endeavors. The development of European identity/citizenship is strongly connected with the "feeling of belonging to a common social and cultural space" (Lopez-Duarte et al., 2021). Besides contributing to the development of European identity, for some students, mobility experiences help to develop a vision of themselves as global citizens, broadening their perspectives and enhancing their global mindset (Krzaklewska and Cuzzocrea, 2024; Cuzzocrea and Krzaklewska, 2022). Moreover, mobility strengthens the European labor market, improving employability and developing a skilled workforce (brain gain) by encouraging the retention of international students after graduation (Lopez-Duarte et al., 2021; Rodríguez González et al., 2011). On the economic front, in traditional mobility students have the opportunity to organize different meetings, trips and travel (Konakli et al., 2024; Pineda et al., 2008; Sousa et al., 2024), which is known as "academic tourism" (Unurlu, 2021 in Sousa et al., 2024), even "compared to the joy of holidays" by some students (Krzaklewska and Cuzzocrea, 2024). These activities promote intercultural dialogue, enhance knowledge sharing, and stimulate local economies through tourism-related engagements.

Additionally, mobility initiatives can raise awareness of climate-related challenges, promoting sustainable travel (green) practices and contributing to broader societal efforts to mitigate climate impact (European Commission, 2024; Rumbley et al., 2024).

4. Discussion

This paper presents the Institutional Student Mobility Ecosystem (ISME), which is based on a systematic literature review of 321 initially retrieved papers from the Web of Science (WoS) database. Out of these, 22 papers met the quality criteria and were analyzed in detail, supplemented by an analysis of 11 policy and professional papers. Through the process of coding and categorizing the data from these analyses, the concept of ISME for student credit mobility is proposed, addressing the research question: "What are the main elements of the Institutional Student Mobility Ecosystem (ISME)?" Four primary ISME pillars are identified: 1) student decisional factors, 2) supporting mechanisms, 3) stakeholders, and 4) outcomes.

Student decisional factors, which constitute the first ISME pillar, are subdivided into motivational factors for mobility, barriers, and situational factors. The main characteristics of these factors influence a student's decision to participate in credit mobility. Among these, situational factors are the element most frequently investigated in the papers analyzed, which focus largely on motivations and barriers to student mobility (e.g., Cuzzocrea and Krzaklewska, 2022; Heirweg et al., 2020; Kmiotek-Meier et al., 2019; Sousa et al., 2024; Souto-Otero et al., 2013). The present study reaffirms previous classifications of motivational factors into personal, professional, and academic/educational categories (Krzaklewska and Cuzzocrea, 2024; Sousa et al., 2024). However, the categorization of barriers is less straightforward. For instance, Heirweg et al. (2020) identify financial, technical, organizational, linguistic, psychological, and practical barriers, particularly for students with disabilities. In contrast, Souto-Otero et al. (2013) categorize self-identified barriers to participation in the Erasmus Program into awareness/information, personal background, financial barriers, Erasmus conditions and higher education system comparability. Within the ISME framework, the main recognized barriers are categorized as personal and academic, following the same logic used for categorizing motivational factors. Country, city, and institutional aspects are recognized as motivational factors in some previous research (Sousa et al., 2024). In contrast, the ISME considers these factors as situational and emphasizes that both the home and host institutions play a significant role in the decision-making process for student mobility. However, these situational factors also come into play at the individual student level, influencing the decision to study abroad for a semester. The literature clearly outlines the factors influencing student decision-making. To provide a more detailed breakdown of certain elements, Figures 3, 4, and 5 address the main factors influencing student mobility decisions. The proposed ISME does not rank those factors according to their importance, but serves, rather, as a solid basis for further research into the significance of certain student decisional factors in different contexts (i.e., different HEIs and countries).

In addition to individual-level factors influencing student decisions, supporting mechanisms serve as the foundation containing all elements that support, and to some extent affect, institutional levels. Supporting mechanisms are the second ISME pillar. Each HEI has a set of processes that facilitate student mobility. This encompasses everything from promoting programs and managing student applications to preparing students for their study abroad period, providing support during mobility, and assisting with reintegration upon their return. These internal processes at HEIs are closely connected to partner-oriented tasks that support specific functions. For example, all activities related to signing and managing the Inter-Institutional Agreement (IIA) among partner institutions serve as inputs for creating mobility opportunities for students and facilitating their subsequent application process. To enhance the effectiveness of these processes, they are supported by various documents (e.g., student application forms, Online Learning Agreements) and digital tools (e.g., the Erasmus Without Paper platform). Finally, the frameworks and initiatives that establish the basis for student credit mobility are essential prerequisites around which all other elements at the institutional level are organized. Although the supporting mechanisms are recognized in the literature analyzed here, they are not extensively examined compared to the student decisional factors. Therefore, further investigation into various supporting mechanisms, their interconnections within the higher education system, and their impact on student mobility could be beneficial.

The third pillar of ISME consists of stakeholders who, from various roles and perspectives, influence student decision-making factors and the supporting mechanisms. Students, HEIs, policy makers, and society (including i.e. NGOs and ESN) participate in a network of interconnected processes and activities that are part of student mobility. To develop a clearer understanding of the ISME, it is essential to examine the role and impact of each stakeholder in more detail.

Finally, outcomes from student mobility represent the fourth ISME pillar. According to the systematic review from Roy et al. (2019), outcomes from student mobility are categorized into cultural, personal, and employment/career outcomes, while the present study used the same categorization as for the motivational factors: personal, professional and academic. As explained previously, factors that are seen as motivational before student mobility can be perceived as benefits or outcomes gained from study abroad experience. Additionally, ISME recognizes outcomes for society, i.e. developing European identity/citizenship (Mocanu and Llurda, 2024; Krzaklewska and Cuzzocrea, 2024, Lopez-Duarte et al., 2021; Rodríguez González et al., 2011) and outcomes for HEIs. Sustainability related to student mobility is the outcome factor that requires special attention (Shields and Lu, 2024), as student mobility can have both positive and negative impacts on the environment. On the negative side are increased travel and other mobility-related activities that contribute to global warming. On the positive side, however, international experiences can enhance students' awareness of sustainable practices and encourage them to adopt more sustainable behaviors. Shields and Lu (2024) highlight the need for increased awareness of the climate impact associated with international student mobility, along with the unique challenges, opportunities, and complexities present in the European higher education landscape.

Although student mobility has been previously researched in the literature, this paper contributes to the field by proposing a comprehensive view of the student mobility ecosystem, structuring its elements according to the four recognized main pillars. The proposed structure of the ISME follows the structure of the University-Business Ecosystem (UBC), as outlined by Davey et al. (2011), and contributes to the development of tertiary education ecosystems as one of its subsystems (Hazelkorn and Locke, 2023), focused on student mobility as an important part of the higher education system. Beyond its practical implications, the research also advances theoretical development by demonstrating the applicability of the UBC model in a new context within higher education ecosystems, by showing how this conceptual approach can be transferred and adapted to student mobility. *Key stakeholders, Supporting mechanisms* and *Outcomes* as the main UBC elements were retained in ISME, while *Influencing factors* from the UBC are called *Student decisional factors* in the ISME. However, the *Economic development* as an indirect outcome is not recognized within ISME. At a more detailed level, differences emerge between the two models: while the UBC is designed to explain university–business cooperation, ISME is tailored to student credit mobility, which shapes the interpretation and operationalization of each element. For example, within the UBC the *Supporting mechanisms* include strategies, structures and approaches, activities, and framework conditions, while in the ISME these are grouped as processes, tasks, tools, documents, and initiatives (frameworks). In addition, the ISME provides a more detailed explanation of elements that are specifically relevant to the context of student credit mobility. Referring to the systemic, organizational, and individual levels of student mobility (European Commission, 2018), the proposed ISME addresses all three—primarily the individual level through student decisional factors, the organizational level through supporting mechanisms, and the systemic level partially through stakeholders and outcomes. However, the current version of the ISME does not yet illustrate the interconnections and dependencies between these pillars, which remain to be further analyzed and developed in subsequent research. As such, this study provides a solid foundation for researchers and practitioners looking to enhance student mobility within higher education.

5. Conclusion

The proposed ISME presented in this paper serves as the foundation for future development of the student mobility ecosystem. Further work on the development of ISME can focus on deeper understanding of specific ISME elements on different levels – i.e. the importance of specific student decisional factors or deeper understanding of supporting mechanism at HEIs. In order to gain deeper insight into the connection between individual ISME elements, it is important to include all relevant stakeholders who are recognized as part of the student mobility ecosystem, including students, HEIs representatives (administrative, academic and managerial staff), representatives of governmental institutions and wider society, i.e. the Erasmus student network. Future work on the further development of specific ISME elements will involve conducting a case study at different HEIs along with a student survey and workshops/focus groups with relevant stakeholders within the "SuMoS - Strengthening the ecosystem for sustainable student mobility" project.

In addition to its theoretical contribution, the findings of this study point to the need for policies that enhance inclusiveness in student mobility due to the recognized socio-economic barriers for students. In particular, targeted measures such as EU- or national-level funding schemes and institutional scholarships for students from disadvantaged or underrepresented backgrounds could help address existing inequalities and ensure that mobility opportunities are accessible to all. On the HEI level, more effort should be placed on the academic recognition of the study abroad period and supporting initiatives such as the mobility window. Less rigidity in the academic recognition process would go far in encouraging more students to undertake a semester abroad. Finally, HEIs should dedicate more resources (both human and financial) to International Relations Offices that are often understaffed and overworked, due to the large number of processes they are involved in. The decision-makers heading mobility programmes such as Erasmus+ should go further in streamlining processes, so that IROs can devote more of their time to tasks with high added value, such as individual student advising, as their role is crucial in supporting the entire student mobility lifecycle.

While the proposed ISME provides a strong foundation for researchers in the field of student mobility, it is important to emphasize its research limitations as well. On the one hand, the involvement of several researchers in reading articles and identifying ISME elements can be seen as a limitation. Although there was a meeting to agree on a standardized methodology for analyzing the articles, there is always room for subjectivity and bias among different authors. However, this diversity can also add value to the research, as the researchers come from various backgrounds. They include professionals in international relations with over 20 years of experience at a strategic level, a vice-dean in international relations, international relations officers with over 15 years of operational experience related to student mobility, as well as academic advisors and professors. Additionally, the researchers represent five project partner countries, each contributing their expertise from different educational systems. This variety enriches the study's findings.

Acknowledgments

This work was supported by the project "SuMoS - Strengthening the ecosystem for sustainable student mobility", co-funded from the Erasmus+ Programme of the European Union, within KA220-HED – Cooperation partnerships in higher education (2024-1-HR01-KA220-HED-000254853). The sole responsibility for the content of this article lies with the authors. It does not necessarily reflect the opinion of the European Union.

References

Bulut-Sahin, B. (2023). International Office Professionals: An Example of Street-Level Bureaucrats in Higher Education. *Education Sciences*. 13(9), 890. https://doi.org/10.3390/educsci13090890

Cirkvenčić, F., & Lončar, J. (2021). International Mobility and the Quality of Life of Foreign Students in the City of Zagreb. *Croatian Journal of Education: Hrvatski časopis za odgoj i obrazovanje*, 23(2), 405–445. https://doi.org/10.15516/cje.v23i2.3958

Council of the European Union. (2024). *Council Recommendation 'Europe on the Move' – learning mobility opportunities for everyone.* https://data.consilium.europa.eu/doc/document/ST-9804-2024-INIT/en/pdf

Cuzzocrea, V., & Krzaklewska, E. (2023). Erasmus students' motivations in motion: understanding super-mobility in higher education. *Higher Education.* 85, 571–585. https://doi.org/10.1007/s10734-022-00852-6

Davey, T., Baaken, T., Galan Muros V., Meerman, A. (2011). *The State of European University- Business Cooperation. Final Report - Study on the cooperation between Higher Education Institutions and public and private organisations in Europe*. Science-to-Business Marketing Research Centre. https://www.ub-cooperation.eu/pdf/final_report.pdf

Erasmus Student Network. (2023). *European Parliament Erasmus+ 2021-2027 Implementation Report - ESN's contribution. The Experience of Higher Education Students: Contribution of the Erasmus Student Network*. https://www.esn.org/news/ep-erasmus-21-27-implementation-report

European Commission. (2018). *Mid-term evaluation of the Erasmus+ programme (2014-2020).* Brussels. https://ec.europa.eu/assets/eac/erasmus-plus/eval/swd-e-plus-mte.pdf

European Commission. (2023). *Erasmus+annual report 2022*. Publications Office of the European Union. https://data.europa.eu/doi/10.2766/211791

European Commission. (2024). *Education and training monitor 2024 – Comparative report.* Publications Office of the European Union. https://data.europa.eu/doi/10.2766/815875

Fleaca, E., Fleaca, B., Negoita, O. D., & Negoita, O. I. (2015). Strengthening the implementation of mobility of individuals in higher education with the aid of business process modeling. *In International Conference on Management and Industrial Engineering*. Issue 7. 471–481

Granato, S., Havari, E., Mazzarella, G., & Schnepf, S. V. (2024). Study abroad programmes and student outcomes: Evidence from Erasmus. *Economics of Education Review*, 99. https://doi.org/10.1016/j.econedurev.2024.102510

Granato, S., & Schnepf, S. V. (2024). Why are lower socioeconomic background students underrepresented in Erasmus? A focus on the selection into mobility and degree course organization. *Studies in Higher Education*. 50(3) 638-652 https://doi.org/10.1080/03075079.2024.2349963

Hazelkorn, E. and Locke, W. (2023). Researching tertiary education ecosystems. *Policy Reviews in Higher Education*. 7(2). 123-126. https://doi.org/10.1080/23322969.2023.2235647

Heirweg, S., Carette, L., Ascari, A., & Van Hove, G. (2020). Study Abroad Programmes for All? Barriers to Participation in International Mobility Programmes Perceived by Students with Disabilities. *International Journal of Disability, Development and Education*. 67(1). 73–91. https://doi.org/10.1080/1034912X.2019.1640865

Kitchenham, B. (2004) Procedures for Performing Systematic Literature Reviews. Joint Technical Report. Keele University Technical Report TR/SE-0401 and NICTA Technical Report 0400011T.1.

Kmiotek-Meier, E., Skrobanek, J., Nienaber, B., Vysotskaya, V., Samuk, S., Ardic, T., Pavlova, I., Dabasi-Halázs, Z., Diaz, C., Bissinger, J., Schlimbach, T., & Horvath, K. (2019). Why is it so hard? And for whom? Obstacles to intra-European mobility. *Migration Letters*. 16(1), 31–44. https://doi.org/10.59670/ml.v16i1.627

Konakli, T., Akman, T. A., Mert, M., & Yuksel, D. (2024). The perceptions of study abroad coordinators about digitalisation and Erasmus+ virtual mobility. *Research in Comparative and International Education*. 20(1), 27-48. https://doi.org/10.1177/17454999241301617

Kosmas, P., Parmaxi, A., Perifanou, M., Economides, A. A., & Zaphiris, P. (2020). Creating the Profile of Participants in Mobility Activities in the Context of Erasmus+: Motivations, Perceptions, and Linguistic Needs. In: Zaphiris, P., Ioannou, A. (eds*) Learning and Collaboration Technologies. Designing, Developing and Deploying Learning Experiences*. HCII 2020. Lecture Notes in Computer Science(), vol 12205. Springer, Cham. https://doi.org/10.1007/978-3-030-50513-4_37

Kratz, F., & Netz, N. (2018). Which mechanisms explain monetary returns to international student mobility? *Studies in Higher Education*. 43(2). 375–400. https://doi.org/10.1080/03075079.2016.1172307

Krzaklewska, E., & Cuzzocrea, V. (2024). Young People Experiencing Multiple Mobilities: In Search of an Oasis of Youth Across Europe. *Sociological Research Online*. 29(4). 998–1015. https://doi.org/10.1177/13607804231224998

López-Duarte, C., Maley, J. F., & Vidal-Suárez, M. M. (2021). Main challenges to international student mobility in the European arena. *Scientometrics*. 126, 8957–8980. https://doi.org/10.1007/s11192-021-04155-y

López-Duarte, C., Maley, J. F., & Vidal-Suárez, M. M. (2023). International mobility in higher education: students' attitude to international credit virtual mobility programs. *European Journal of Higher Education*. 13(4). 468–487. https://doi.org/10.1080/21568235.2022.2068637

Martins, D., Silva, S., & Silva, C. (2016). Knowledge Transfer Between European Universities Through the Erasmus Students Programme. In *European Conference on Knowledge Management.* 559–566

Mocanu, V., & Llurda, E. (2024). Multilingual ELF and European identity - contributions from study abroad through the Erasmus programme. *Journal of Multilingual and Multicultural Development*, 1–15. https://doi.org/10.1080/01434632.2024.2357716

Nada, C. I., Ploner, J., & Esteki, L. (2023). They Just Signed and Stamped Papers: Understanding the Erasmus Student Experiences. *Journal of International Students*. 13(2), 114–132. https://doi.org/https://doi.org/10.32674/jis.v13i2.4296

Perez-Encinas, A., Rodriguez-Pomeda, J., & Josek, M. (2017). Problematic Areas of Host University Support Services for Short-Term Mobility Students. *Journal of International Students*. 7(4). 1030–1047. https://doi.org/10.5281/zenodo.1035959

Pineda, P., Moreno, V., & Belvis, E. (2008). The mobility of university students in Europe and Spain. *European Educational Research Journal.* 7(3). 273–288. https://doi.org/10.2304/eerj.2008.7.3.273

Rodríguez González, C., Bustillo Mesanza, R., & Mariel, P. (2011). The determinants of international student mobility flows: an empirical study on the Erasmus programme. *Higher Education*. 62, 413–430. https://doi.org/10.1007/s10734-010-9396-5

Roy, A., Newman, A., Ellenberger, T., & Pyman, A. (2019). Outcomes of international student mobility programs: a systematic review and agenda for future research. *Studies in Higher Education*. 44(9). 1630–1644. https://doi.org/10.1080/03075079.2018.1458222

Rumbley, L. E., & Hoekstra-Selten, J. (2024). *EAIE Barometer: Internationalisation in Europe* (3rd ed.). European Association for International Education. Retrieved from: https://www.eaie.org/resources/barometer.html

Rumbley, L. E., Hoekstra-Selten, J., Morrison, G., & Lamont, A. (2024). *EAIE Barometer: Internationalisation in Europe. Third edition. Environmental sustainability and climate action*. The European Association for International Education (EAIE). Retrieved from: https://www.eaie.org/resources/barometer.html

Saldaña, J. (2013.) The Coding Manual for Qualitative Researchers. Second Edition. Sage

Shanaida, V., Vitenko, T., Droździel, P., & Madleňák, R. (2018). International academic mobility of students: analysis of the problems of organization this process in ECTS conditions. *In 12th International Technology, Education and Development Conference (INTED2018) Proceedings*. 4055–4060. https://doi.org/10.21125/inted.2018.0789

Shields, R., Lu, T. (2024). Uncertain futures: climate change and international student mobility in Europe. *Higher Education.* 88. 1791–1808. https://doi.org/10.1007/s10734-023-01026-8

Silva, S., Silva, C., & Martins, D. (2017). Erasmus students expectations: a qualitative study in Portuguese context. In *11th International Technology, Education and Development Conference (INTED 2017) Proceedings.* 3591-3600. https://doi.org/10.21125/inted.2017.0886

Sin, C., Tavares, O., & Neave, G. (2017). Student Mobility in Portugal: Grappling With Adversity. *Journal of Studies in International Education*. 21(2). 120–135. https://doi.org/10.1177/1028315316669814

Sousa, M., Fontão, E., & Barata, A. (2024). Why Students Choose the European Project Semester Program for Academic Mobility: A Case Study at an Engineering School. *Education Sciences*. 14(7). 735. https://doi.org/10.3390/educsci14070735

Souto-Otero, M., Huisman, J., Beerkens, M., de Wit, H., & Vujić, S. (2013). Barriers to International Student Mobility: Evidence From the Erasmus Program. *Educational Researcher*. 42(2). 70–77. https://doi.org/10.3102/0013189X12466696

Standards and guidelines for quality assurance in the European Higher Education Area (ESG) (2015). European Association for Quality Assurance in Higher Education (ENQA), European Students' Union (ESU), European University Association (EUA), European Association of Institutions in Higher Education (EURASHE). Brussels, Belgium

Van Maele, J., Vassilicos, B., & Borghetti, C. (2016). Mobile students' appraisals of keys to a successful stay abroad experience: Hints from the IEREST project. *Language and Intercultural Communication*. 16(3). 384–401. https://doi.org/10.1080/14708477.2016.1168050

Appendix 1 - 22 research papers included in analysis

1. Cirkvenčić, F., & Lončar, J. (2021). International Mobility and the Quality of Life of Foreign Students in the City of Zagreb. *Croatian Journal of Education: Hrvatski časopis za odgoj i obrazovanje*, 23(2), 405–445. https://doi.org/10.15516/cje.v23i2.3958
2. Cuzzocrea, V., & Krzaklewska, E. (2023). Erasmus students' motivations in motion: understanding super-mobility in higher education. *Higher Education.* 85, 571–585. https://doi.org/10.1007/s10734-022-00852-6
3. Fleaca, E., Fleaca, B., Negoita, O. D., & Negoita, O. I. (2015). Strengthening the implementation of mobility of individuals in higher education with the aid of business process modeling. *In International Conference on Management and Industrial Engineering*. Issue 7. 471–481
4. Granato, S., Havari, E., Mazzarella, G., & Schnepf, S. V. (2024). Study abroad programmes and student outcomes: Evidence from Erasmus. *Economics of Education Review*, 99. https://doi.org/10.1016/j.econedurev.2024.102510
5. Granato, S., & Schnepf, S. V. (2024). Why are lower socioeconomic background students underrepresented in Erasmus? A focus on the selection into mobility and degree course organization. *Studies in Higher Education*. 50(3) 638-652 https://doi.org/10.1080/03075079.2024.2349963
6. Kmiotek-Meier, E., Skrobanek, J., Nienaber, B., Vysotskaya, V., Samuk, S., Ardic, T., Pavlova, I., Dabasi-Halázs, Z., Diaz, C., Bissinger, J., Schlimbach, T., & Horvath, K. (2019). Why is it so hard? And for whom? Obstacles to intra-European mobility. *Migration Letters*. 16(1), 31–44. https://doi.org/10.59670/ml.v16i1.627
7. Konakli, T., Akman, T. A., Mert, M., & Yuksel, D. (2024). The perceptions of study abroad coordinators about digitalisation and Erasmus+ virtual mobility. *Research in Comparative and International Education*. 20(1), 27-48. https://doi.org/10.1177/17454999241301617
8. Kosmas, P., Parmaxi, A., Perifanou, M., Economides, A. A., & Zaphiris, P. (2020). Creating the Profile of Participants in Mobility Activities in the Context of Erasmus+: Motivations, Perceptions, and Linguistic Needs. In: Zaphiris, P., Ioannou, A. (eds*) Learning and Collaboration Technologies. Designing, Developing and Deploying Learning Experiences*. HCII 2020. Lecture Notes in Computer Science(), vol 12205. Springer, Cham. https://doi.org/10.1007/978-3-030-50513-4_37
9. Krzaklewska, E., & Cuzzocrea, V. (2024). Young People Experiencing Multiple Mobilities: In Search of an Oasis of Youth Across Europe. *Sociological Research Online*. 29(4). 998–1015. https://doi.org/10.1177/13607804231224998

10. López-Duarte, C., Maley, J. F., & Vidal-Suárez, M. M. (2021). Main challenges to international student mobility in the European arena. *Scientometrics*. 126, 8957–8980. https://doi.org/10.1007/s11192-021-04155-y
11. López-Duarte, C., Maley, J. F., & Vidal-Suárez, M. M. (2023). International mobility in higher education: students' attitude to international credit virtual mobility programs. *European Journal of Higher Education*. 13(4). 468–487. https://doi.org/10.1080/21568235.2022.2068637
12. Martins, D., Silva, S., & Silva, C. (2016). Knowledge Transfer Between European Universities Through the Erasmus Students Programme. In *European Conference on Knowledge Management.* 559–566
13. Mocanu, V., & Llurda, E. (2024). Multilingual ELF and European identity - contributions from study abroad through the Erasmus programme. *Journal of Multilingual and Multicultural Development*, 1–15. https://doi.org/10.1080/01434632.2024.2357716
14. Nada, C. I., Ploner, J., & Esteki, L. (2023). They Just Signed and Stamped Papers: Understanding the Erasmus Student Experiences. *Journal of International Students*. 13(2), 114–132. https://doi.org/https://doi.org/10.32674/jis.v13i2.4296
15. Perez-Encinas, A., Rodriguez-Pomeda, J., & Josek, M. (2017). Problematic Areas of Host University Support Services for Short-Term Mobility Students. *Journal of International Students*. 7(4). 1030–1047. https://doi.org/10.5281/zenodo.1035959
16. Pineda, P., Moreno, V., & Belvis, E. (2008). The mobility of university students in Europe and Spain. *European Educational Research Journal.* 7(3). 273–288. https://doi.org/10.2304/eerj.2008.7.3.273
17. Rodríguez González, C., Bustillo Mesanza, R., & Mariel, P. (2011). The determinants of international student mobility flows: an empirical study on the Erasmus programme. *Higher Education*. 62, 413–430. https://doi.org/10.1007/s10734-010-9396-5
18. Shanaida, V., Vitenko, T., Droździel, P., & Madleňák, R. (2018). International academic mobility of students: analysis of the problems of organization this process in ECTS conditions. *In 12th International Technology, Education and Development Conference (INTED2018) Proceedings*. 4055–4060. https://doi.org/10.21125/inted.2018.0789
19. Silva, S., Silva, C., & Martins, D. (2017). Erasmus students expectations: a qualitative study in Portuguese context. In *11th International Technology, Education and Development Conference (INTED 2017) Proceedings.* 3591-3600. https://doi.org/10.21125/inted.2017.0886
20. Sin, C., Tavares, O., & Neave, G. (2017). Student Mobility in Portugal: Grappling With Adversity. *Journal of Studies in International Education*. 21(2). 120–135. https://doi.org/10.1177/1028315316669814
21. Sousa, M., Fontão, E., & Barata, A. (2024). Why Students Choose the European Project Semester Program for Academic Mobility: A Case Study at an Engineering School. *Education Sciences*. 14(7). 735. https://doi.org/10.3390/educsci14070735
22. Van Maele, J., Vassilicos, B., & Borghetti, C. (2016). Mobile students' appraisals of keys to a successful stay abroad experience: Hints from the IEREST project. *Language and Intercultural Communication*. 16(3). 384–401. https://doi.org/10.1080/14708477.2016.1168050

Appendix 2 - 11 policy and professional documents included in analysis

1. Kalinova, N., Dias, R., & Gabriels, W. (n.d.). *The student perspective to the Erasmus+ 2021–2027 mid-term evaluation.* Erasmus Student Network (ESN). https://esn.org/sites/default/files/news/the_student_perspective_to_the_erasmus_2021-2027_mid-term_evaluation.pdf
2. Erasmus Student Network (ESN). (n.d.). *Call for Evidence*: *Final evaluation of the 2014–2020 Erasmus+ Programme and mid-term evaluation of the Erasmus+ Programme 2021-2027*. ESN. https://esn.org/sites/default/files/news/final_and_midterm_evaluation_erasmus.pdf
3. Green Erasmus Partnership. (2022). *Research on the habits of Erasmus students: consumer, daily life, and travel habits of Erasmus students from the perspective of their environmental attitudes and beliefs.* Diekmann, A. & Karaiskos, G. (eds.). https://doi.org/10.5281/zenodo.10654122
4. European Commission: Directorate-General for Education, Youth, Sport and Culture (2015). *ECTS users' guide 2015.* Publications Office of the European Union. https://data.europa.eu/doi/10.2766/87192

5. Dias, R., Buseyne, S., Haadad, Y., Topi, K., Dear, C., Mistretta, L., Rayón González, J., Gabriels, W. (2024). *ESNsurvey – 15th edition: Making Quality Mobility a Reality for All* (Report). Dias, R., Buseyne, S., Pelan, G., Lepore, S., & Kotkievicz, G. (eds.) Erasmus Student Network AISBL. https://esn.org/sites/default/files/news/xv-esnsurvey_final-report.pdf
6. Rumbley, L. E., & Hoekstra-Selten, J. (2024). *EAIE Barometer: Internationalisation in Europe* (3rd ed.). European Association for International Education. Retrieved from: https://www.eaie.org/resources/barometer.html
7. Rumbley, L. E., Hoekstra-Selten, J., Morrison, G., & Lamont, A. (2024). *EAIE Barometer: Internationalisation in Europe. Third edition. Environmental sustainability and climate action*. The European Association for International Education (EAIE). Retrieved from: https://www.eaie.org/resources/barometer.html
8. Erasmus Student Network. (2023). *European Parliament Erasmus+ 2021-2027 Implementation Report - ESN's contribution. The Experience of Higher Education Students: Contribution of the Erasmus Student Network*. https://www.esn.org/news/ep-erasmus-21-27-implementation-report
9. European Commission. (2018). *Mid-term evaluation of the Erasmus+ programme (2014-2020).* Brussels. https://ec.europa.eu/assets/eac/erasmus-plus/eval/swd-e-plus-mte.pdf
10. European Commission. (2024). *Education and training monitor 2024 – Comparative report.* Publications Office of the European Union. https://data.europa.eu/doi/10.2766/815875
11. European Association for International Education (EAIE). (n.d.). *The international student voice* [Issue of *Forum* Magazine]. Retrieved from: https://www.eaie.org/

Journal of Information and Organizational Sciences
Volume 49, Number 2 (2025)
Journal homepage: jios.foi.hr

JIOS

DOI: 10.31341/jios.49.2.8
Open Access

UDC 004.05:004.774:35:004:006.032ISO-047.37(594)
Original Scientific Paper

Assessing User Satisfaction of Local Government Websites Through ISO 25010 and Technology Acceptance Model (TAM): A SmartPLS and IPMA-Based Study in Lombok Tengah

Reza Adyaputra[1], Khairul Imtihan[1]* and Maemun Saleh[1]

[1]Bachelor of Information Systems Program, STMIK Lombok, Praya, Lombok Tengah, Indonesia
*Correspondence: khairulimtihan31@gmail.com

PAPER INFO

Paper history:
Received 17 April 2025
Accepted 25 October 2025

Citation:
Adyaputra, R., Imtihan, K. & Saleh M. (2025). Assessing User Satisfaction of Local Government Websites Through ISO 25010 and Technology Acceptance Model (TAM): A SmartPLS and IPMA-Based Study in Lombok Tengah. In Journal of Information and Organizational Sciences, vol. 49, no. 2, pp. 291-312

ABSTRACT

Despite continued efforts to digitize public services, many local government websites in emerging contexts still underperform in delivering satisfactory user experiences. This study develops an integrated evaluation framework that combines the ISO 25010 software quality model with the Technology Acceptance Model (TAM) to jointly assess system quality and user acceptance. We analyzed survey data from 524 users in Lombok Tengah, Indonesia, using Partial Least Squares Structural Equation Modeling (PLS-SEM) and Importance Performance Map Analysis (IPMA). The results indicate that functional suitability, usability, and reliability significantly shape perceived usefulness, whereas reliability, security, and performance efficiency drive perceived ease of use. Both perceived usefulness and perceived ease of use positively influence user satisfaction and behavioral intention, with satisfaction emerging as the strongest predictor. IPMA highlights performance efficiency and security as priority areas for improvement. The study contributes to e-government literature by proposing a dual layer model that links system level attributes to user-level perceptions and outcomes, and by translating statistical effects into actionable priorities for local governments seeking to enhance the quality and adoption of digital public services in semi urban developing regions.

Keywords: ISO 25010, Technology Acceptance Model, SmartPLS, IPMA, E-government, User satisfaction, Local government website

1. Introduction

In the evolving landscape of digital governance, the evaluation of public-facing information systems demands not only technical scrutiny but also a deeper understanding of user-centered behavioral responses. While substantial progress has been made in assessing e-government platforms from either a system engineering or behavioral standpoint, there remains a methodological gap in integrating both perspectives into a unified analytical framework (Fadhel et al., 2020; Granić & Marangunić, 2019). This study addresses that gap by combining the ISO 25010 software quality model and the Technology Acceptance Model (TAM), further enhanced with Importance-Performance Map Analysis (IPMA), to assess digital service delivery in local government websites. By focusing on Lombok Tengah, a semi-urban region in a developing country, this research situates itself within the broader discourse of digital inclusion, highlighting the intersection of technology performance, citizen satisfaction, and sustainable digital engagement. This integrative approach aims not only to evaluate system effectiveness but also to inform policy design through actionable, user-driven insights. Digital transformation has fundamentally reshaped the landscape of public service delivery worldwide. Governments across both developed and developing countries are increasingly leveraging

Information and Communication Technology (ICT) to enhance the efficiency, transparency, and accessibility of public services. In Indonesia, the implementation of the Electronic-Based Government System (Sistem Pemerintahan Berbasis Elektronik-SPBE) encourages local governments to develop web-based information systems as the backbone of modern public administration (Lee-Geiller & Lee, 2019; Ma & Zheng, 2019; Nguyen et al., 2020). Government websites serve as the primary gateway for administrative services, information dissemination, and interactive communication between government entities and the public. High-quality websites have great potential to foster citizen engagement, improve public trust, and accelerate access to digital services (Chan et al., 2021).

In Lombok Tengah, a semi urban regency in Indonesia, persistent structural constraints in local digital governance remain visible: uneven broadband infrastructure beyond district centers, limited technical capacity among implementing units, and insufficient citizen digital literacy and user training during the SPBE rollout. Recent official statistics show that over 20% of villages in Lombok Tengah still experience unstable internet connectivity (Ministry of Communication and Information Technology, 2023), and a 2022 provincial audit found that only 43% of local public service websites meet usability and accessibility standards (Provincial Government of West Nusa Tenggara, 2022). Common issues include slow page loading, lack of mobile responsiveness, and outdated content. Taken together, these indicators justify the selection of Lombok Tengah as the empirical setting and underline the urgency of improving digital service quality.

Nevertheless, the quality of local government websites in Indonesia continues to face significant challenges. Prior studies have highlighted persistent issues related to usability, accessibility, and system reliability, which often fall short of user expectations (Hidayah and Setyaningsih, 2019; Nishant et al., 2019; Santoso et al., 2023). In addition, many government websites do not conform to international software quality standards such as ISO 25010, resulting in suboptimal user experiences and low levels of acceptance of digital services (Debnath et al., 2021; Fadhel et al., 2019, 2020). These issues are particularly evident in local government contexts, such as in the Lombok Tengah Regency. Preliminary observations in this region reveal a range of limitations, including the lack of available online services, poor transparency of public information, non-responsive web designs, and minimal interactive features such as real-time communication or social media integration (Al-Sakran & Alsudairi, 2021; Ilyas et al., 2022; Paul & Das, 2020). These deficiencies directly hinder user engagement and diminish the perceived benefits of digital platforms (Bournaris, 2020; Irawan and Nizar Hidayat, 2022; Santoso et al., 2023).

Given these conditions, it is essential to develop a comprehensive evaluation model that integrates both technical and user-behavioral perspectives in assessing the quality of government websites. The ISO/IEC 25010 standard provides a robust framework for assessing software quality from a technical standpoint, while the Technology Acceptance Model (TAM) offers theoretical insight into users' cognitive and behavioral responses toward technology adoption. Unlike previous studies that have applied either ISO 25010 or TAM in isolation (Fadhel et al., 2019; Muchran and Ahmar, 2019; Salloum et al., 2019; Pereira et al., 2024), such an integrated perspective addresses the multidimensional nature of user satisfaction and behavioral intention toward local government websites and enables a more holistic and practical evaluation, which remains relatively underexplored in the context of digital public service delivery at the local government level. Practically, this study provides actionable insights and recommendations to enhance digital service quality in local governments. The findings are expected to guide strategic improvements, inform policy development, and offer a replicable model for semi-urban regions in developing countries. By highlighting priority service areas grounded in user perceptions and international benchmarks (Bervell et al., 2023; Chawla & Joshi, 2021; Yapp & Yeap, 2023), the study supports more effective resource allocation for policymakers.

In line with these objectives, we empirically evaluate an integrated ISO/IEC 25010, TAM model and apply IPMA to translate the estimated effects into actionable priorities for local governments. The study is guided by the following research questions: (1) How do technical quality attributes of government websites, as defined by ISO 25010, influence perceived usefulness (PU) and perceived ease of use (PEOU)? (2) To what extent do PU and PEOU influence user satisfaction and behavioral intention toward local government websites? (3) Which website quality attributes should be prioritized for improvement based on their relative importance and performance, as identified through Importance Performance Map Analysis (IPMA)?

This paper contributes theoretically by articulating a dual layer ISO/IEC 25010, TAM lens for local e-government evaluation, and practically by using IPMA to convert model estimates into actionable priorities for semi-urban governments under resource constraints. The remainder of this paper is organized as follows: the second section presents the literature review and hypotheses development; the third section outlines the research methodology; the fourth section discusses the findings and analysis; the fifth section highlights theoretical and managerial implications; and the final section offers conclusions, limitations, and directions for future research.

2. Literature Review

The evaluation of public sector websites has increasingly adopted frameworks that integrate both technical and user-centered quality dimensions. One of the most robust standards in this domain is the ISO/IEC 25010 software quality model, which outlines eight core characteristics: functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability (Fadhel et al., 2019; Dewi et al., 2020; Peters and Aggrey, 2020). Various studies have employed ISO 25010 to assess web-based platforms such as human resource systems (Geoloni & Agushinta, 2023), scholarship information systems (Nurunnisa et al., 2024), e-commerce sites (Wattiheluw et al., 2020), and enterprise resource planning systems (Peters & Aggrey, 2020). These findings affirm that ISO 25010 provides a reliable benchmark for evaluating the technical quality of web systems. In the broader e-government context, Al-Adwan et al. (2024) emphasize the importance of meta-governance in facilitating integrated policy development across digital systems, which complements the technical assessment dimensions of ISO 25010.

Building on these technical perspectives, recent studies have shown a growing tendency to integrate ISO/IEC 25010 with the Technology Acceptance Model (TAM), aiming to connect standardized software quality evaluation with the dynamics of user technology adoption. For example, Nurunnisa et al. (2024) applied the ISO 25010 TAM framework to evaluate scholarship information systems and found that usability and functional suitability directly influence perceived usefulness. Pereira et al. (2024) extended this integration in the context of higher education portals, demonstrating that system quality attributes mediated through TAM constructs significantly affect user satisfaction and continuance intention. More recently, Al-Adwan et al. (2024) emphasized the importance of meta governance in aligning technical quality standards with user adoption frameworks in public sector digital systems. Despite these contributions, studies explicitly addressing semi-urban government websites remain scarce. Most existing research focuses on national level or urban-centered platforms, leaving a critical gap in understanding how infrastructural limitations, digital literacy challenges, and socio-cultural contexts affect the relationship between system quality and user acceptance. This study addresses that gap by situating the ISO 25010–TAM integration within the semi-urban context of Lombok Tengah and complementing it with Importance Performance Map Analysis (IPMA) for strategic prioritization.

However, technical quality alone is insufficient to guarantee successful technology adoption or user satisfaction. User experience (UX) and usability are equally critical in determining the effectiveness of public-facing websites. Research has shown that visual design and usability significantly influence user perceptions, engagement, and loyalty (Martínez-Sala et al., 2020; Jongmans et al., 2022). Studies employing user-centered methods such as the questionnaire, and heuristic evaluations (Quiñones & Rusu, 2019) emphasize the importance of integrating ergonomic and cognitive dimensions into usability assessments. Furthermore, user culture has been found to affect usability perceptions, suggesting the need for culturally sensitive evaluation frameworks (Alexander et al., 2021). In a Southeast Asian context, Munajat & Irawati (2025) highlight how cultural and infrastructural factors in countries like Indonesia and Malaysia shape user interaction with public web platforms, reinforcing the need for localized usability models.

To understand user behavior in technology adoption, the Technology Acceptance Model (TAM) remains one of the most widely used theoretical models (Davis & Davis, 1989). TAM posits that perceived usefulness and perceived ease of use are key determinants of users' intention to adopt technology. Its applicability has been validated across multiple domains, including mobile health applications (Cheah et al., 2023), e-learning (Rahmawati, 2019; Salloum et al., 2019), online sports streaming (Sun & Zhang, 2021), and internet banking (Vuković et al., 2019). Several systematic reviews have reaffirmed the model's robustness in predicting user behavior (Granić and Marangunić, 2019; Al-Qaysi et al., 2020; Zaineldeen et al., 2020). Despite its popularity, TAM has been criticized for its limited consideration of external variables and contextual factors, necessitating its integration with complementary models such as ISO 25010, as implemented in this study. Supporting this integrated perspective, Alkhwaldi & Al-Ajaleen (2022) demonstrate that mobile government adoption is driven not only by perceived usefulness and ease of use but also by trust and contextual readiness, especially in the context of developing nations.

To enhance diagnostic precision in evaluating user satisfaction, researchers have increasingly utilized Importance-Performance Map Analysis (IPMA). IPMA helps identify and prioritize factors based on their importance to a target construct and their current performance levels (Sarstedt et al., 2024). This approach has been widely applied in contexts such as digital wallet adoption (Chawla & Joshi, 2021), travel applications (Fakfare & Manosuthi, 2023), customer loyalty (Siregar & Rachman, 2024), and hospital branding (Wiyono & Antonio, 2024). In the public service domain, Bervell et al. (2023) evaluated student portals using IPMA, while Sop et al. (2025) applied it to measure satisfaction with tourism destinations. These studies underscore IPMA's practical value in data-driven decision-making.

While PLS-SEM and IPMA have been widely applied in information systems research, their integrated use alongside ISO 25010 and the Technology Acceptance Model (TAM) remains limited particularly within the context of semi-urban public sector websites in developing countries (Chuang et al., 2023; Ali et al., 2024; Krithika and Vasantha, 2024). Previous studies have typically employed ISO 25010 to assess technical quality (Fadhel et al., 2019; Dewi et al., 2020) or TAM to explain user acceptance (Rahmawati, 2019; Salloum et al., 2019), yet rarely have they combined these frameworks to capture the multidimensional nature of system evaluation. These single-model approaches fall short in addressing the complex interplay between system quality, user perceptions, and continued usage intentions. This study addresses that theoretical gap by integrating ISO 25010, TAM, PLS-SEM, and IPMA into a unified evaluative framework. Practically, it responds to real-world challenges faced by local governments particularly in semi-urban regions such as Lombok Tengah by providing data-driven insights and prioritized recommendations for digital service improvement. The analytical procedures are detailed in the methodology section, and the framework developed herein offers both a robust theoretical contribution and a replicable tool for policymakers seeking to enhance public digital service quality and usability.

Given the diversity of existing evaluation frameworks, a clear rationale is needed to justify the theoretical choices adopted in this study. To clarify the conceptual positioning of this study and justify the theoretical foundation, it is important to distinguish the key constructs and methodological choices adopted. This study adopts Perceived Usefulness (PU) from the Technology Acceptance Model (TAM) as a core construct to explain user acceptance behavior. Although the Unified Theory of Acceptance and Use of Technology (UTAUT) introduces a similar construct termed Performance Expectancy, this research does not adopt UTAUT formally. Instead, it treats PU as the principal belief-based variable due to its well established role in explaining behavioral intention across various digital service contexts. While both PU and Performance Expectancy capture users' perceptions of technology's utility, PU focuses more narrowly on individual cognitive assessments within voluntary usage environments, which aligns with the setting of this study citizen interaction with non-mandatory government websites.

Moreover, this research deliberately chooses to integrate ISO/IEC 25010 and TAM, rather than combining TAM with the DeLone and McLean IS Success Model or adopting UTAUT, for several theoretical and contextual reasons: (1) ISO 25010 provides a comprehensive set of technical quality attributes that are especially suitable for evaluating public sector information systems. Unlike the D&M IS Success Model, which emphasizes constructs such as information quality and service quality at a more abstract level, ISO 25010 offers operational definitions that allow for direct system level assessments (Dewi et al., 2020; Fadhel et al., 2019). (2) TAM is preferred over UTAUT because the target users in this study citizens are not employees bound by organizational structures or mandates. UTAUT's constructs such as *effort expectancy*, *social influence*, and *facilitating conditions* are more appropriate in organizational or mandatory settings. In contrast, TAM is more parsimonious and has been proven robust in explaining voluntary technology adoption in public service platforms (Granić & Marangunić, 2019; Salloum et al., 2019). Since citizen interaction with local government websites is voluntary and not mandated, TAM's core constructs (perceived usefulness and perceived ease of use) are more suitable for capturing user acceptance behavior in this context. By contrast, UTAUT is more relevant to organizational or mandatory environments, while SERVQUAL lacks the technical specificity to evaluate software system quality. (3) The integration of ISO 25010 and TAM, enhanced by IPMA, offers a dual-layered framework that connects system-level quality with behavioral-level perceptions, while also enabling strategic prioritization. This methodological combination is theoretically grounded and practically valuable for local governments in emerging economies, where both technical capability and user-centered design are critical for service success. While alternative models such as the Unified Theory of Acceptance and Use of Technology (UTAUT) and SERVQUAL offer valuable perspectives, they were not adopted in this study due to contextual and theoretical misalignment. UTAUT, which includes constructs like effort expectancy and social influence, is primarily designed for organizational environments with mandatory technology use, making it less suitable for examining voluntary citizen interactions with government websites. Similarly, SERVQUAL emphasizes intangible service dimensions such as empathy and assurance, but lacks the technical specificity required to evaluate software system quality. In contrast, ISO/IEC 25010 provides concrete system-level metrics covering functional suitability, usability, and performance efficiency (Dewi et al., 2020; Fadhel et al., 2019), while the Technology Acceptance Model (TAM) robustly explains behavioral intention in various digital service contexts, particularly in voluntary. Therefore, the integration of ISO 25010 and TAM offers a more context appropriate and methodologically coherent framework to assess public digital services in semi urban regions.

This study explicitly synthesizes ISO/IEC 25010, the Technology Acceptance Model (TAM), and Importance Performance Map Analysis (IPMA) to provide a comprehensive framework for evaluating semi-urban government websites. ISO 25010 offers the technical foundation, TAM explains how these attributes shape perceived usefulness and ease of use, while IPMA identifies improvement priorities by combining importance with performance levels. This integration not only links system level quality with user-centered perceptions and behavioral intentions but also addresses the practical need for actionable guidance in digital service development. In operationalizing ISO 25010, the study focuses on five user-salient attributes functionality, usability, reliability, security, and performance efficiency because they are directly perceivable by citizens. The other three characteristics, namely compatibility, maintainability, and portability, are excluded as they are developer centric or already represented through performance measures such as responsiveness.

Although national level e-government initiatives have been widely promoted by the central government, semi urban regions such as Lombok Tengah continue to face structural and contextual challenges that hinder full digital transformation. Prior studies highlight that infrastructural limitations, low levels of digital literacy, and the absence of user training significantly influence how citizens adopt and interact with online government platforms (Alexander et al., 2021; Munajat & Irawati, 2025). This context provides a representative case to examine the interplay between system design quality and user perceptions in digital public service delivery. Accordingly, this study not only bridges the theoretical gap in evaluating semi-urban e-government platforms through integrated frameworks but also delivers actionable insights to guide policy-level digital transformation strategies.

3. Hypothesis development

To ensure a more holistic understanding of user interaction with local government websites, this study adopts an integrated theoretical framework that combines the ISO 25010 software quality model, the Technology Acceptance Model (TAM), and Importance Performance Map Analysis (IPMA). ISO 25010 offers a structured assessment of technical system quality, including functionality, usability, reliability, security, and performance efficiency, attributes critical for evaluating the design and capability of web-based public services. Meanwhile, TAM explains how users form behavioral intentions based on their perceptions of usefulness and ease of use, serving as a well-established lens for assessing user acceptance. The inclusion of IPMA complements this integration by providing a prescriptive diagnostic tool that prioritizes improvement areas based on the constructs' importance and performance. Collectively, these three frameworks are synthesized into a dual-layered model: ISO 25010 forms the technical foundation that feeds into TAM's psychological constructs, while IPMA is applied post-structural modeling to guide managerial action. This framework is especially relevant in developing country contexts like Lombok Tengah, where aligning technical quality with user perceptions is essential for sustainable e-government adoption.

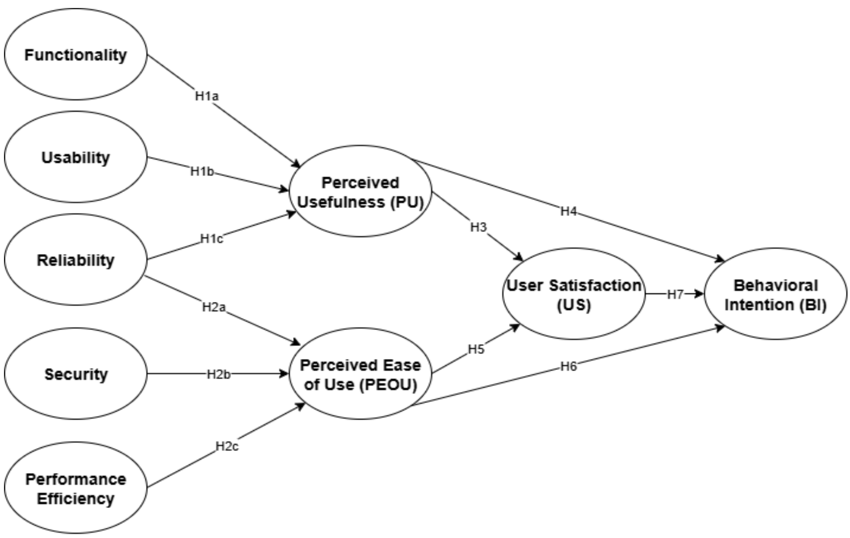

Figure 1. Research model

To enhance clarity and conceptual visualization, Figure 1 illustrates the dual-layered research model employed in this study, showing how ISO 25010 attributes influence TAM constructs, which subsequently affect user satisfaction and behavioral intention, with IPMA serving as an extension for managerial prioritization.

In assessing user acceptance of public digital services, the interrelation between system quality and user perception remains central to both system effectiveness and sustainable engagement. To provide conceptual clarity and strengthen the theoretical framework, this study formulates seven hypotheses that explicitly link system quality dimensions from ISO 25010 namely functionality, usability, reliability, security, and performance efficiency to the core constructs of TAM: perceived usefulness (PU), perceived ease of use (PEOU), user satisfaction (US), and behavioral intention (BI). These hypotheses reflect a multidimensional approach that connects technical features to user perception, affective response, and behavioral outcomes. Building upon the ISO 25010 standard, which articulates technical quality dimensions such as functionality, usability, reliability, security, and performance efficiency, this study links these attributes to key cognitive constructs derived from the Technology Acceptance Model (TAM), namely perceived usefulness (PU) and perceived ease of use (PEOU). Functionality, defined as the completeness and relevance of system features, has been shown to influence users' judgments about the utility of web-based platforms (Vila et al., 2021). Likewise, usability, encompassing intuitive design and ease of navigation, is a critical determinant of user-perceived value, especially in public systems where digital literacy varies (Alexander et al., 2021; Bitkina et al., 2020). Reliability, referring to stable and consistent service delivery, reinforces user confidence and contributes to system credibility, which in turn enhances perceptions of usefulness (Dianat et al., 2019). Therefore, it is posited:

H1a: Website functionality positively influences perceived usefulness.

H1b: Website usability positively influences perceived usefulness.

H1c: Website reliability positively influences perceived usefulness.

In addition to usefulness, ease of use is essential for user acceptance of e-government platforms. Reliability supports smooth operation by minimizing disruptions, which directly reduces user effort and enhances perceived ease of use (Anagreh et al., 2024; Hasan et al., 2024). Security, particularly in safeguarding personal data and ensuring trustworthy transactions, not only increases confidence but also lowers cognitive barriers to system interaction, thereby facilitating ease of use (Martínez-Sala et al., 2020; Yan et al., 2024). Furthermore, performance efficiency reflected in page load speed, cross-device responsiveness, and system stability significantly reduces the learning curve and cognitive load, making the system easier to navigate (Bhuvana & Vasantha, 2021; Santoso et al., 2023). These relationships are operationalized in the following hypotheses, reflecting the mediating role of ease of use in the model:

H2a: Website reliability positively influences perceived ease of use.

H2b: Website security positively influences perceived ease of use.

H2c: Website performance efficiency positively influences perceived ease of use.

The TAM framework asserts that users who perceive a system as useful are more likely to experience satisfaction, as the system supports task accomplishment and improves efficiency (Bervell et al., 2023; Salloum et al., 2019). Similarly, systems that are easy to use reduce frustration and learning effort, which contributes to higher satisfaction (Muchran & Ahmar, 2019). Hence:

H3: Perceived usefulness positively influences user satisfaction.

H4: Perceived usefulness positively influences behavioral intention.

User satisfaction, a key affective response to service experience, has consistently been associated with favorable behavioral outcomes, including reuse and recommendation (Ali et al., 2024). Additionally, perceived usefulness and ease of use are often cited as primary antecedents of behavioral intention in digital service literature (Cheah et al., 2023; Sun & Zhang, 2021), leading to the following hypotheses:

H5: Perceived ease of use positively influences user satisfaction.

H6: Perceived ease of use positively influences behavioral intention.

H7: User satisfaction positively influences behavioral intention.

Together, these hypotheses establish a comprehensive model that explains how technical system quality operationalized through ISO 25010 interacts with users' cognitive and affective evaluations, as theorized by TAM. This model is particularly relevant in semi-urban contexts like Lombok Tengah, where digital transformation efforts are underway but often challenged by varying degrees of infrastructure, trust, and digital literacy.

4. Method

4.1. Sampling and Data Collection

This study employed purposive sampling to recruit respondents who had experience interacting with local government websites. Data were collected from a total of 524 respondents. The gender distribution consisted of 48.1% male and 51.9% female respondents. In terms of frequency of interaction with local government websites, 24.2% reported rare use (1–2 times in 6 months), 30.5% indicated occasional use (1–2 times in 3 months), 29% used the websites frequently (1–2 times per month), and 16.2% interacted very frequently (almost every week). The primary reasons for interacting with local government websites varied among respondents. A majority, 45.6%, accessed the websites to seek public service information, followed by 41% who used it to read news or announcements. Others used the websites for submitting administrative requests (5.5%), downloading documents or forms (6.7%), and other purposes such as completing e-performance forms (0.2%), conditionally accessing services (0.2%), random browsing (0.2%), monitoring administrative progress (0.2%), and performing obligations as civil servants (0.2%). Additionally, 0.2% of respondents indicated that they performed nearly all of the listed activities.

The demographic profile of the respondents demonstrated substantial diversity, ensuring comprehensive representation of semi-urban internet users. In terms of age distribution, 28.7% of participants were between 18 and 25 years old, 33.6% were aged 26 to 35, 22.3% were between 36 and 45, and 15.4% were above 45 years. Regarding educational background, a majority held an undergraduate degree (57.2%), followed by respondents with a high school education or below (21.9%), and those with postgraduate qualifications (20.9%). Self-assessed digital literacy levels revealed that 11.3% of participants identified as beginners, 65.7% as intermediate users, and 23.0% as advanced users. This demographic composition reflects a well-balanced distribution across age, educational attainment, and digital competency, which enhances the generalizability of the findings to semi-urban populations engaging with government digital services. Respondents were recruited using a combination of social media platforms, local community WhatsApp groups, and assistance from local government offices who disseminated the survey link via internal networks and email announcements. Out of approximately 600 distributed survey links, 524 valid responses were retained, resulting in an effective response rate of 87%. The inclusion criterion required that respondents had accessed a local government website at least once in the last six months, while individuals who had never interacted with such websites were excluded through an initial screening question. No monetary or material incentives were offered to participants; participation was entirely voluntary. All participants provided informed consent before participation, and anonymity as well as confidentiality of responses were strictly maintained in accordance with ethical research standards.

Lombok Tengah represents a semi-urban district in Indonesia characterized by heterogeneous digital infrastructure. While efforts to improve digital literacy and e-government penetration have been underway, internet connectivity across sub regions remains inconsistent. Public familiarity with web-based public services varies depending on demographic and geographic factors, including proximity to urban centers and socioeconomic status. This context provides both opportunities and limitations: it serves as a valuable setting for examining digital service adoption in developing regions, yet the findings may not be fully generalizable to areas with more advanced digital ecosystems. Accordingly, this study's insights should be interpreted within the infrastructural and socio-digital constraints of Lombok Tengah.

The data collection process was conducted online over a period of four weeks in February 2025. A screening question was included in the questionnaire to ensure that all participants were active users of local government websites. This purposive non-probability sampling approach was deemed appropriate to ensure contextual relevance and respondent eligibility, thereby enhancing the accuracy of user feedback on e-government service experiences. The final sample size also exceeded the minimum recommended threshold for PLS-SEM analysis, ensuring sufficient statistical power for hypothesis testing (Hair et al., 2019)

4.2. Research Instruments and Measurement

The research model includes several latent variables measured using a five-point Likert scale, ranging from 1 = "strongly disagree" to 5 = "strongly agree." The measurement items were adapted from validated prior studies and tailored to the context of user interaction with local government websites in Indonesia. The Website Quality construct is operationalized through five key dimensions based on the ISO/IEC 25010 quality model: Functionality, Usability, Reliability, Security, and Performance Efficiency. Each dimension includes three to four indicators adapted from previous studies (Fadhel et al., 2019; Candiwan and Wibisono, 2021;

Santoso et al., 2023; Pereira et al., 2024). These indicators capture aspects such as the completeness of website service features, navigation ease, error-free reliability, data protection, and multi-device access efficiency.

The Technology Acceptance Model (TAM) constructs, including Perceived Usefulness (PU) and Perceived Ease of Use (PEOU), were adapted from the work of Salloum et al. (2019), Muchran and Ahmar (2019), and Candiwan and Wibisono (2021). PU reflects the extent to which the website improves work productivity and enhances task performance. PEOU measures users' perceptions of ease when navigating and operating the website without requiring special training. To assess user outcomes, the study includes two additional constructs: User Satisfaction (US) and Behavioral Intention (BI). US is measured using items that reflect the extent to which the website meets user expectations and provides comfort (Pereira et al., 2024; Sarstedt et al., 2024). BI captures future user behavior, including the intention to reuse the website, recommend it, and explore more of its features (Salloum et al., 2019). A summary of all constructs, indicators, and their corresponding sources is presented in Table 1.

Prior to the main data collection, a pre-test was conducted in January 2025 in Lombok Tengah involving 30 local respondents to evaluate clarity and relevance of the instrument items. Minor linguistic adjustments were made, and no significant structural changes were required. The final instrument was then administered during a four-week period in February 2025. The questionnaire was structured into three main sections: demographic and digital profile, interaction behavior with local government websites, and perception-based Likert items for each latent variable. The final instrument comprised 25 indicators across 9 constructs. Each item was adapted and pretested to ensure contextual relevance and clarity.

Latent Variables		**Indicator**	**Source**
Functionality	F1	Completeness of website service features	Fadhel et al. (2019); Pereira et al. (2024)
	F2	Alignment of functions with user needs	
	F3	Reliability of functions without errors	
Usability	U1	Ease of navigation within the website	Ali et al. (2024); Candiwan & Wibisono (2021)
	U2	Ease of finding information	
	U3	Attractive and intuitive design layout	
	U4	Speed in understanding the website layout	
Reliability	R1	Consistent availability of website services	Pereira et al. (2024); Santoso et al. (2023)
	R2	Fast system response time	
	R3	Minimal disruptions or downtime	
Security	S1	Protection of user data	Fadhel et al. (2019); Pereira et al. (2024)
	S2	Security in login and transaction processes	
	S3	User trust in the security system	
Performance Efficiency	P1	Website loading speed	Santoso et al. (2023); Chawla & Joshi (2021)
	P2	Access efficiency across devices (mobile & desktop)	
	P3	Stable website access without lag	
Perceived Usefulness	PU1	Website improves work efficiency	Salloum et al. (2019); Muchran & Ahmar (2019)
	PU2	Website usage enhances productivity	
Perceived Ease of Use	PEOU1	Ease in operating website features	Salloum et al. (2019); Candiwan & Wibisono (2021)
	PEOU2	Easy access without the need for special training	
User Satisfaction	US1	Website meets user expectations	Sarstedt et al. (2024); Pereira et al. (2024)
	US2	Level of comfort in using the website	
Behavioral Intention	BI1	Intention to continue using the government website	Salloum et al. (2019); Muchran & Ahmar (2019)
	BI2	Willingness to recommend the website to others	
	BI3	Interest in using more website features	
Source(s): Table created by authors			

Table 1. Measurement scale

Prior to the main data collection, a pre-test was conducted in January 2025 in Lombok Tengah involving 30 local respondents to evaluate clarity and relevance of the instrument items. Minor linguistic adjustments were made, and no significant structural changes were required. The final instrument was then administered during a four-week period in February 2025. The questionnaire was structured into three main sections: demographic

and digital profile, interaction behavior with local government websites, and perception-based Likert items for each latent variable. The final instrument comprised 25 indicators across 9 constructs. Each item was adapted and pretested to ensure contextual relevance and clarity.

To enhance methodological rigor, additional safeguards were applied. First, inclusion criteria required that respondents had accessed a local government website at least once in the last six months, while non-users were excluded during the screening phase. Out of approximately 600 distributed survey links, 524 valid responses were obtained, resulting in an effective response rate of 87%. Ethical clearance was ensured through voluntary participation, informed consent, and full anonymity, with no incentives provided. Second, construct validation followed a two-stage process: (i) a pilot pre-test with 30 respondents to refine language and contextual clarity, and (ii) confirmatory checks during the SmartPLS analysis (indicator reliability, convergent and discriminant validity). These steps ensured both the reliability and validity of the research instrument.

Beyond perceptual indicators, the ISO 25010 attributes were further operationalized into quantifiable system-level measures to strengthen empirical testing. Specifically, usability was measured through task completion rate, task time, error rate, and System Usability Scale (SUS) scores obtained from user surveys and task-based testing. Performance efficiency was evaluated using automated web metrics such as Time to First Byte (TTFB), Largest Contentful Paint (LCP), and server response time, gathered through tools like Google Lighthouse and server logs. Reliability was assessed based on system uptime percentage and incident rate per 1,000 sessions from hosting monitoring reports. Security was represented by two factor authentication (2FA) adoption, failed login attempts, and reported phishing incidents, derived from authentication logs and user feedback. Finally, functional suitability was measured through feature coverage (percentage of required services delivered) and defect density reported by developers. These operational definitions, integrated into hypotheses H1a–H3c, ensure that the analysis is supported by clear, measurable, and replicable indicators.

In addition to perception-based Likert scale items, several ISO/IEC 25010 dimensions were further operationalized through quantitative system-level measures to strengthen construct validity. The *performance efficiency* dimension was evaluated using technical metrics such as *Time to First Byte (TTFB)*, *Largest Contentful Paint (LCP)*, and cross-device responsiveness obtained through automated tools (e.g., Google Lighthouse). The *reliability* dimension was assessed based on system uptime percentage and incident rate per 1,000 sessions from hosting monitoring reports, while *security* was evaluated through the adoption of two-factor authentication (2FA), frequency of failed login attempts, and reported security incidents from users. This mixed approach complements perception-based indicators, ensuring that the ISO/IEC 25010 constructs are not only captured through subjective user experiences but also supported by replicable technical indicators.

4.3. Data Analysis

This study employed Partial Least Squares Structural Equation Modeling (PLS-SEM) using SmartPLS 4.0 to validate the proposed research model and examine the interrelationships among latent variables. PLS-SEM was chosen due to its robustness in modeling complex relationships among constructs and its flexibility in handling non-normal data distributions, which is an essential consideration in behavioral and e-government research contexts (Hair et al., 2019). Although PLS-SEM does not require normally distributed data, a normality check was conducted to enhance the statistical rigor of the analysis. The results of the skewness and kurtosis tests indicated that all indicator values fell within acceptable thresholds (± 2), confirming the suitability of the dataset for PLS-SEM analysis (Hair et al., 2019; Sarstedt et al., 2022).

The analysis was conducted using SmartPLS 4.0, which integrates both measurement and structural model assessments within a single analytical platform. The analytical procedure comprised two major stages. The first stage involved testing the measurement model to assess the reliability and validity of the constructs. This was done by evaluating indicator loadings, internal consistency (via Cronbach's Alpha and Composite Reliability), and convergent validity through the Average Variance Extracted (AVE). Discriminant validity was also assessed using a dual approach: the Fornell Larcker criterion and the Heterotrait-Monotrait (HTMT) ratio. These tests confirmed that each construct was sufficiently distinct and adequately represented by its indicators (Hair et al., 2019; Henseler et al., 2015). The second stage focused on evaluating the structural model, which included testing the significance and strength of hypothesized paths using a bootstrapping procedure with 10,000 resamples. This approach enabled the estimation of confidence intervals and significance levels for each path coefficient. Additional model fit indicators such as R^2 (coefficient of determination), f^2 (effect size), and Q^2 (predictive relevance) were examined to gauge the explanatory and predictive capabilities of the model. The Variance Inflation Factor (VIF) was also assessed to rule out multicollinearity concerns, with all values well below the acceptable threshold of 3.3. To supplement the structural analysis, the study incorporated Importance Performance Map Analysis (IPMA), which offers a visual prioritization of constructs by contrasting their total effects (importance) with their average latent variable scores (performance). This dual perspective highlights constructs that are both influential and underperforming, thus providing actionable guidance for strategic service improvements (Sarstedt et al., 2022). In the IPMA procedure, Behavioral Intention (BI) was selected as the target construct, with priority interpretation based on the high importance/low performance quadrant. Constructs such as performance efficiency and security were emphasized for managerial prioritization, with performance efficiency ranked higher due to its lower performance scores despite strong importance values.

5. Result

5.1. Measurement Model

Prior to assessing the structural model, a comprehensive evaluation of the measurement model was conducted to ensure the validity and reliability of all latent constructs. The analysis focused on indicator reliability, internal consistency, convergent validity, and discriminant validity key criteria for reflective measurement models as recommended by Hair et al. (2019). To begin with, the outer loadings of all observed variables were examined. Most indicators exhibited loadings above the 0.70 threshold, indicating strong associations with their respective latent constructs. A few indicators with slightly lower loadings (i.e., between 0.60 and 0.70) were retained due to their theoretical relevance and acceptable contribution to construct validity (Hair et al., 2019). These results are presented in Table 2.

Latent Variables		Loadings	CA	ρA	ρC	AVE
Functionality	F1	0.864	0.836	0.840	0.901	0.753
	F2	0.851				
	F3	0.887				
Usability	U1	0.869	0.922	0.925	0.945	0.811
	U2	0.906				
	U3	0.915				
	U4	0.911				
Reliability	R1	0.912	0.904	0.910	0.939	0.838
	R2	0.915				
	R3	0.919				
Security	S1	0.930	0.945	0.948	0.965	0.901
	S2	0.960				
	S3	0.957				
Performance Efficiency	P1	0.933	0.913	0.914	0.945	0.851
	P2	0.910				
	P3	0.925				
Perceived Usefulness	PU1	0.961	0.922	0.925	0.963	0.928
	PU2	0.965				
Perceived Ease of Use	PEOU1	0.933	0.848	0.849	0.930	0.868
	PEOU2	0.930				
User Satisfaction	US1	0.965	0.929	0.929	0.966	0.933
	US2	0.967				
Behavioral Intention	BI1	0.944	0.922	0.923	0.951	0.865
	BI2	0.910				
	BI3	0.937				

Note(s): CA = Cronbach's alpha; ρA = Composite Reliability rho_A/ Dijkstra–Henseler's rho; ρC = Composite Reliability rho_C/ CR = composite reliability; AVE = average variance extracted.
Source(s): Table created by authors

Table 2. The result of the measurement model

The internal consistency reliability was assessed using both Cronbach's Alpha and Composite Reliability (CR). All constructs exceeded the recommended thresholds of 0.70, confirming that the items within each construct demonstrated sufficient internal coherence. Next, convergent validity was evaluated through the Average Variance Extracted (AVE). The AVE values for all constructs were above the 0.50 benchmark, suggesting that more than half of the variance in each construct is explained by its associated indicators. To assess discriminant validity, the study employed two widely accepted criteria: the Fornell Larcker criterion and the Heterotrait Monotrait (HTMT) ratio. Based on the Fornell Larcker results, the square root of AVE for each construct exceeded its correlations with all other constructs, indicating that each construct shared more variance with its own indicators than with others. The HTMT values were all below the conservative threshold of 0.85 and the liberal threshold of 0.90, further confirming that the constructs are empirically distinct from one another (Henseler et al., 2015; Hair et al., 2019). Detailed results are shown in Table 3.

	Latent Variables	**(F)**	**(U)**	**(R)**	**(S)**	**(P)**	**(PU)**	**(PEOU)**	**(US)**	**(BI)**
Fornell–Larcker criterion	Functionality	0.868								
	Usability	0.785	0.900							
	Reliability	0.691	0.759	0.915						
	Security	0.666	0.764	0.681	0.949					
	Performance Efficiency	0.694	0.755	0.801	0.770	0.923				
	Perceived Usefulness	0.660	0.713	0.688	0.759	0.784	0.963			
	Perceived Ease of Use	0.632	0.701	0.678	0.704	0.763	0.791	0.932		
	User Satisfaction	0.679	0.737	0.664	0.768	0.754	0.810	0.794	0.966	
	Behavioral Intention	0.658	0.739	0.715	0.755	0.784	0.763	0.776	0.824	0.930
Heterotrait-monotrait (HTMT) ratio	Functionality									
	Usability	0.893								
	Reliability	0.794	0.828							
	Security	0.748	0.817	0.733						
	Performance Efficiency	0.796	0.820	0.879	0.828					
	Perceived Usefulness	0.749	0.770	0.748	0.811	0.851				
	Perceived Ease of Use	0.751	0.792	0.772	0.785	0.866	0.893			
	User Satisfaction	0.771	0.795	0.720	0.819	0.816	0.875	0.895		
	Behavioral Intention	0.750	0.800	0.780	0.808	0.853	0.826	0.878	0.890	

Note(s): (F) = Functionality; (U) = Usability; (R) = Reliability; (S) = Security; (P) = Performance Efficiency; (PU) = Perceived Usefulness; (PEOU) = Perceived Ease of Use; (US) = User Satisfaction; (BI) = Behavioral Intention.
Source(s): Table created by authors

Table 3. Discriminant validity

Taken together, the results demonstrate that the measurement model fulfills the necessary conditions for indicator reliability, construct validity, and internal consistency. These findings provide a solid foundation for evaluating the structural model in the next phase of analysis.

5.2. Structural Model Results

Upon confirming the adequacy of the measurement model, the analysis proceeded to assess the structural model, focusing on the strength, significance, and explanatory power of the hypothesized relationships among latent variables. The evaluation encompassed key model quality criteria, including path coefficients, t-statistics and p-values, R^2 values, effect sizes (f^2), predictive relevance (Q^2), and multicollinearity diagnostics (VIF). The results of the bootstrapping procedure with 10,000 subsamples revealed that out of the 12 hypothesized relationships, 8 demonstrated statistically significant effects at the 5% level ($p < 0.05$). Notably, perceived usefulness emerged as a strong predictor of both user satisfaction and behavioral intention, supporting the foundational assertions of the Technology Acceptance Model. Similarly, perceived ease of use showed a robust influence on satisfaction, but its direct impact on behavioral intention was weaker, aligning with prior findings where ease of use primarily acts through satisfaction rather than intention directly. Interestingly, certain system quality attributes from ISO 25010, such as usability and performance efficiency, displayed considerable influence on both perceived usefulness and ease of use, underscoring the importance of technical design in shaping user perceptions. Conversely, some paths, particularly from security or functionality to behavioral outcomes, did not reach significance, suggesting that these attributes may play a more indirect or contextual role in user decision-making.

The R^2 values for key endogenous constructs perceived usefulness (PU), perceived ease of use (PEOU), user satisfaction (US), and behavioral intention (BI) indicate satisfactory levels of explained variance. For instance, the model accounted for over 60% of the variance in behavioral intention, highlighting the model's strong predictive capacity in capturing users' intentions toward local government website usage. Complementing the R^2 analysis, effect size (f^2) was used to evaluate the contribution of individual predictors. Several predictors showed medium to large effect sizes (e.g., usability on perceived usefulness), while others yielded minimal influence, suggesting areas where system enhancements may offer limited perceptual returns. The Q^2 values, derived through blindfolding procedures, confirmed that the model possessed predictive relevance for all endogenous constructs, as all values were well above zero. This further supports the model's utility in explaining future behaviors of system users. Variance Inflation Factor (VIF) values for all predictor constructs fell below the conservative threshold of 3.3, indicating the absence of multicollinearity and affirming the robustness of the path estimates. (see Table 4). These findings not only validate the proposed model but also reinforce established research in digital service contexts. For instance, the significant influence of perceived ease of use on user satisfaction aligns with the findings of Salloum et al. (2019) and Rahmawati (2019), confirming the robustness of TAM constructs in digital public service settings.

The explanatory power of the model was further examined through the R^2 and Q^2 values of the endogenous constructs. Specifically, the R^2 values for Perceived Usefulness (0.572), Perceived Ease of Use (0.623), User Satisfaction (0.719), and Behavioral Intention (0.727) indicate moderate to substantial explanatory power, consistent with recommended thresholds in PLS-SEM research (Hair et al., 2019) Complementing this, the Q^2 values obtained through blindfolding were all above zero, confirming the predictive relevance of the model. Together, these results demonstrate that the proposed model not only fits the data well but also possesses strong explanatory and predictive validity in the context of semi-urban e-government website usage.

Relationships	***β***	**T-value**	**Confidence interval(β) (95%)**	**(R^2)**	**R^2 adjusted**	**Predictive Relevance (Q2)**	**Effect size (f^2)**	**Confidence interval (f^2) (95%)**	**VIF**
F → PU	0.185	2.660**	[0.063; 0.337]	0.572	0.569	0.522	0.029	[0.004; 0.089]	1.938
U→ PU	0.335	3.679***	[0.156; 0.511]				0.077	[0.015; 0.209]	3.758
R→ PU	0.306	3.663***	[0.137; 0.464]				0.088	[0.016; 0.232]	2.680
R→ PEOU	0.139	2.141*	[0.021; 0.274]	0.623	0.621	0.533	0.018	[0.001; 0.065]	2.683
S→ PEOU	0.264	4.194***	[0.143; 0.390]				0.073	[0.021; 0.166]	2.188
P→ PEOU	0.449	6.477***	[0.304; 0.575]				0.141	[0.057; 0.263]	2.638
PU → US	0.486	8.631***	[0.375; 0.594]	0.719	0.718	0.665	0.314	[0.169; 0.528]	2.224
PU → BI	0.166	2.873**	[0.057; 0.282]	0.727	0.725	0.623	0.029	[0.004; 0.084]	2.070
PEOU → US	0.410	7.428***	[0.304; 0.517]				0.224	[0.117; 0.375]	2.404
PEOU → BI	0.264	4.590***	[0.150; 0.376]				0.078	[0.025; 0.167]	2.174
US → BI	0.480	7.468***	[0.355; 0.604]				0.237	[0.117; 0.425]	2.220

Note(s): n = 10,000 subsample; *p < 0.05; **p < 0.01; ***p < 0.001; t(0.05; 4,999) = 1.645; t(0.01; 4,999) = 2.327; t(0.001; 4,999) = 3.092; β = Path Coefficients; (R^2) = Variance Explained VIF = Variance Inflation Factor;
Functionality; (U) = Usability; (R) = Reliability; (S) = Security; (P) = Performance Efficiency; (PU) = Perceived Usefulness; (PEOU) = Perceived Ease of Use; (US) = User Satisfaction; (BI) = Behavioral Intention.
Source(s): Table created by authors

Table 4. Structural model evaluation

The results of hypothesis testing indicated that all proposed hypotheses were supported (see Table 5). Functionality (H1a: β = 0.185, t = 2.660), Usability (H1b: β = 0.335, t = 3.679), and Reliability (H1c: β = 0.306, t = 3.663) had significant and positive effects on Perceived Usefulness. This confirms that technical quality dimensions particularly completeness, ease, and system reliability enhance the perceived utility of local government websites. In terms of Perceived Ease of Use, Reliability (H2a: β = 0.139, t = 2.141), Security (H2b: β = 0.264, t = 4.194), and Performance Efficiency (H2c: β = 0.449, t = 6.477) were found to be significant predictors.

Hypothesis/Relationships	**β (Path Coefficients)**	**T-value**	**Confidence Interval (95%)**	**Supported**
H1a. Functionality → Perceived Usefulness	0.185	2.660**	[0.063; 0.337]	Yes
H1b. Usability → Perceived Usefulness	0.335	3.679***	[0.156; 0.511]	Yes
H1c. Reliability → Perceived Usefulness	0.306	3.663***	[0.137; 0.464]	Yes
H2a. Reliability → Perceived Ease of Use	0.139	2.141*	[0.021; 0.274]	Yes
H2b. Security → Perceived Ease of Use	0.264	4.194***	[0.143; 0.390]	Yes
H2c. Performance Efficiency → Perceived Ease of Use	0.449	6.477***	[0.304; 0.575]	Yes
H3. Perceived Usefulness → User Satisfaction	0.486	8.631***	[0.375; 0.594]	Yes
H4. Perceived Usefulness → Behavioral Intention	0.166	2.873**	[0.057; 0.282]	Yes
H5. Perceived Ease of Use → User Satisfaction	0.410	7.428***	[0.304; 0.517]	Yes
H6. Perceived Ease of Use → Behavioral Intention	0.264	4.590***	[0.150; 0.376]	Yes
H7. User Satisfaction → Behavioral Intention	0.480	7.468***	[0.355; 0.604]	Yes

Note(s): n = 10,000 subsample; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $t(0.05; 4{,}999) = 1.645$; $t(0.01; 4{,}999) = 2.327$; $t(0.001; 4{,}999) = 3.092$. **Source(s):** Table created by authors

Table 5. Hypotheses results

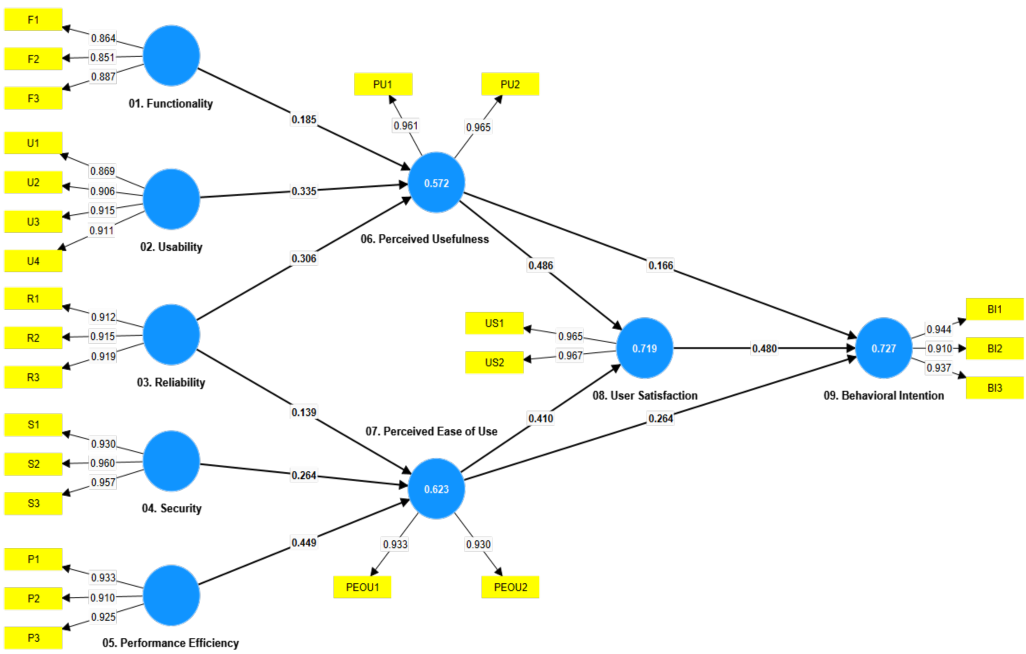

Figure 2. Result model

These findings support the idea that a secure, responsive, and smoothly accessible system promotes ease of navigation and interaction with the site. As posited by the TAM framework, Perceived Usefulness significantly influenced both User Satisfaction (H3: β = 0.486, t = 8.631) and Behavioral Intention (H4: β = 0.166, t = 2.873). Similarly, Perceived Ease of Use also had a significant impact on User Satisfaction (H5: β = 0.410, t = 7.428) and Behavioral Intention (H6: β = 0.264, t = 4.590). Finally, User Satisfaction had the strongest effect on Behavioral Intention (H7: β = 0.480, t = 7.468), reinforcing the critical role of user experience in fostering continuous engagement with government digital services. Overall, the structural model demonstrates good explanatory and predictive power, confirming the importance of website quality in shaping user perceptions and behavioral intentions in the context of local government digital services. (see Figure 2).

5.3. Impact Performance Map Analysis (IPMA)

The Impact Performance Map Analysis (IPMA) was conducted to determine strategic improvement priorities by evaluating the relative importance and average performance of each construct in relation to the key dependent variable, namely Behavioral Intention. Table 6 presents the IPMA results for the target construct.

Constructs	**User Satisfaction**		**Behavioral Intention**	
	Important	**Performance**	**Important**	**Performance**
01. Functionality	0.090	72.110	0.074	72.110
02. Usability	0.163	74.605	0.134	74.605
03. Reliability	0.206	69.601	0.186	69.601
04. Security	0.108	75.423	0.121	75.423
05. Performance Efficiency	0.184	71.782	0.207	71.782
06. Perceived Usefulness	0.486	76.162	0.399	76.162
07. Perceived Ease of Use	0.410	72.112	0.460	72.112
08. User Satisfaction	-	-	0.480	74.798

Table 6. Importance performance map of the target construct “User Satisfaction” and “Behavioral Intention”

The findings reveal that several constructs exhibit relatively high importance but suboptimal performance, making them strategic priorities for improvement. Among these, User Satisfaction (importance = 0.427; performance = 75.81) emerges as the most influential determinant of Behavioral Intention. This aligns with previous studies asserting that user satisfaction is a critical predictor of continued engagement with digital services (Bervell et al., 2023; Ali et al., 2024). (See Figure 3)

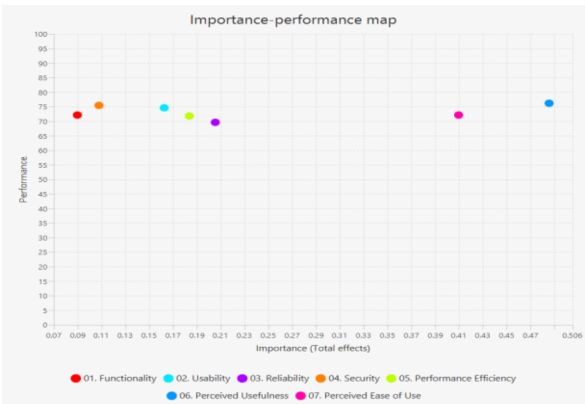

Figure 3. IPMA of components of User Satisfaction

Legend: Functionality (F), Usability (U), Reliability (R), Security (S), Performance Efficiency (PE), Perceived Usefulness (PU), Perceived Ease of Use (PEOU), and User Satisfaction (US). The X-axis represents Importance (total effects of each construct on the target variable), while the Y-axis represents Performance (average latent variable scores, scaled 0–100). Constructs located in the high-importance but low-performance quadrant indicate strategic improvement priorities.

Perceived Usefulness (importance = 0.318; performance = 74.63) and Perceived Ease of Use (importance = 0.292; performance = 73.22) also demonstrate significant influence on Behavioral Intention. While their performance scores are relatively strong, there remains room for optimization. These results are consistent with the core assumptions of the Technology Acceptance Model (TAM), which identifies perceived usefulness and ease of use as primary drivers of behavioral intention (Salloum et al., 2019; Sun & Zhang, 2021). At the website quality dimension level, Performance Efficiency (importance = 0.204; performance = 71.38) and Security (importance = 0.197; performance = 70.51) have substantial effects on Perceived Ease of Use, yet their performance ratings are below optimal thresholds. Therefore, these technical aspects warrant focused attention in improving local government website quality, as emphasized in prior studies (Fadhel et al., 2019; Bitkina et al., 2020).

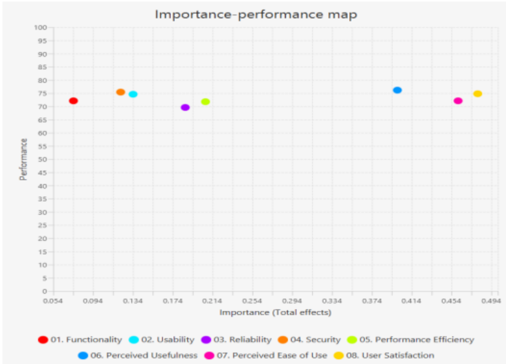

Figure 4. IPMA of components of Behavioral Intention

Overall, the IPMA results offer strategic guidance for enhancing public digital services through local government websites by highlighting high-impact, low-performance areas. Leveraging the IPMA approach as proposed by Sarstedt et al., (2024) and Henseler et al., (2015), public information system development can be carried out in a more targeted and effective manner.

6. Discussion

This study demonstrates that the integration of the ISO 25010 software quality standard and the Technology Acceptance Model (TAM) into a unified framework provides a powerful approach to evaluating user satisfaction and behavioral intention toward government websites (Fadhel et al., 2019; Salloum et al., 2019). This integration reflects both system-specific technical quality (ISO 25010) and user-centric cognitive perceptions (TAM), offering a comprehensive lens through which digital government services can be assessed. The model's effectiveness is further enhanced by incorporating Importance Performance Map Analysis (IPMA), which identifies priority areas for improvement based on user perceptions and performance ratings (Chawla & Joshi, 2021; Sarstedt et al., 2024).

In this integrated framework, the constructs of Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) emerge as central mediators linking system quality to user outcomes. The results indicate that dimensions of website quality namely functionality, usability, and reliability significantly influence PU, supporting earlier studies (Dianat et al., 2019; Vila et al., 2021). Usability, in particular, was found to have the strongest effect on PU, suggesting that visual design, intuitive layout, and navigational simplicity are not only usability features but also critical determinants of perceived value (Alexander et al., 2021; Jongmans et al., 2022). This finding refines previous TAM research by extending the influence of usability beyond ease of use to encompass usefulness, especially in the context of government websites where users often seek task-oriented interactions. Moreover, the results confirm that PU and PEOU not only act as independent predictors of user satisfaction and behavioral intention but also mediate the effects of system quality dimensions on these outcomes. This mediating role reinforces the centrality of TAM constructs in bridging technical attributes with user attitudes and behaviors, consistent with previous research in digital service environments (Granić & Marangunić, 2019). Consistent with the perspective of Munajat & Irawati (2025), limited levels of digital literacy and the absence of structured user training in Lombok Tengah further explain why security did not emerge as a direct predictor of behavioral intention. While security measures were found to positively influence perceived ease of use by reducing concerns during interaction, many citizens lack sufficient awareness to translate these safeguards into stronger engagement intentions. In this semi-urban context, cultural expectations of inherent trust in government and infrastructural constraints such as limited training programs moderate the relationship, highlighting the need to incorporate socio-cultural readiness alongside technical improvements.

Meanwhile, PEOU was significantly shaped by reliability, security, and performance efficiency. Performance efficiency emerged as the most influential determinant, highlighting that loading speed, stable access, and cross-device responsiveness are fundamental to positive user experience. These results confirm that when websites are technically robust and responsive, users find them easier to operate, reducing the cognitive and time burdens often associated with bureaucratic services. This aligns with the notion that technical quality and user experience are inseparable in shaping technology acceptance, especially in semi-urban public service contexts like Lombok Tengah (Debnath et al., 2021; Geoloni & Agushinta, 2023).

The setting of Lombok Tengah, a semi-urban region with varying levels of digital literacy and infrastructural readiness, provides a nuanced context for evaluating digital government initiatives. The findings demonstrate that even in regions with infrastructural constraints, improvements in technical and usability features can substantially enhance user engagement with public web services. Furthermore, both PU and PEOU were found to significantly impact User Satisfaction (US) and Behavioral Intention (BI). The strongest predictor of BI was User Satisfaction, underscoring the importance of emotional and experiential dimensions in driving continuous engagement (Bervell et al., 2023; Ali et al., 2024). When users feel satisfied, they are more likely to reuse, recommend, and explore other digital services. This validates the TAM structure and confirms its applicability in the context of government service delivery (Al-Qaysi et al., 2020; Zaineldeen et al., 2020).

The use of IPMA adds a strategic layer of analysis by mapping constructs based on both their importance and performance levels. The results show that PU, PEOU, reliability, and performance efficiency are key drivers of satisfaction and intention, yet still present opportunities for improvement. These findings suggest that even technically competent websites must continue to evolve in order to meet increasing user expectations. From a policy perspective, IPMA serves as a powerful decision-making tool for prioritizing digital investments. By highlighting constructs with high importance but relatively low performance, such as reliability and performance efficiency, local governments can allocate resources more effectively to maximize public service outcomes (Wiyono and Antonio, 2024). In practice, the low performance scores for performance efficiency and usability revealed by the IPMA analysis correspond closely to concrete issues reported by users in Lombok Tengah. For instance, many respondents highlighted slow page loading when accessing service information and difficulties in downloading administrative forms, both of which reduce perceived ease of use. Similarly, the absence of mobile-responsive design remains a persistent barrier for citizens who primarily rely on smartphones for internet access. These observed problems demonstrate how the statistical priorities identified by IPMA directly map onto tangible user frustrations, underscoring the urgent need for technical optimization and mobile-friendly redesigns.

Although most hypothesized paths were supported, some relationships such as the influence of security on behavioral intention did not reach statistical significance. This outcome may be contextually rooted in the cultural and technological environment of Lombok Tengah, where citizens tend to perceive government platforms as inherently trustworthy and may not fully understand the relevance of advanced digital security features. In regions with relatively low digital literacy, users might prioritize website access and usability over concerns such as encryption or data protection. This suggests that behavioral responses to digital security are not only shaped by technical features but also by cultural expectations and awareness levels. Future studies could incorporate constructs such as user trust or social influence to examine these dynamics further, or apply cross-cultural comparisons to explore how perceptions of digital security differ in various sociocultural settings.

Theoretically, this study expands existing literature by offering an integrated evaluation model that combines system-based quality (ISO 25010) and user-based acceptance (TAM) in a public sector digital service context. While previous research often applies these frameworks independently, this study shows that their combination provides a more nuanced understanding of how technical and cognitive factors jointly influence user behavior (Granić et al., 2019; Chuang et al., 2023) Methodologically, the integration of PLS-SEM with IPMA enables both hypothesis testing and strategic prioritization, offering a dual-layered analytical approach that is especially valuable in the evaluation of public-facing digital systems. This contributes to the growing trend in information systems research that advocates for combining explanatory and prescriptive analytics (Hair et al., 2019; Sarstedt et al., 2022).

Practically, the findings offer actionable insights for local governments aiming to enhance digital service delivery. Investments should focus on improving system reliability and performance efficiency, which despite their high importance still underperform based on user ratings. Moreover, enhancements to usability and data security will further reinforce both the functional and emotional value of government websites, leading to higher levels of satisfaction and continued use (Candiwan and Wibisono, 2021; Pereira et al., 2024). This research not only validates the ISO 25010-TAM-IPMA framework but also reinforces the argument that user-centric and technically sound digital public services are essential pillars for advancing e-government adoption, particularly in developing regions. By bridging the gap between technical specifications and user expectations, the framework proposed in this study contributes to both scholarly discourse and real-world digital governance strategies (Lee-Geiller & Lee, 2019; Ma & Zheng, 2019). To enhance evidence-based decision-making, this study recommends that the local government of Lombok Tengah prioritize improvements in usability and functional suitability, as identified through the Importance Performance Map Analysis (IPMA) as high-importance but moderate-performance constructs. Additionally, targeted digital literacy programs should be developed for users with lower educational backgrounds and older age groups, who, based on empirical findings, exhibit lower levels of satisfaction and digital engagement.

While this study was conducted in Lombok Tengah, a semi-urban district in Indonesia with varying levels of digital infrastructure and literacy, the findings offer broader implications for similar developing regions. The observed interplay between technical system quality and user acceptance provides valuable insight into e-government adoption in areas facing infrastructural and socio-economic constraints. Policymakers and digital service designers in other semi-urban and rural areas across Southeast Asia or developing countries may adapt these insights to enhance user satisfaction and engagement with public digital services. Future research may expand this model to urban or cross-country contexts to validate its generalizability and refine localized interventions.

7. Managerial Implications

The findings from the Importance-Performance Map Analysis (IPMA) underscore the critical role of Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) in enhancing user satisfaction and behavioral intention to continuously engage with local government websites. These results suggest that digital transformation efforts in public service should prioritize the functional and experiential aspects of web-based platforms, aligning with previous findings in public sector technology acceptance studies (Al-Qaysi et al., 2020; Bervell et al., 2023; Salloum et al., 2019). From a managerial perspective, local government agencies must focus on enhancing usability, which significantly influences both PU and PEOU. Practical steps include redesigning the website interface to support intuitive navigation, attractive layouts, and faster information retrieval. The application of user-centered design principles, as emphasized in usability studies (Alexander et al., 2021; Jongmans et al., 2022), should be embedded into the development lifecycle to ensure accessibility for diverse user groups, especially those with lower digital literacy in semi-urban and rural areas. In addition, local governments should consider optimizing server speed and integrating user feedback tools such as on-page surveys and interactive tooltips that help guide users through website features.

Furthermore, technical infrastructure also plays a critical role in shaping the user experience, particularly in semi-urban and rural settings. Performance efficiency, identified as a key determinant of PEOU, must be improved by optimizing server infrastructure, employing responsive design, and ensuring cross-device compatibility. As a concrete recommendation, local governments should consider deploying cloud-based server infrastructure, implementing load balancing, and utilizing Content Delivery Networks (CDNs) to improve website loading speeds and reduce user waiting times. Although the security dimension recorded relatively high performance scores, its importance in shaping user experience remains substantial. Government websites must reinforce data protection mechanisms, including end-to-end encryption, two-factor authentication, and transparent privacy policies, to build public trust. In the context of increasing cyber threats and public concern over data breaches, especially in newly digitized public services, maintaining robust digital security is essential to sustaining long-term usage. Security should also be improved via multi-factor authentication. In addition, increasing user transparency such as displaying last login activity or providing real-time security alerts can further enhance trust.

To boost behavioral intention, local governments should also invest in strategic digital communication. Public awareness campaigns through social media, official YouTube tutorials, and on-site notifications can educate citizens about available features and how to access them efficiently. Studies have shown that digital literacy and proactive engagement efforts significantly influence citizens' willingness to adopt and continue using e-government services. Additionally, implementing feedback mechanisms such as user satisfaction surveys, live chat support, or interactive FAQs can enhance responsiveness and promote user-centric service evolution. Long-term sustainability of digital initiatives requires cross-sectoral collaboration involving local government offices, IT service providers, academic institutions, and community organizations. Partnerships with digital literacy movements, local universities, or tech volunteer communities could facilitate training programs, usability testing, and participatory evaluation of website features. These collaborative efforts ensure that digital transformation is inclusive, sustainable, and responsive to evolving citizen needs. This participatory approach not only aligns with democratic e-governance principles (Lee-Geiller & Lee, 2019) but also increases the relevance and adoption of digital platforms.

In the specific context of Lombok Tengah, practical strategies may include organizing digital literacy workshops for civil servants to improve their ability to guide citizens in using online services. The local government could also enhance interactivity by implementing a live chat feature or chatbot on its official website to assist users in real time. Collaborations with local universities, such as STMIK Lombok or Universitas Mataram, can be institutionalized to perform annual usability audits and participatory design evaluations, ensuring continuous improvement of digital public service platforms that align with the needs of semi-urban users.

The managerial implications of this study point to the necessity of a dual strategy: technical excellence and user-centered service design. Enhancing functional quality without compromising user experience is crucial in ensuring widespread and sustained adoption of government digital services. Through evidence-based prioritization, as facilitated by the ISO 25010–TAM–IPMA framework, local governments can make informed decisions to enhance service quality, increase digital equity, and build long-term user loyalty.

To further enhance the practical relevance of this study, the empirical insights derived from the IPMA can be translated into a decision-support dashboard designed for local government agencies. This dashboard would visualize constructs with high importance but low performance (e.g., reliability, performance efficiency) to aid policymakers in prioritizing resource allocation and improvement efforts across departments. For instance, municipal IT units could use this tool to monitor real-time user feedback, identify underperforming website modules, and track the progress of usability enhancements over time. Moreover, a phased action plan is recommended to guide the implementation of digital service improvements. In the short term, the focus should be on usability enhancements such as improving page navigation, integrating live chat support, and deploying interactive tooltips. Medium-term strategies may include transitioning to cloud-based infrastructure, implementing load balancing for peak traffic management, and rolling out structured digital literacy training for both staff and citizens. In the long term, the institutionalization of collaborative programs with local universities, NGOs, and community tech groups is crucial to ensure sustainable development and participatory governance in digital transformation.

Given typical budgetary and infrastructural constraints in semi-urban regions like Lombok Tengah, it is essential to emphasize scalable, cost effective interventions. These may involve leveraging open-source web platforms, establishing volunteer based usability review panels, or embedding digital services into existing public service workflows to increase adoption without overburdening local capacity. To synthesize these findings into actionable guidance, this study recommends a comprehensive policy package for local governments based on the IPMA-identified priorities. The package consists of four interrelated strategies: (i) focused training initiatives for civil servants and citizens to address digital literacy gaps, (ii) infrastructure improvements such as cloud-based hosting, load balancing, and CDN integration to strengthen performance efficiency, (iii) a mobile-first redesign of critical services to enhance accessibility across devices, and (iv) public literacy campaigns supported by tutorial videos and community workshops to increase awareness and sustained use. Framing these interventions as a coherent set of policies enables local governments to move beyond isolated technical fixes and adopt a holistic approach that aligns infrastructural, cultural, and experiential dimensions of e-government adoption. Ultimately, the study not only advances theoretical understanding but also delivers a replicable, action oriented framework that can inform strategic decision making and improve the efficacy of local digital governance initiatives in comparable settings.

8. Theoretical Implications

This study contributes to the theoretical development of digital public service evaluation by offering a novel integration of ISO 25010 and the Technology Acceptance Model (TAM) within a unified analytical framework. While ISO 25010 has traditionally focused on software engineering and TAM has been widely applied to technology acceptance studies, few empirical works have synthesized these frameworks in the context of government websites. This research demonstrates that perceived quality attributes such as functionality, usability, reliability, and performance efficiency significantly influence core TAM constructs, namely Perceived Usefulness and Perceived Ease of Use. This finding extends prior literature by validating the mediating role of TAM in linking system-based quality to behavioral intention and satisfaction in public digital services.

Moreover, by applying Importance-Performance Map Analysis (IPMA), this study goes beyond explanatory modeling and introduces a prescriptive dimension to the evaluation framework. This dual-layered approach reinforces recent theoretical calls in information systems research to integrate both predictive and diagnostic perspectives. The findings underscore the theoretical relevance of usability as not only a technical metric but also a cognitive perception of value, thereby refining existing TAM assumptions. The contextualization in a semi-urban region like Lombok Tengah expands the boundary conditions of TAM and ISO 25010 applicability, suggesting that these models remain robust even in settings with limited infrastructure and digital literacy. This contributes to the growing literature on technology acceptance in developing countries and validates the model's generalizability across diverse socio-technical environments.

9. Limitations and Future Research

This study successfully developed an integrated evaluation model based on ISO 25010, TAM, and IPMA to assess the quality and user acceptance of local government websites. However, several limitations should be acknowledged. First, the model did not account for other external variables such as user trust and social influence, which may also shape satisfaction and behavioral intention. Future research may incorporate constructs such as user trust and social influence, or adopt models like UTAUT and the Expectation Confirmation Model (ECM) to enhance explanatory power. Second, the study was limited to a single region, Lombok Tengah, which may restrict the generalizability of the findings. Third, the data were collected using self-reported survey instruments, which may introduce potential biases such as social desirability effects and common method bias. Although procedural remedies (e.g., ensuring anonymity, randomizing questionnaire items) were applied, and statistical checks (HTMT inference) indicated that CMB was not a major concern, these biases cannot be fully excluded. Future studies are encouraged to triangulate perceptual survey data with system level evidence such as website usage logs, telemetry data, or automated performance monitoring to enhance robustness. In addition, future research could adopt a longitudinal design to observe changes in user perceptions before and after IPMA-driven website improvements. Comparative studies across other semi-urban regions in Indonesia or Southeast Asia would also enhance external validity by accounting for infrastructural and cultural differences. Furthermore, experimental approaches such as A/B testing of priority features (e.g., mobile responsiveness, server speed optimization) could provide causal insights into which interventions most effectively improve satisfaction and behavioral intention. Qualitative methods such as interviews or focus group discussions are also recommended to uncover deeper insights into evolving user behavior, particularly in semi-urban or digitally underserved communities.

These limitations underscore the need for caution in generalizing the findings beyond the study context. Future studies are encouraged to validate this model using larger and more diverse samples, and to further test the robustness of the integrated ISO 25010–TAM–IPMA framework in different public service domains.

10. Conclusion

This study developed and validated an integrated model combining ISO 25010, the Technology Acceptance Model (TAM), and Importance-Performance Map Analysis (IPMA) to evaluate user satisfaction and behavioral intention toward local government websites. Using survey data from 524 users in Lombok Tengah, Indonesia, and analysis via SmartPLS 4.0, the study confirmed that system quality dimensions functionality, usability, reliability, security, and performance efficiency significantly influence perceived usefulness (PU) and perceived ease of use (PEOU), which in turn affect user satisfaction and behavioral intention. Among the findings, usability was the strongest predictor of PU, offering a novel insight into TAM by demonstrating that usability contributes not only to ease of use but also to perceived usefulness. This extends TAM's applicability in the context of public digital services and reflects the critical role of user-centered design in enhancing service quality. The study also highlights the practical utility of IPMA in prioritizing system improvements. Constructs such as PU, PEOU, reliability, and performance efficiency were identified as high-impact yet underperforming dimensions, providing local governments with actionable directions for resource allocation and system enhancement. Theoretically, the integrated ISO 25010–TAM–IPMA model contributes to the literature by bridging system-level quality standards and user-level acceptance factors in a single, comprehensive framework. This is particularly relevant for e-government services in developing regions, where user satisfaction and sustainable engagement depend heavily on both technical performance and perceived usability. The research underscores that technically sound and user-centered digital government platforms are essential for advancing e-government adoption. By aligning technical quality with user expectations, public institutions can enhance digital inclusion and service effectiveness.

References

Al-Adwan, A. S., Al Masaeed, S., Yaseen, H., Balhareth, H., Al-Mu'ani, L., & Pavlíková, M. (2024). Navigating the roadmap to meta-governance adoption. Global Knowledge, Memory and Communication, ahead-of-print. doi: 10.1108/GKMC-02-2024-0105

Alexander, R., Thompson, N., McGill, T., & Murray, D. (2021). The influence of user culture on website usability. International Journal of Human-Computer Studies, 154, 102688. doi: 10.1016/j.ijhcs.2021.102688

Ali, M., Amir, H., & Ahmed, M. (2024). The role of university switching costs, perceived service quality, perceived university image and student satisfaction in shaping student loyalty. Journal of Marketing for Higher Education, 34(1), 201–222. doi: 10.1080/08841241.2021.1975184

Alkhwaldi, A. F., & Al-Ajaleen, R. T. (2022). Toward a Conceptual Model for Citizens' Adoption of Smart Mobile Government Services during the COVID-19 Pandemic in Jordan. Information Sciences Letters, 11(2), 573–579. doi: 10.18576/isl/110225

Al-Qaysi, N., Mohamad-Nordin, N., & Al-Emran, M. (2020). Employing the technology acceptance model in social media: A systematic review. Education and Information Technologies, 25, 4961–5002. doi: 10.1007/s10639-020-10197-1

Al-Sakran, H. O., & Alsudairi, M. A. (2021). Usability and accessibility assessment of Saudi Arabia mobile e-government websites. IEEE Access, 9, 48254–48275. https://doi.org/10.1109/ACCESS.2021.3068917

Bervell, B., Owuyaw, K. E., Armah, J. K., Mireku, D. O., Asante, A., & Somuah, B. A. (2023). Modelling distance education students' satisfaction and continuous use intention of students' portal: An importance-performance map analysis. Information Development, 02666669231206774. doi: 10.1177/02666669231206774

Bitkina, O. V., Kim, H. K., & Park, J. (2020). Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges. International Journal of Industrial Ergonomics, 76, 102932. doi: 10.1016/j.ergon.2020.102932

Bournaris, T. (2020). Evaluation of e-government web portals: The case of agricultural e-government services in Greece. Agronomy, 10(7), 932. doi: 10.3390/agronomy10070932

Candiwan, C., & Wibisono, C. (2021). Analysis of the influence of website quality to customer's loyalty on e-commerce. International Journal of Electronic Commerce Studies, 12(1), 83–102. doi: 10.7903/ijecs.1892

Chan, F. K. Y., Thong, J. Y. L., Brown, S. A., & Venkatesh, V. (2021). Service design and citizen satisfaction with e-government services: a multidimensional perspective. Public Administration Review, 81(5), 874–894. doi: 10.1111/puar.13308

Chawla, D., & Joshi, H. (2021). Importance-performance map analysis to enhance the performance of attitude towards mobile wallet adoption among Indian consumer segments. Aslib Journal of Information Management, 73(6), 946–966. doi: 10.1108/AJIM-03-2021-0085

Cheah, W.-H., Jusoh, N. M., Aung, M. M. T., Ab Ghani, A., & Rebuan, H. M. A. (2023). Mobile technology in medicine: development and validation of an adapted system usability scale (SUS) questionnaire and modified technology acceptance model (TAM) to evaluate user experience and acceptability of a mobile application in MRI safety screening. Indian Journal of Radiology and Imaging, 33(01), 36–45. doi: 10.1055/s-0042-1758198

Chuang, S.-H., Ou, C.-C., & Wang, Y.-X. (2023). Exploring the Influence of Smart Applications And Website Service Quality by Smart PLS Software on Continuous Online Shopping Intention. 2023 IEEE 3rd International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB), 36–40. doi: 10.1109/ICEIB57887.2023.10170377

Davis, F., & Davis, F. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13, 319. doi: 10.2307/249008

Debnath, N., Peralta, M., Salgado, C., Baigorria, L., Riesco, D., Montejano, G., & Mazzi, M. (2021). Digital transformation: A quality model based on ISO 25010 and user experience. EPiC Series in Computing, 75, 11–21.

Dewi, M. R., Ngaliah, N., & Rochimah, S. (2020). Maintainability measurement and evaluation of myits mobile application using iso 25010 quality standard. 2020 International Seminar on Application for Technology of Information and Communication (ISemantic), 530–536. doi: 10.1109/iSemantic50169.2020.9234283

Dianat, I., Adeli, P., Jafarabadi, M. A., & Karimi, M. A. (2019). User-centred web design, usability and user satisfaction: The case of online banking websites in Iran. Applied Ergonomics, 81, 102892. doi: 10.1016/j.apergo.2019.102892

Fadhel, I. E. I., Idrus, S. Z. B. S., Abdullah, M. S. Y., Ibrahim, A. A. E. A., Omar, M., & Khred, A. (2020). A new perspective of web-based systems quality engineering measure by using software engineering theory (ISO 25010): An initial study. Journal of Physics: Conference Series, 1529(2), 022004. doi: 10.1088/1742-6596/1529/2/022004

Fadhel, I. E. I., Idrus, S. Z. B. S., Abdullah, M. S. Y., Ibrahim, A. A. E. A., Omar, M., & Saad, S. (2019). Nias-mukalla web based systems success measurement and students satisfaction evaluation based on security factor of systems quality engineering theory (ISO 25010) and other factors. Independent Journal of Management & Production, 10(6), 2102–2123. doi: 10.14807/ijmp.v10i6.967

Fakfare, P., & Manosuthi, N. (2023). Examining the influential components of tourists' intention to use travel apps: the importance–performance map analysis. Journal of Hospitality and Tourism Insights, 6(3), 1144–1168. doi: 10.1108/JHTI-02-2022-0079

Geoloni, G. D., & Agushinta, D. (2023). Web-Based Human Resource System Quality Analysis Using ISO 25010: 2011 Method Based on Usability Characteristics. Sistemasi: Jurnal Sistem Informasi, 12(2), 573–584. doi: 10.32520/stmsi.v12i2.2791

Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context: A systematic literature review. British Journal of Educational Technology, 50(5), 2572–2593. doi: 10.1111/bjet.12864

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. In European Business Review (Vol. 31, Issue 1, pp. 2–24). Emerald Group Publishing Ltd. doi: 10.1108/EBR-11-2018-0203

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115–135. doi: 10.1007/s11747-014-0403-8

Hidayah, N. A., & Setyaningsih, F. (2019). Combining webqual and importance performance analysis for assessing a government website. 2019 7th International Conference on Cyber and IT Service Management (CITSM), 7, 1–6. doi: 10.1109/CITSM47753.2019.8965408

Ilyas, A., Wajid, S. H., & Muhammad, A. (2022). Usability Evaluation of E-Government Website: A Use of System Usability Scale. Pakistan Journal of Engineering and Technology, 5(1), 11–15. doi: 10.51846/vol5iss1pp11-15

Irawan, B., & Nizar Hidayat, M. (2022). Evaluating local government website using a synthetic website evaluation model. International Journal of Information Science and Management (IJISM), 20(1). doi: 20.1001.1.20088302.2022.20.1.32.5

Jongmans, E., Jeannot, F., Liang, L., & Dampérat, M. (2022). Impact of website visual design on user experience and website evaluation: the sequential mediating roles of usability and pleasure. Journal of Marketing Management, 38(17–18), 2078–2113. doi: 10.1080/0267257X.2022.2085315

Krithika, P., & Vasantha, S. (2024). Evaluating the quality of website design aspects and its effect on functional perspectives using Smart-PLS in Web-based apparel shopping environment. Salud, Ciencia y Tecnología-Serie de Conferencias, 3, 894. doi: 10.56294/sctconf2024894

Lee-Geiller, S., & Lee, T. D. (2019). Using government websites to enhance democratic E-governance: A conceptual model for evaluation. Government Information Quarterly, 36(2), 208–225. doi: 10.1016/j.giq.2019.01.003

Ma, L., & Zheng, Y. (2019). National e-government performance and citizen satisfaction: a multilevel analysis across European countries. International Review of Administrative Sciences, 85(3), 506–526. doi: 10.1177/0020852317703

Martínez-Sala, A.-M., Monserrat-Gauchi, J., & Alemany-Martínez, D. (2020). User Usable Experience: A three-dimensional approach on usability in tourism websites and a model for its evaluation. Tourism Management Perspectives, 33, 100579. doi: 10.1016/j.tmp.2019.100579

Muchran, M., & Ahmar, A. S. (2019). Application of TAM model to the use of information technology. ArXiv Preprint ArXiv:1901.11358 doi: 10.48550/arXiv.1901.11358

Munajat, M. E., & Irawati, I. (2025). Digital Sociocracy: Navigating Governance Challenges in Southeast Asia. Policy & Governance Review, 9(1), 106–124. doi: 10.30589/pgr.v9i1.1220

Nguyen, T. T., Phan, D. M., Le, A. H., & Nguyen, L. T. N. (2020). The determinants of citizens' satisfaction of E-government: an empirical study in Vietnam. The Journal of Asian Finance, Economics and Business, 7(8), 519–531. doi: 10.13106/jafeb.2020.vol7.no8.519

Nishant, R., Srivastava, S. C., & Teo, T. S. H. (2019). Using polynomial modeling to understand service quality in e–government websites. MIS Quarterly, 43(3), 807-A7. doi: 10.25300/MISQ/2019/12349

Nurunnisa, F., Hamzah, M. L., & Saputra, E. (2024). Analyzing the Quality of Web-Based Scholarship Information System Using ISO/IEC 25010 Standard. 2024 International Conference on Circuit, Systems and Communication (ICCSC), 1–6. doi: 10.1109/ICCSC62074.2024.10617354

Paul, S., & Das, S. (2020). Accessibility and usability analysis of Indian e-government websites. Universal Access in the Information Society, 19(4), 949–957. doi: 10.1007/s10209-019-00704-8

Pereira, J., Alturas, B., & Marques, C. (2024). Users' satisfaction evaluation of telemedicine mobile applications based on ISO standards. International Journal of Mobile Communications, 23(3), 376–392. doi: 10.1504/IJMC.2024.137852

Peters, E., & Aggrey, G. K. (2020). An ISO 25010 based quality model for ERP systems. Advances in Science, Technology and Engineering Systems Journal, 5(2), 578–583. doi: 10.25046/aj050272

Quiñones, D., & Rusu, C. (2019). Applying a methodology to develop user eXperience heuristics. Computer Standards & Interfaces, 66, 103345. doi: 10.1016/j.csi.2019.04.004

Rahmawati, R. N. (2019). Self-efficacy and use of e-learning: A theoretical review technology acceptance model (TAM). American Journal of Humanities and Social Sciences Research, 3(5), 41–55.

Salloum, S. A., Alhamad, A. Q. M., Al-Emran, M., Monem, A. A., & Shaalan, K. (2019). Exploring students' acceptance of e-learning through the development of a comprehensive technology acceptance model. IEEE Access, 7, 128445–128462. doi: 10.1109/ACCESS.2019.2939467

Santoso, N. P. L., Sunarjo, R. A., & Fadli, I. S. (2023). Analyzing the factors influencing the success of business incubation programs: A smartpls approach. ADI Journal on Recent Innovation, 5(1), 60–71. doi: 10.34306/ajri.v5i1.985

Sarstedt, M., Hair, J. F., Pick, M., Liengaard, B. D., Radomir, L., & Ringle, C. M. (2022). Progress in partial least squares structural equation modeling use in marketing research in the last decade. Psychology & Marketing, 39(5), 1035–1064. doi: 10.1002/mar.21640

Sarstedt, M., Richter, N. F., Hauff, S., & Ringle, C. M. (2024). Combined importance–performance map analysis (cIPMA) in partial least squares structural equation modeling (PLS–SEM): a SmartPLS 4 tutorial. Journal of Marketing Analytics, 1–15. doi: 10.1057/s41270-024-00325-y

Siregar, K. R., & Rachman, A. A. (2024). Importance-Performance Between Mobile Service Shopping Quality and Zalora Customer Loyalty. Proceedings of the International Conference on Sustainable Collaboration in Business, Technology, Information, and Innovation (SCBTII 2024)., 107. doi: 10.2991/978-94-6463-558-4_7

Sop, S. A., Ongun, U., & Arıcı, H. E. (2025). How does destination natural quality influence second homeowners' life satisfaction: a mediation model and importance-performance map analysis. Current Issues in Tourism, 28(2), 287–303. doi: 10.1080/13683500.2023.2301458

Sun, Y., & Zhang, H. (2021). What motivates people to pay for online sports streaming? An empirical evaluation of the revised technology acceptance model. Frontiers in Psychology, 12, 619314. doi: 10.3389/fpsyg.2021.619314

Vila, T. D., González, E. A., Araújo Vila, N., & Fraiz Brea, J. A. (2021). Indicators of website features in the user experience of e-tourism search and metasearch engines. Journal of Theoretical and Applied Electronic Commerce Research, 16(1), 18–36. doi: 10.4067/S0718-18762021000100103

Vuković, M., Pivac, S., & Kundid, D. (2019). Technology acceptance model for the internet banking acceptance in split. Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy, 10(2), 124–140. doi: 10.2478/bsrj-2019-022

Wattiheluw, F. H., Rochimah, S., Fatichah, C., & Abidin, K. Z. (2020). Development of a Quality Model Based on ISO 25010 Using Fuzzy and PSO for E-commerce Websites. 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 250–254. doi: 10.1109/ECTI-CON49241.2020.9158323.

Wiyono, S. P., & Antonio, F. (2024). Importance Performance Map Analysis on Hospital Branding toward Hospital Choice Intention. JMMR (Jurnal Medicoeticolegal Dan Manajemen Rumah Sakit), 13(3), 315–331. doi: 10.18196/jmmr.v13i3.481

Zaineldeen, S., Hongbo, L., Koffi, A. L., & Hassan, B. M. A. (2020). Technology acceptance model'concepts, contribution, limitation, and adoption in education. Universal Journal of Educational Research, 8(11), 5061–5071. doi: 10.13189/ujer.2020.081106

Journal of Information and Organizational Sciences
Volume 49, Number 2 (2025)
Journal homepage: jios.foi.hr

JIOS

DOI: 10.31341/jios.49.2.9
Open Access

UDC 37.091.33-047.37
Original Scientific Paper

Evaluating The Flipped Classroom Approach in Computer Science Curricula

Ruben Picek[1] and Samuel Picek[2*]

[1]Faculty of Organization and Informatics, University of Zagreb, Varaždin, Croatia
[2]InputLayer GmbH, Zürich, CH
[*]Correspondence: sam@inputlayer.ai

PAPER INFO

Paper history:
Received 14 May 2025
Accepted 18 November 2025

Citation:
Picek, R. and Picek, S. (2025). Evaluating The Flipped Classroom Approach in Computer Science Curricula. In Journal of Information and Organizational Sciences, vol. 49, no. 2, pp. 313-327

ABSTRACT

Active, technology-supported learning accelerated during and after COVID-19, yet evidence from non-programming computer science courses remains limited. This paper contributes (i) a focused review of flipped classroom (FC) studies in CS program (2020-2024) and (ii) a three-year case study of how the flipped classroom enhances the teaching of IT Service Management (ITSM) as a discipline in the computer science program in an online university environment, during and after the COVID-19 pandemic. The FC design combined pre-lecture micro-videos and auto graded quizzes with in-classroom clarification and post classroom activities (project).

Using LMS telemetry, course outcomes, and an end of semester survey across three academic years (2021/22-2023/24), we examined engagement-achievement links with non-parametric, rank based correlations (Spearman ρ), regularized logistic regression, and comparisons across empirically defined engagement tertiles. Results show consistent, practically meaningful associations between quality weighted engagement (quiz participation and performance) and both passing and final grades, with survey perceptions aligning to the behavioral signals. While strictly non-causal, the pattern is robust across methods and suggests actionable uses: early identification of at-risk students and design guidance that emphasizes short, well scaffolded videos and steady formative assessment.

Keywords: flipped classroom, flipped learning case study, computer science education, IT Service Management (ITSM), learning analytics, engagement, logistic regression, experience and achievements using flipped classroom

1. Introduction

Digital technologies and the COVID-19 period accelerated the adoption of online and blended teaching models in higher education. Approaches such as flipped learning, problem based learning, and project based learning are now widely used, supported by Learning Management Systems (LMS) that provide continuous access to resources and traceable engagement data.

The flipped classroom is an innovative teaching and learning strategy where students first engage with materials (e.g., short videos and quizzes) before class and then use class time for guided practice and discussion. Prior studies in computing report that this model is associated with greater student activity and opportunities for interaction, while also noting variability across courses, teachers, and implementations (e.g., O'Flaherty & Phillips, 2015; Long et al., 2017; Sosa Díaz et al., 2021; Gong et al., 2023). Implementations range from fully flipped courses to selectively flipped units, often relying on 5-10 minute videos, low-stakes pre-class assessments, and collaborative in-class tasks.

Despite growing use in programming and other CS subjects, evidence from non-programming CS courses remains limited. This paper contributes a focused review of flipped-learning studies in computer science (2020-2024) and a three year case study (2021-2023) of an IT Service Management (ITSM) course delivered in online/blended modes. Our analyses rely on observational data; we examine associations rather than causal effects.

We address two questions:

- RQ1: To what extent is engagement in flipped activities associated with achievement in an ITSM course?
- RQ2: What are students' perceptions and experiences of learning CS topics through flipped learning?

2. Methodology

The goal of this research is to provide both a theoretical and practical contribution. First, we conducted a comprehensive literature review to identify and synthesize key factors related to implementing the flipped classroom (FC) in higher education computer science (CS) programs. Second, we carried out a case study applying the FC approach in the IT Service Management (ITSM) course taken in the final year of the vocational study program Information Technology and Business Digitalization at the Faculty of Organization and Informatics, University of Zagreb. The case study spans three academic years: 2021/22, 2022/23, and 2023/24. A flipped learning design was developed for the course, covering both conceptual and assessment components.

2.1. Literature review

We followed the three phase approach of Divjak et al. (2022): P1 - Extraction, P2 - Abstract Analysis, and P3 - Detailed Examination. We searched Scopus, Web of Science Core Collection (WoS), and the IEEE Xplore Digital Library on 31 May 2024.

Time window and fields. To capture developments during and after the COVID-19 period, we limited the search from January 2020 to May 2024. Searches targeted Title, Abstract, and Keywords (using database-specific field tags).

Query terms. The core concept combined the flipped classroom with the CS domain. Representative queries were:

- *Scopus:* `TITLE-ABS-KEY ( flipped AND classroom AND computer AND science ) AND PUBYEAR > 2019 AND LIMIT-TO ( SUBJAREA, "COMP" )`
- *WoS:* `TI=(flipped AND classroom AND computer AND science) OR AB=(flipped AND classroom AND computer AND science) OR AK=(flipped AND classroom AND computer AND science); Timespan: 2020-01-01 to 2024-05-31; Category: Computer Science-Interdisciplinary Applications`
- *IEEE:* `(("Document Title":flipped AND "Document Title":classroom AND "Document Title":computer AND "Document Title":science) OR ("Abstract":flipped AND "Abstract":classroom AND "Abstract":computer AND "Abstract":science) OR ("Author Keywords":flipped AND "Author Keywords":classroom AND "Author Keywords":computer AND "Author Keywords":science)) Timespan:2020-2024`

Eligibility criteria. We included peer reviewed journal articles and conference papers in English that (i) focus on higher-education CS or closely related computing courses, and (ii) empirically examine FC implementations, experiences, outcomes, or design. We excluded editorials, posters, theses, and papers outside higher education or outside CS.

Screening and deduplication. Records from all sources were merged and deduplicated (by DOI/title). Titles/abstracts were screened in P2, and full texts examined in P3, with disagreements resolved by discussion.

Yield. The searches returned 112 papers (Scopus 80; WoS 12; IEEE Xplore 20). After deduplication and Phase-1 screening (abstract review), 36 studies remained (Scopus 33; WoS 1; IEEE Xplore 2). If a reader is interested in obtaining the full list of papers analyzed in this study, they may either replicate the search queries or request the author to provide the corresponding BibTeX files.

2.2. Case study

We implemented the flipped classroom (FC) in IT Service Management (ITSM), a final year course in the undergraduate vocational program *Information Technology and Business Digitalization* at the Faculty of Organization and Informatics, University of Zagreb. The case study includes 3 generations of students in academic years 2021/22-2023/24; N = 147; enrollments: 45, 49, 53 students. Delivery was fully online in 2021/22 (Moodle + BigBlueButton) and blended in 2022/23 and 2023/24.

To achieve learning outcomes, in addition to attending lectures, students in a team of 3 members have to solve weekly assignments (projects) related to the concepts of IT service management, which are based on ITIL best practices and the service lifecycle. After the introductory topics, the classes were organized using the FC approach. The FC activities can be categorized into three segments: pre-lecture, in-classroom, and post-classroom activities.

Pre-lecture activities: Over 10 weeks, students had to prepare in advance for each lecture by watching the prepared video content (public YouTube videos uploaded by IT consultants who conduct ITIL training or courses for ITIL certification preparation) and accessing assessments before the lecture, set on the Learning Management System (Moodle). The course follows a model of continuous monitoring/assessment (applicable only to the exam period for continuous assessment), according to which a student can earn up to 10% of the total points for participating in the FC activities. The use of FC was not mandatory, and with each assessment students could earn up to 1 point (10 points in total). Students who did not participate in FC at all did not receive any points from this category during the continuous assessment exam period. All other exam periods followed the traditional oral and written format. The video materials used for FC were also available alongside the exam materials, so students were able to watch them, but of course now without the possibility of completing the assessments.

Each assessment was based on 3 questions (the database for each assessment contained 6 questions), in which the student had to show that he or she understood the new concepts to be considered in the lecture that would follow and apply them in team based weekly tasks.

In-classroom activities: Immediately before the lecture, the teacher analyzed the success of solving each (of the 6) questions and paid special attention to those parts of the material that had the lowest percentage of accuracy.

Post-classroom activities: After each lecture, students solved weekly assignments as a team, which were evaluated at the end of the semester.

In this course, students' knowledge was assessed using multiple techniques and instruments, providing a comprehensive evaluation of both theoretical understanding and practical application. Assessment methods included two written online quizzes (colloquia), which resembled certification tests but were simplified. The final colloquium also included oral questions to further evaluate students' comprehension. Additionally, students, as a team, completed a project in a Virtual IT Company, which involved weekly tasks aligned with the lectures. During the semester, each team presented one completed task according to the schedule. Both the team tasks and presentations were graded, and the project concluded with a team-based oral assessment covering the entire project. Summative evaluation followed the continuous-assessment model: project 40%, two written colloquia 50%, and FC activity assessment points 10%.

For our case study analysis, we derived a set of engagement metrics from FC activity assessment and behavior:

- **Base metrics:** participation rate, number of tests attempted, average test score, total test score, composite tests engagement score (participation and performance), and three consistency measures (overall, stability, early-late).
- **Percentile analogues:** per-metric within group percentiles to enable scale free comparisons.

Achievement was captured by **final grade** (1-5 scale) and **passed** (binary) variables.

We conducted descriptive and inferential analyses, with test selection driven by distributional diagnostics:

1. **Distribution checks.** We assessed normality with the Shapiro-Wilk test. Given pervasive non-normality in outcomes, all bivariate associations were estimated with Spearman's rank correlation (ρ). We checked two-sided p-values and interpreted effect sizes by the magnitude of $|\rho|$ (≈ 0.10 small, ≈ 0.30 moderate, ≥ 0.50 large).
2. **Engagement-achievement correlations.** For each engagement metric (base and percentile variants) paired with each achievement metric, we computed Spearman's ρ with two-sided significance tests and summarized effect sizes via $|\rho|$.
3. **Grouped comparisons by engagement level.** To examine distributional differences in outcomes across engagement strata, we formed Low/Middle/High groups using the 33rd and 67th percentiles

of the composite engagement score within the pooled sample (i.e., ≤33rd, 33rd–67th, >67th). Because outcomes were non-normal, we used Kruskal-Wallis tests with epsilon-squared (ε^2) as effect size (≈0.01 small, 0.06 medium, 0.14 large). We report group n, means/medians/SDs, p-values, ε^2, and lay summaries (e.g., pass-rate differences).

4. **Pass/fail prediction (operational framing).** To assess whether engagement signals separate pass vs. fail cases, we fit an L2-regularized logistic regression with standardized percentile-scaled engagement predictors (to reduce collinearity and unit effects) and a passed variable as the target. We used an 80/20 stratified split (seed = 42) and 5-fold stratified CV. Metrics included accuracy, precision, recall, F1, and AUC-ROC; we also report the cross-validated AUC mean ± SD as a stability check. Coefficients and odds ratios are presented for interpretability (positive coefficients indicate higher odds of passing, holding other features constant).

Other notes on the case study analysis:

- **Multiple comparisons.** We report unadjusted p-values and focus interpretation on effect sizes and consistency across methods (correlations, group tests, classification).
- **Robustness.** Emphasis is placed on rank-based statistics and Spearman's ρ where normality is not supported.
- **Non-causal interpretation.** All analyses are observational and associational; language and conclusions avoid causal attribution.

2.3. Perception survey

At the end of the last lecture of the semester (in all three academic years), and before the evaluation of the project and the final exam, an online survey was conducted among all enrolled undergraduate students to gather their opinions about their experience using the flipped classroom. The instrument comprised 16 statements across four sections, adapted from prior studies to ensure content validity: 7 Likert-type items (1-10), 7 multiple-choice items, and 2 open-ended prompts. Items covered (i) familiarity with modern teaching concepts, (ii) perceptions of flipped classroom (FC) materials and experience, (iii) views on FC assessments, and (iv) self-reported impact and suggestions. Participation was voluntary and approved by the Faculty Ethics Committee.

3. Results

This section reports two complementary strands: (i) a focused review of flipped classroom (FC) studies in computer science programs (2020-2024) and (ii) a three year case study of an online/blended IT Service Management (ITSM) course that adopted an FC approach. We first synthesize the literature to contextualize our setting, and then present quantitative and qualitative evidence from the case study.

3.1. Review of flipped classroom studies in computer science programs

This review synthesizes recent work on flipped classrooms (FC) in university computer science programs from various perspectives. Across studies, a common pattern emerges: short pre-class videos paired with brief quizzes, in-class clarification and active work, and post-class practice or projects (e.g., Aldalur et al., 2022; Gong et al., 2023). Videos are typically concise (≈5-10 minutes), often hosted on YouTube, though length varies by topic. Evidence from a STEM meta-analysis indicates a modest positive effect of FC over traditional formats, especially when pre-class checks, mixed individual/group in-class activities, and post-class quizzes are used (Gong et al., 2023). From a program and instructor perspective, Aldalur et al. highlight those outcomes and workload depend on scope (single course vs. multi-course rollout). Students report higher engagement and continuity, while teachers value richer contact time but face substantial upfront effort, particularly for producing clear, concise videos. Design and technology choices matter: LMS-based quizzes and analytics are common, and studies often blend FC with gamification to sustain motivation (e.g., Algayres et al., 2021; Olivindo et al., 2021). Challenges recur, balancing depth with short videos, supporting large or heterogeneous groups, and maintaining interaction online (Steinmaurer & Gütl, 2023; Ossovski, 2022). Instructor adoption is shaped more by experience and enabling conditions (time, support, tools) than by age or gender (Bakheet & Gravell, 2021b,c).

Most empirical work centers on programming (CS1/CS2) but also spans databases, software engineering/testing, distributed systems, data management, and broader STEM (Aldalur et al., 2022; Araujo et al., 2020; Zamora-Hernandez et al., 2022). Design guidance converges on careful learning design: align FC elements with outcomes, scaffold pre-class work, structure active in-class tasks, and use analytics for timely feedback (Mithun & Luo, 2020). The literature supports FC as a viable approach in CS when thoughtfully designed and institutionally supported, while noting practical constraints, chiefly instructor time, video quality and length, and sustaining student engagement outside class. Table 1 presents the main findings of the literature review.

Category	Findings
Student Impact	FC improves the perception of learning, engagement, and continuity across academic years (Aldalur et al., 2022). Students prefer flexibility: they can pause, rewatch, and come to class prepared (Gong et al., 2023). Quizzes, group and individual activities during class enhance performance (Gallaugher, 2023). Gamification and rewards positively affect engagement (Algayres et al., 2021). Some skepticism remains regarding engagement without direct classroom instruction (Urquiza-Fuentes, 2023)
Teacher Perspective	FC improves the teaching experience but demands significant resources (Aldalur et al., 2022). Creating video content is the most demanding part (Bakheet & Gravell, 2021b). Teaching experience (more than age or gender) influences FC adoption (Bakheet & Gravell, 2021c). Eleven factors were identified that influence FC adoption, including technical self-efficacy and student motivation (Bakheet & Gravell, 2021c).
Pedagogical Design and Course Structure	Successful FC implementation depends on careful instructional design and material adaptation (Aldalur et al., 2022). All studies follow a similar structure: pre-class (video + quiz), in-class (discussion, practice) and post-class (assignments) (Gong et al., 2023). Iterative adaptation is needed based on subject, students, and available resources (Aldalur et al., 2022). Video materials typically last 5–10 minutes (Mithun & Luo, 2020).
Technological and Organizational Aspects	The use of YouTube and other tools is common, but challenges include technical issues and content balancing (Steinmaurer & Gütl, 2023). Asynchronous-synchronous combinations (video + live stream) are well accepted (Aldalur et al., 2022). Face-to-face communication remains irreplaceable (Steinmaurer & Gütl, 2023). In large groups with heterogeneous prior knowledge, FC helps with customization (Ossovski, 2022).
Limitations and Challenges	Difficult content is sometimes better taught in traditional settings than with FC (Elgrably & Oliveira, 2022). Preparing video content and learning materials is time-consuming for teachers (Bakheet & Gravell, 2021a). Some students show lower enjoyment and higher absence rates (Bakheet & Gravell, 2021b). Barriers include gender, language, and time availability (Bakheet & Gravell, 2021b).

Table 1. Main findings of the conducted review.

3.2. Case study - analytic roadmap

Our dataset consists of LMS telemetry data, derived engagement metrics (base and percentile-scaled), and course outcomes, complemented by an end-of-semester survey. We first describe the variables and their empirical distributions, including normality checks and handling of missing data. We then analyze *engagement ↔ achievement* associations using correlation analysis. Next, we examine the predictive strength of those correlations to verify implementation of early interventions. We then assess distributional separation and compare outcomes across engagement tertiles, defined by empirical 32^{nd} and 68^{th} percentiles, using non-parametric tests and effect sizes. Finally, we triangulate the results with survey responses to align student perceptions with observed engagement and achievement patterns. All analysis can be found in the following Github repository: *https://github.com/rubenpicek/fc*.

3.2.1. Variable descriptions and measurements

All assessment variables summarize student activity across ten weekly flipped classroom assessments administered during weeks 4-13. Based on distributional diagnostics and our analytic stance, histograms and Shapiro-Wilk tests indicate non-normality for most variables (notably, participation and attempts are left-skewed; grades are discrete; pass is Bernoulli) (Figure 1a, Figure 1b, Figure 2). Accordingly, we prioritized non-parametric procedures throughout (Spearman's ρ for associations; Kruskal-Wallis for group contrasts).

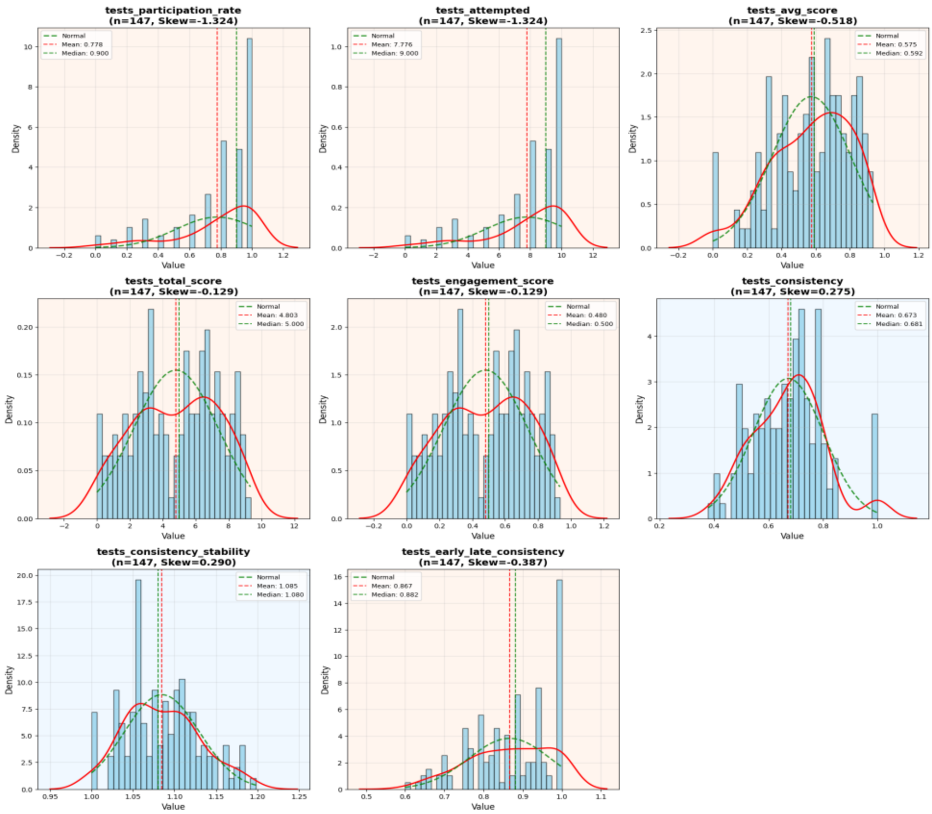

Figure 1a. Histograms of participation, performance and consistency metrics with normality assessment

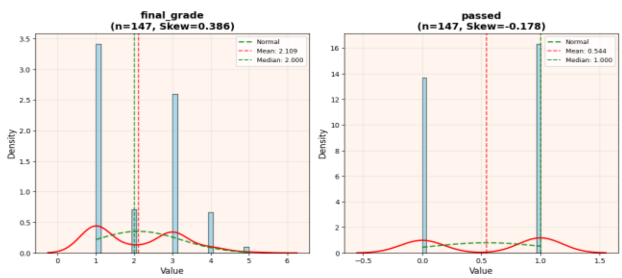

Figure 1b. Histograms of achievement metrics with normality assessment

Together, these variables capture complementary facets of flipped classroom engagement - whether students participated, how well they performed, and how steadily their performance evolved across the ten assessments - alongside two course outcomes (final grade and pass). The histograms and normality tests support our use of robust, rank-based statistics (Spearman's ρ; Kruskal-Wallis) across analyses. These distributional facts also anchor interpretation in the sections that follow (correlation mapping, predictive modeling, tertile comparisons, and survey triangulation), avoiding causal claims and focusing on association strength and practical separability.

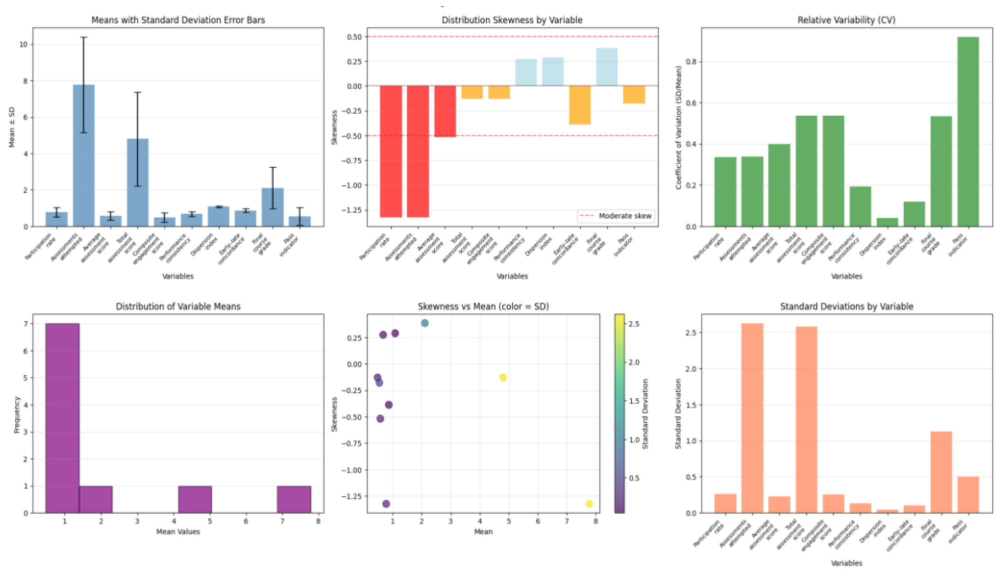

Figure 2. Dataset distribution summary statistics

Key Measurement Insights: Sample size: N = 147. All variables showed non-normal distributions (Shapiro-Wilk), with participation rate and assessments attempted exhibiting strong negative skew (< -1.0), indicating left-heavy engagement patterns. The pass rate was 54.4%, with students attempting an average of 77.8% of the ten assessments. The engagement score (participation × average score) combines frequency and quality of assessment activity, serving as the primary predictor for subsequent analyses.

3.2.2. Correlation analysis

We next examine how each engagement indicator associates with the achievement outcomes using Spearman's rank correlation (ρ). This choice follows the distributional diagnostics: variables are bounded and non-normal, making rank-based measures preferable to parametric alternatives. Spearman's ρ captures monotonic relationships without assuming linearity and is robust to outliers and scale differences. For interpretability, we show ρ in the correlation heatmaps (Figure 3a) and report ρ^2 as an effect-size analogue (Figure 3c) in the rank domain in the companion panels (that is, the proportion of variability in the ranks of the outcome that is monotonically associated with the ranks of the predictor).

Across engagement-outcome pairs, two-sided p-values indicated that core activity/performance metrics are highly significant ($p < .001$). Measures that combine frequency and quality - the total assessment score and the composite engagement index (participation rate × mean score) - show the strongest monotonic relationships with achievement (final grade: $\rho \approx 0.48$–0.49, $\rho^2 \approx 0.24$-0.25; pass/fail: $\rho \approx 0.36$-0.37, $\rho^2 \approx 0.13$-0.14; medium effects). Pure activity indicators (assessments attempted, participation rate) are also significant ($p \leq .001$) but smaller in magnitude (final grade: $\rho \approx 0.40$, $\rho^2 \approx 0.15$; pass/fail: $\rho \approx 0.29$, $\rho^2 \approx 0.08$-0.09). The average score sits between composite and activity-only measures (final grade: $\rho \approx 0.40$, $\rho^2 \approx 0.15$-0.16; pass/fail: $\rho \approx 0.31$, $\rho^2 \approx 0.08$-0.09). Raw temporal regularity features contribute little: raw consistency and consistency stability are almost negligible ($\rho = 0.07$; $p = 0.014$; $\rho^2 = 0.000$ / $\rho^2 = 0.006$), and early-late consistency has a weak relation (final grade: $\rho = 0.18$-0.19, $p = 0.014$, $\rho^2 = 0.047$; pass: $\rho = 0.14$-0.15, $p = 0.01$, $\rho^2 = 0.022$), indicating that once overall participation and mean score are taken into account, the timing pattern of performance (earlier vs. later assessments) adds minimal explanatory value.

Expressing engagement variables as percentiles (Figure 3b and Figure 3d) produces a modest uplift in association strength, with the improvement concentrated in activity indicators such as participation and attempts. The substantive ordering of predictors is unchanged - the composite engagement index and total assessment score remain the strongest correlates of both outcomes. Approximately two-thirds of the 32 engagement-outcome pairs are statistically significant at $p < .05$ under the percentile specification, mirroring the raw-scale results.

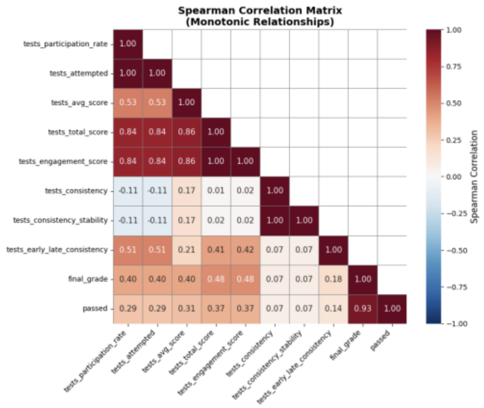

Figure 3a. Correlations of values

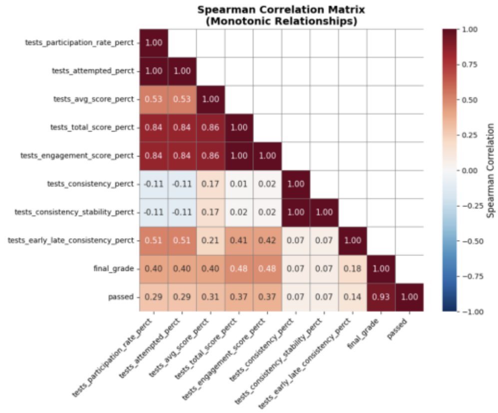

Figure 3b. Correlations of percentiles

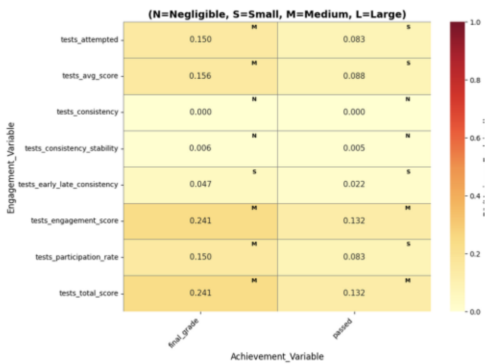

Figure 3c. Effect sizes of values

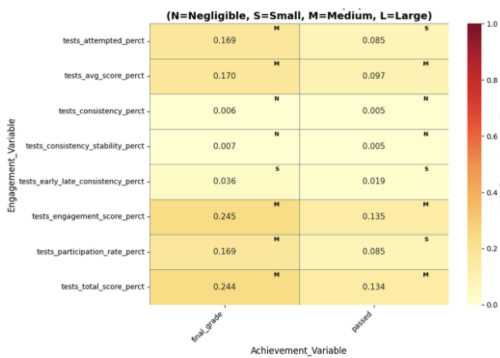

Figure 3d. Effect size of percentiles

3.2.3. Predictive modeling

Having established robust bivariate relationships between engagement signals and outcomes using Spearman's ρ (reported as both correlation coefficients and ρ^2 as rank-domain variance measures), the next step is to examine whether these signals retain unique, conditional value when considered together. Correlations are informative but inherently pairwise; they neither control for the strong interdependence among the engagement variables (e.g., participation, total score, composite engagement) nor indicate how their joint combination translates into actionable predictions. Because one of our outcomes is binary (pass/fail) and most predictors are non-normal, an L2-regularized logistic regression provides an appropriate multivariate framework: it models the log-odds of passing without assuming normality of predictors, quantifies direction and magnitude through interpretable odds ratios, and allows assessment of out-of-sample discrimination (ROC-AUC) for early-warning purposes. Regression enables us to move from "are these variables associated?" to "which signals add independent information, and how well do they work together to identify students at risk," while remaining strictly predictive rather than causal.

We estimated an L2-regularized logistic regression to predict passed boolean value, from the percentile-scaled engagement indicators (participation rate, attempts, average score, total score, composite engagement, and the three regularity measures). The modelling set comprised 147 complete cases (54.4% passed). Using a stratified 80/20 split preserved class balance (train 64/53 pass/fail; test 16/14).

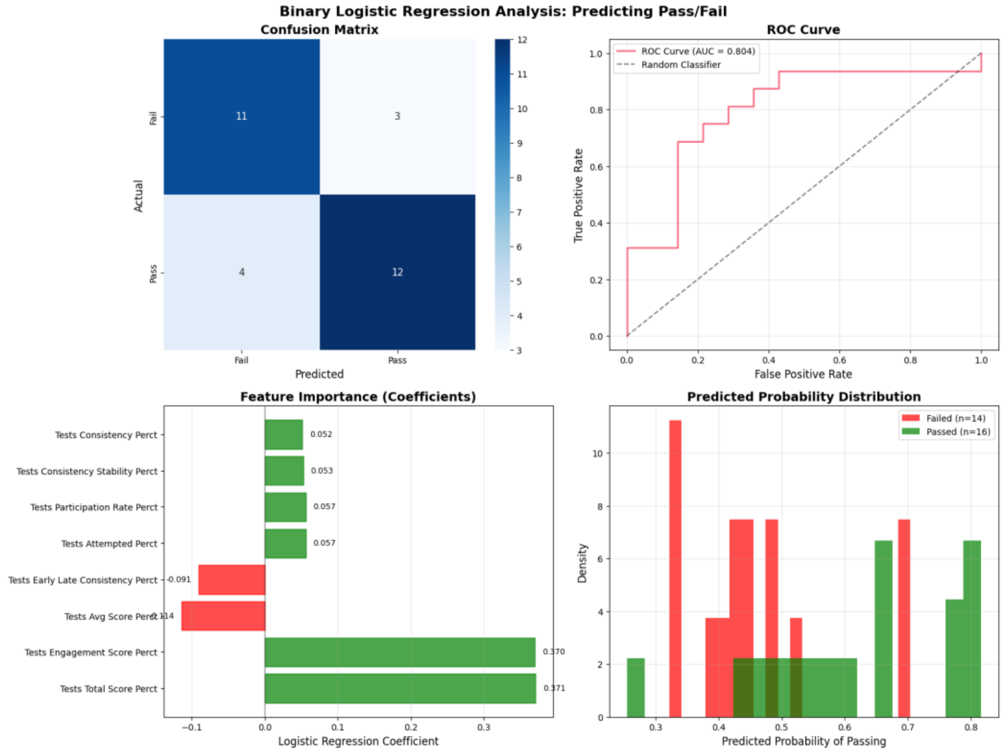

Figure 4. Binary classification model performance evaluation

The feature importance chart (Figure 4) reveals a key insight: quality matters more than quantity. The total assessment score and composite engagement are approximately three to four times more influential than simple participation counts. This suggests that merely completing assessments is insufficient - understanding and mastery, reflected in quiz scores, are crucial. On the held-out fold, the model achieved ROC-AUC ≈ 0.804 with accuracy ≈ 0.77-0.78 (23/30 correct; precision ≈ 0.77, recall ≈ 0.77, F1 ≈ 0.77) (Figure 4).

A stratified 5-fold cross-validation yielded a mean AUC ≈ 0.716, indicating consistent discrimination across folds (Figure 5). Model diagnostics support healthy generalization. Learning curves show the training and validation AUCs converging near 0.73 as sample size grows, with no widening gap - evidence against overfitting. The regularization path is smooth; performance plateaus for C in the 1-10 range, with no signs of dramatic under or over-penalization. Predicted probability distributions exhibit clear separation: most non-pass cases cluster around 0.30-0.50, while pass cases concentrate around 0.60-0.80. The model is slightly conservative (few extreme probabilities), which is desirable for calibrated early warning use.

Coefficient patterns align with the correlation results and provide an interpretable ranking of signals. Two predictors carry the largest positive weights (Figure 4): total assessment score (β ≈ 0.371; odds ratio ≈ 1.45 per percentile unit) and the composite engagement index (β ≈ 0.370; OR ≈ 1.45). Participation intensity remains beneficial once performance is controlled for (attempts and participation rate both have small positive coefficients, β ≈ 0.057). The regularity features contribute modestly: tests_consistency and consistency_stability are positive but small, while early-late consistency is slightly negative. Average score becomes weakly negative after conditioning on total score and composite engagement - an expected suppression effect given strong inter-correlations among activity/performance indicators. Taken together, these results indicate that joint quantity and quality engagement signals dominate predictive value; simple activity counts add only a small increment; timing regularity matters, albeit minimally. The ROC curve, stable learning curves, balanced confusion matrix, and coherent feature weights all point to a well-performing, well-calibrated model that is useful for risk identification (Figures 4 and 5).

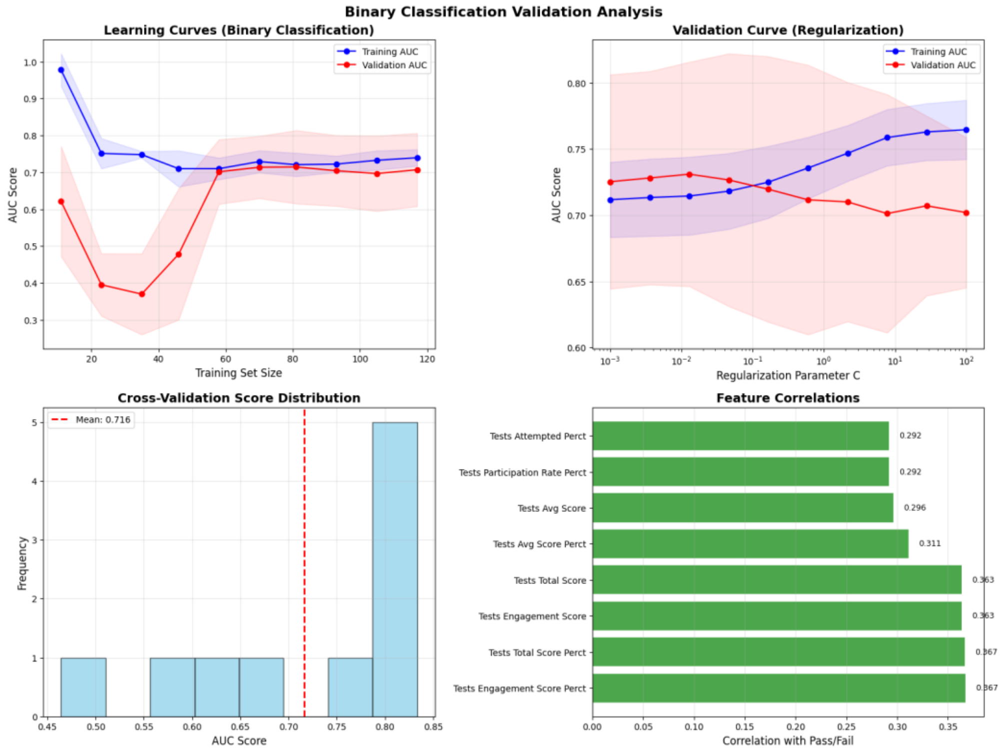

Figure 5. Model validation and generalization assessment

The model's accuracy (ROC-AUC ≈ 0.80) indicates that it correctly ranks students by risk approximately 80% of the time. If one randomly selects a passing and a failing student, the model would correctly identify which is which four out of five times. This performance is strong enough for practical use: flagging the bottom 20-30% of students for proactive outreach would capture most at-risk cases while keeping the intervention manageable.

3.2.4. Tertile comparison

To translate the earlier associations into a form that is directly interpretable for teaching practice, we segmented students into three engagement levels and compared outcomes across these groups. We used the composite engagement score and created approximate tertiles with empirical cut points at the 32nd and 68th percentiles (0.333 and 0.634) to avoid boundary ties. This yielded Low (n = 51), Mid (n = 49), and High (n = 47) groups with mean engagement scores of 0.190, 0.502, and 0.772, respectively.

Because all focal variables are non-normal, we tested group differences with Kruskal-Wallis and reported epsilon-squared (ε^2) as effect size (≈.01 small, ≈.06 medium, ≥ .14 large). The goal here is descriptive: to assess how strongly achievement varies across pragmatic engagement bands.

The tertile split produces clearly separated strata (manipulation check: $H = 129.78$, $p < .001$; $\varepsilon^2 = .887$). Achievement varies monotonically with engagement. Final grades increase from 1.65 (Low) to 1.94 (Mid) to 2.83 (High) (Figure 6); the overall difference is highly significant ($H = 27.43$, $p < .001$) with a large effect ($\varepsilon^2 = .179$). Pass rates show the same pattern - 39.2% (Low), 48.9% (Mid), 78.7% (High) - again significant ($H = 16.42$, $p < .001$) with a medium effect ($\varepsilon^2 = .102$). Process indicators confirm the segmentation: participation rate rises from 0.531 to 0.843 to 0.977 ($H = 86.71$, $p < .001$; $\varepsilon^2 = .588$), and average quiz score from 0.371 to 0.602 to 0.791 ($H = 90.55$, $p < .001$; $\varepsilon^2 = .628$).

In practical terms, students in the top engagement band earn, on average, 1.18 more grade points (on a 1-5 scale) and are about 39.5 percentage points more likely to pass than those in the bottom band. These differences are substantial and consistent with the correlation and regression findings, while remaining strictly associational.

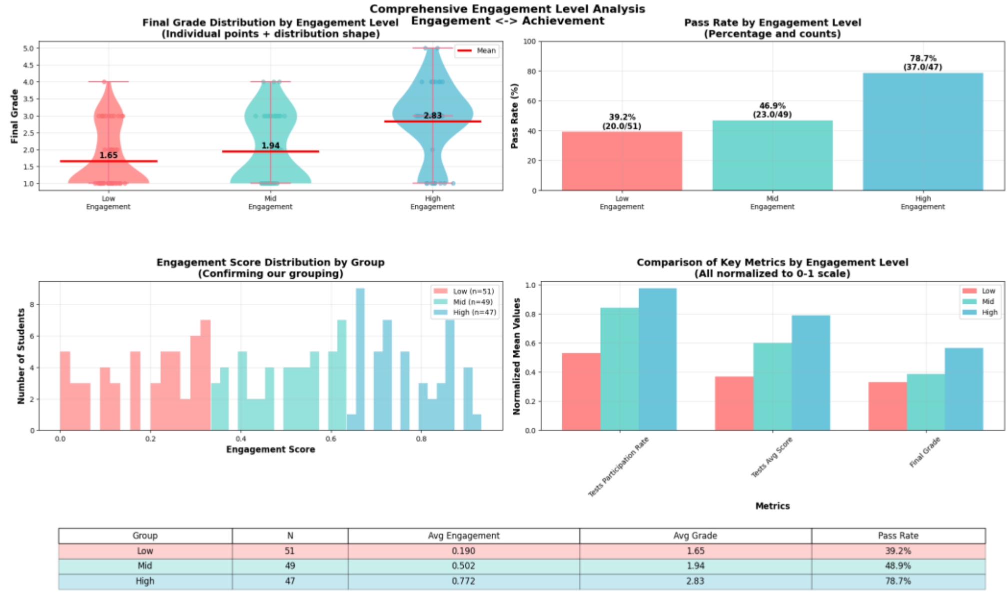

Figure 6. Engagement impact on academic performance

3.2.5. Survey triangulation

To triangulate findings across analytical methods, we administered a post-course perception survey (N=147) comprising 14 substantive items (excluding self-reported impact and suggestion). The instrument captured: familiarity with FC pedagogy, perceived learning benefits and materials quality, self-reported engagement behaviors, assessment experiences, and overall evaluation. Questions used mixed formats (Likert 1-10, categorical, count, percentage) administered in Croatian. Repository (https://github.com/rubenpicek/fc) provides complete definitions and survey questions. Two items (materials_quality, materials_relevance) formed a validated materials perception scale ($\alpha = 0.845$, $\rho = 0.706$). Twelve individual indicators (num_learning_methods, knowledge_before, knowledge_after, learning_style_pref, quiz_difficulty, video_watching_freq, quiz_timing_pref, quiz_helpfulness, time_spent_learning, quiz_participation_rate, overall_satisfaction, perceived_fc_help) captured distinct FC dimensions.

Triangulation across methods:

Correlation ↔ Survey: Students rating materials favorably reported greater FC impact on grades (perceived_fc_help: $\rho = +0.35$, $p < .001$), mirroring correlation findings where composite engagement (participation × score) showed strongest association with final grade ($\rho \approx 0.48$–0.49). The parallel effect sizes - moderate correlations in both behavioral ($\rho^2 \approx 0.24$) and perceptual domains ($\rho^2 \approx 0.12$) - suggest students accurately perceive the achievement benefit of quality engagement.

Regression ↔ Survey: Quiz_participation_rate dominated satisfaction ratings ($\rho = +0.59$, $p < .001$) over materials perception ($\rho = +0.05$, ns), confirming regression results where total_score and composite_engagement carried largest coefficients ($OR \approx 1.45$), while simple participation added minimal unique variance. Students' satisfaction thus tracks the conditional importance hierarchy revealed by multivariate modeling: what you achieve in assessments matters more than mere attendance.

Tertile ↔ Survey: Students self-reporting higher quiz_participation_rate also reported materials helped task completion ($\rho = 0.57$-$.59$) and exam preparation ($\rho = 0.61$-0.62). This aligns with tertile findings: High-engagement students (participation 0.98, score 0.79) passed at 78.7% versus Low-engagement (participation 0.53, score 0.37) at 39.2%. The magnitude of perceptual differences (0.57-0.62 correlations) mirrors the practical significance of behavioral tertile gaps (39.5 percentage-point pass-rate difference), validating that subjective utility aligns with objective stratification.

Cross-method convergence: All four methods independently identify quality-weighted participation (doing assessments well, not just often) as the dominant signal: correlation magnitude ordering, regression coefficient size, tertile pass-rate gaps, and survey satisfaction drivers converge on the same pattern. Survey perceptions validate behavioral telemetry: students who engaged more consistently both performed better objectively and perceived greater benefit subjectively. This multi-method alignment strengthens confidence in the engagement-achievement relationship while maintaining observational interpretation.

4. Discussion

This study extends flipped classroom (FC) evidence to a non-programming CS course (IT Service Management) and integrates four lenses-distributions, correlations, predictive modeling, and tertile contrasts-triangulated with a perception survey. The findings cohere around a simple message: quality-weighted engagement with the formative FC assessments tracks achievement in a strong and practically meaningful way, while timing regularity adds little. We interpret all links as associational, not causal.

Positioning in prior work. Our focused review (2020-2024) shows that FC research in CS is still dominated by programming courses and converges on a common pattern of pre-class micro-content with short quizzes, in-class clarification, and post-class practice. The present case adopts that pattern in ITSM and reaches conclusions consistent with the review: students value clear, concise pre-class materials; formative, auto-graded checks are workable at scale; and in-class time is reallocated to targeted explanation and application. The analyzed studies also indicate that FC approaches foster motivation and enhance students' acceptance and engagement (Aldalur et al., 2022; Algayres et al., 2021; Olivindo et al., 2021; Ossovski, 2022).

RQ1: Engagement and achievement. Non-parametric associations show that the strongest monotonic relationships with final grade and pass/fail are the total assessment score and a composite engagement index (participation rate $\times$ mean assessment score): for final grade, $\rho \approx 0.48\text{-}0.49$ ($\rho^2 \approx 0.24\text{-}0.25$); for pass/fail, $\rho \approx 0.37\text{-}0.38$ ($\rho^2 \approx 0.13\text{-}0.14$). Pure activity measures (attempted/participation) are smaller but clearly significant (e.g., final grade $\rho \approx 0.40$). Regularity features (consistency, stability, early-late balance) contribute only marginally once overall participation and average score are considered. Percentile scaling yields a small average uplift in ρ^2 without changing the ordering of predictors.

A regularized logistic model trained on percentile-scaled engagement indicators achieves test AUC $\approx$ 0.804 with ~0.77 accuracy and cross-validated AUC $\approx$ 0.716, with stable learning curves and a smooth regularization path - evidence of a healthy, well-calibrated classifier. The largest positive coefficients are again total score and composite engagement (odds ratios $\approx$ 1.45 per standardized unit); participation intensity adds a modest positive increment; regularity features are small; average score turns slightly negative when total/composite are included, a suppression pattern consistent with their inter-correlations. Predicted probabilities separate pass/fail cases sensibly (fail $\approx$ 0.30-0.50; pass $\approx$ 0.60-0.80) and remain conservative rather than overconfident.

To translate these signals into practice, tertile comparisons using empirical 32nd/68th cut points show significant monotonic gaps: final grade rises from 1.65 (Low) to 1.94 (Mid) to 2.83 (High) (Kruskal-Wallis $p < .001$; $\varepsilon^2 \approx 0.18$), and pass rate from 39.2% to 48.9% to 78.7% ($p < .001$; $\varepsilon^2 \approx 0.10$). Process indicators confirm separation (participation 0.53 → 0.84 → 0.98; average assessment score 0.37 → 0.60 → 0.79). Practically, a student in the top engagement band earns about +1.18 grade points (on a 1-5 scale) and is approximately 40 percentage points more likely to pass than a student in the bottom band. Analyzed studies also confirm a general positive influence: FC improves students' final grades and overall performance (Bakheet & Gravell, 2021a), (Aldalur et al., 2022).

Practical use. These results support a simple, actionable workflow: compute a composite engagement index each week from LMS data; flag students in the bottom tertile or with predicted pass probability below ~0.5 for a light-touch intervention (brief check-in, reminder to complete the next quiz, or targeted pointers to the specific video/quiz they missed). Because a one-unit increase (standardized - z-score) in total score or composite engagement raises the modeled odds of passing by ~45% (OR = e^0.371 $\approx$ **1.45, Figure 4**), small, achievable increments-finishing one more assessment or raising an assessment average on the next attempt-can yield meaningful gains. At the course level, teachers can (i) allocate feedback and tutorial time first to low-engagement students, (ii) keep micro-videos tightly aligned with the quiz that follows, and (iii) consider modest incentives (time allowances, brief feedback, limited re-attempts) that encourage steady participation rather than last-minute bursts. Several suggestions for effective FC design are also supported by previous work (Bakheet & Gravell, 2021a), (Sibia & Liut, 2022).

RQ2 - Student perceptions. Survey results align closely with the telemetry and outcomes. Students report that FC increases interest in weekly topics and that the videos materially help them follow lectures and prepare for exams; team-based follow-ups are also viewed as useful. This aligns with research showing that the FC approach improved students' perception of their learning, increased engagement, and ensured valuable continuity across academic years (Aldalur et al., 2022; Gong et al., 2023; Gallaugher, 2023). Behaviors are imperfect-many students skim or do not finish videos-and fewer than 10% consistently watch all materials before attempting the quiz. Still, students who report fuller viewing and more frequent assessment access tend to pass more often and earn higher grades, mirroring the quantitative patterns. Satisfaction with assessment design is high; requests focus on more time, richer feedback, occasional native-language videos, and (limited) multiple attempts rather than on different content.

Implications for design and teaching. The converging evidence suggests several actionable priorities for FC in non-programming CS courses:

- **Weight quality and regularity of formative work.** Total assessment score and the composite engagement index carry the most information; simple counts help but are insufficient.
- **Use engagement signals for early support.** The logistic model's calibrated probabilities (AUC ≈ 0.80) and the tertile gaps provide workable thresholds for nudging and tutorial allocation-strictly for prediction, not attribution.
- **Tune pre-class materials and assessment flow.** Keep micro-videos concise (Mithun & Luo, 2020); align each quiz tightly with its video; consider allowing limited re-attempts and providing short, descriptive feedback. Requests for more time and occasional native-language content are reasonable enhancements.
- **Communicate expectations.** Some students still attempt assessments without finishing videos; clearer guidance and light-touch requirements (or small grade weight) can encourage better sequencing.

Limitations. The study is observational, single-site, and focused on a single course; unmeasured factors (e.g. prior knowledge, motivation) may influence both engagement and outcomes. Grades are discrete and pass/fail is coarse; engagement is operationalized through the ten FC assessments and may not capture all study behaviors. Survey data are self-reported and subject to recall and desirability biases. Results speak to associations in this context, rather than causal effects.

Future work. To move from association to credible causal claims about whether flipped learning outperforms traditional formats, the next phase should be a multi-institution, multi-course study-especially in non-programming CS subjects-with sufficient power for heterogeneity analysis. Designs could include cluster randomization or stepped-wedge rollouts at the section or instructor level, complemented by quasi-experimental strategies (propensity scoring, difference-in-differences, synthetic controls) where randomization is infeasible. A pre-registered causal analysis plan should integrate Bradford Hill considerations-temporality (engagement precedes outcomes), consistency (replication across groups and sites), dose-response (greater sustained engagement aligns with stronger effects), plausibility and coherence (linking mechanisms to learning theory), experiment (randomized or natural experiments), specificity (outcomes most affected are those targeted by FC), and analogy. Richer telemetry (fine-grained video interaction, forum discourse, attendance, project checkpoints) and background covariates (prior GPA, programming experience, workload) would help address confounding and support sensitivity analyses. Parallel work should assess calibration, fairness, and cost-effectiveness of early-warning models and examine which FC design elements (video length, feedback timing, limited retakes) drive the largest marginal gains.

If such evidence shows robust, generalizable benefits, institutions can act at scale: embed FC as a program-level design (rather than course-by-course), fund content production and analytics infrastructure, provide faculty development on learning design and data use, and formalize early-support protocols triggered by engagement signals. If effects are context-dependent, policy can target FC to courses and student segments where it is most effective, adjust assessment weightings (e.g., small stakes, frequent checks), and iterate on specific design levers (chunking, feedback, team tasks) that show causal lift. Either way, a larger, causally informed evidence base would enable a defensible curriculum strategy rather than ad-hoc adoption.

5. Conclusion

This study set out to clarify what flipped learning means for computer science education beyond programming courses and to share practical lessons from introducing it in an IT Service Management course. The literature points to clear benefits when pre-class preparation is paired with active class time; our case study shows that the same pattern is workable and well-received in ITSM. Students generally valued the videos, pre-class quizzes, and team tasks, and they described the format as helpful for following lectures and preparing for assessments.

The broad takeaway for teaching practice is simple: steady, quality-focused engagement on frequent assessments across the semester tracks with better achievement. Teachers can act on this by (1) setting clear weekly expectations, (2) using short, well-scaffolded videos, (3) keeping pre-class quizzes low-stakes but predictable, and (4) using LMS analytics to spot and support students who fall behind early. Students consistently asked for a bit more time on quizzes, clearer feedback, occasional native-language materials, and-in moderation-retakes; these are concrete, low-cost adjustments that align with the spirit of formative assessment. For course teams, the design workload is front-loaded: planning, producing, and curating materials matters as much as what happens in the classroom.

At the institutional level, scaling flipped learning responsibly means supporting content production, offering staff development on learning design and analytics, and setting light-touch policies for continuous assessment and early support. Data governance and simple dashboards that highlight weekly engagement can make interventions timely without being intrusive.

Our findings do not claim causality, but they do offer practical guidance: when flipped learning is thoughtfully designed and consistently supported, students engage more steadily-and that is the kind of behavior teachers and institutions can nurture.

References

Aldalur, I., Markiegi, U., Iturbe, M., Roman, I., & Illarramendi, M. (2024). An experience in the implementation of the flipped classroom instructional model in the computer science degree. Engineering Reports, 6(4). Scopus. https://doi.org/10.1002/eng2.12754

Aldalur, I., Markiegi, U., Valencia, X., Cuenca, J., & Illarramendi, M. (2022). E-Learning Experience with Flipped Classroom Quizzes Using Kahoot, Moodle and Google Forms: A Comparative Study. ACM Int. Conf. Proc. Ser., 82-89. Scopus. https://doi.org/10.1145/3572549.3572563

Algayres, M., Triantafyllou, E., Werthmann, L., Zotou, M., Efthimios, T., Malliarakis, C., Dermentzi, E., Lopez, R., Jatten, E., & Tarabanis, K. (2021). Collaborative Game Design for Learning: The Challenges of Adaptive Game-Based Learning for the Flipped Classroom. In Brooks A., Brooks E.I., & Jonathan D. (Eds.), Lect. Notes Inst. Comput. Sci. Soc. Informatics Telecommun. Eng.: Vol. 367 LNICST (pp. 228-242). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-030-73426-8_13

Araujo, P., Costa, C., Viana, W., De Lima Veras, N., & Farias, E. J. P. (2020). Automatic Personalisation of Study Guides in Flipped Classroom: A Case Study in a Distributed Systems Course. Proc. Front. Educ. Conf. FIE, 2020-October. Scopus. https://doi.org/10.1109/FIE44824.2020.9274186

Bakheet, E. M., & Gravell, A. M. (2021a). Investigating computer science instructors behavioral intention to adopt the flipped classroom applying an extended utaut model: The role of age, gender, and experience. International Journal of Information and Education Technology, 11(12), 631-637. Scopus. https://doi.org/10.18178/IJIET.2021.11.12.1574

Bakheet, E. M., & Gravell, A. M. (2021b). Significant factors influencing computer science instructor's behavioral intentions to adopt the flipped classroom: A global quantitative study. Int. Conf. Educ. Inf. Technol., ICEIT, 150-155. Scopus. https://doi.org/10.1109/ICEIT51700.2021.9375590

Bakheet, E. M., & Gravell, A. M. (2021c). Would Flipped Classroom be My Approach in Teaching Computing Courses: Literature Review. Int. Conf. Inf. Educ. Technol., ICIET, 166-170. Scopus. https://doi.org/10.1109/ICIET51873.2021.9419631

Divjak, B., Rienties, B., Iniesto, F., Vondra, P., & Žižak, M. (2022). Flipped classrooms in higher education during the COVID-19 pandemic: Findings and future research recommendations. International Journal of Educational Technology in Higher Education, 19(1), 9. https://doi.org/10.1186/s41239-021-00316-4

Elgrably, I. S., & Ronaldo Bezerra Oliveira, S. (2022). Using flipped classroom to promote active learning and engagement in a Software Testing subject remotely during the COVID-19 pandemic. Proc. Front. Educ. Conf. FIE, 2022-October. Scopus. https://doi.org/10.1109/FIE56618.2022.9962379

Gallaugher, J. (2023). Physical Computing: Empowering Students with Hardware Programming. Annu. Am. Conf. Inf. Syst., AMCIS. 29th Annual Americas Conference on Information Systems, AMCIS 2023. Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85192898139&partnerID=40&md5=90f6d030fbd09d305eed6448bcaee293

Gong, J., Cai, S., & Cheng, M. (2023). Exploring the Effectiveness of Flipped Classroom on STEM Student Achievement: A Meta-analysis. Technology, Knowledge and Learning. Scopus. https://doi.org/10.1007/s10758-023-09700-7

Hannaoui, M., Janous, Y. E., El-Hassouny, E. H., El Hachhach, J., Askam, A., & Habibi, I. (2023). The Impact of the Online Flipped Classroom on the Learning Outcomes and Motivation of Nursing Students in Computer Science. In Lazaar M., En-Naimi E.M., Zouhair A., Al Achhab M., & Mahboub O. (Eds.), Lect. Notes Networks Syst.: Vol. 625 LNNS (pp. 138-147). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-031-28387-1_13

J. Urquiza-Fuentes. (2023). A Study About the Impact of Rewards on Student's Engagement with the Flipped Classroom Methodology. 2023 IEEE Global Engineering Education Conference (EDUCON), 1-6. https://doi.org/10.1109/EDUCON54358.2023.10125267

Long, T., Cummins, J., & Waugh, M. (2017). Use of the flipped classroom instructional model in higher education: Teachers' perspectives. Journal of Computing in Higher Education, 29(2), 179-200. https://doi.org/10.1007/s12528-016-9119-8

Mithun, S., & Luo, X. (2020). Design and Evaluate the Factors for Flipped Classrooms for Data Management Courses. Proc. Front. Educ. Conf. FIE, 2020-October. Scopus. https://doi.org/10.1109/FIE44824.2020.9274201

Nouri, J. (2016). The flipped classroom: For active, effective and increased learning - especially for low achievers. International Journal of Educational Technology in Higher Education, 13(1), 33. https://doi.org/10.1186/s41239-016-0032-z

O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. The Internet and Higher Education, 25, 85-95. https://doi.org/10.1016/j.iheduc.2015.02.002

Olivindo, M., Veras, N., Viana, W., Cortés, M., & Rocha, L. (2021). Gamifying Flipped Classes: An Experience Report in Software Engineering Remote Teaching. ACM Int. Conf. Proc. Ser., 143-152. Scopus. https://doi.org/10.1145/3474624.3476971

Ossovski, E. (2022). Digitally Supported Introductory University Teaching in Computer Science Considering Heterogeneous Groups. ICER - Proc. ACM Conf. Int. Comput. Educ. Res., 2, 10-11. Scopus. https://doi.org/10.1145/3501709.3544296

Sibia, N., & Liut, M. (2022). The Positive Effects of using Reflective Prompts in a Database Course. In Aivaloglou E., Fletcher G., & Miedema D. (Eds.), Proc. ACM SIGMOD Int. Workshop Data Syst. Educ.: Bridging Educ. Pract. Educ. Res., DataEd (pp. 32-37). Association for Computing Machinery, Inc; Scopus. https://doi.org/10.1145/3531072.3535323

Sosa Díaz, M. J., Guerra Antequera, J., & Cerezo Pizarro, M. (2021). Flipped Classroom in the Context of Higher Education: Learning, Satisfaction and Interaction. Education Sciences, 11(8), 416. https://doi.org/10.3390/educsci11080416

Steinmaurer, A., & Gütl, C. (2023). Implementation and Experiences of a Flipped Lecture Hall-A Fully Online Introductory Programming Course. In Auer M.E., Pachatz W., & Rüütmann T. (Eds.), Lect. Notes Networks Syst.: Vol. 633 LNNS (pp. 832-843). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-031-26876-2_79

Yu, Y., Fu, Y., Chen, Y., & Li, M. (2023). MOOCs Based Blending Teaching Reform for Integrated College Computer Course. In Hong W. & Weng Y. (Eds.), Commun. Comput. Info. Sci.: Vol. 1811 CCIS (pp. 293-304). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-981-99-2443-1_26

Zamora-Hernandez, I., Rodriguez-Paz, M. X., Gonzalez-Mendivil, J. A., & Palomares-Moctezuma, J. A. (2022). A Teaching Model for a Five-week Electric Circuits Course for Engineering Students. ACM Int. Conf. Proc. Ser., 76-81. Scopus. https://doi.org/10.1145/3572549.3572562

Journal of Information and Organizational Sciences
Volume 49, Number 2 (2025)
Journal homepage: jios.foi.hr

JIOS

DOI: 10.31341/jios.49.2.10
Open Access

UDC 004.8:378:159.9-047.37
Original Scientific Paper

The Impact of Generative AI on University Students' Learning Experience: A Study on Cognitive and Affective Outcomes

Sajeda Ben Otman[1], Sourur Ben Otman[1] and Gökçe Karahan Adalı[2*]

[1]Department of Management Information Systems, Halic University
[2]Department of Software Engineering, İstanbul Sabahattin Zaim University
[*]Correspondence: gokce.karahanadali@izu.edu.tr

PAPER INFO

Paper history:
Received 29 August 2025
Accepted 29 November 2025

Citation:
Otman, S. B., Otman, S. B & Adali, G. K. (2025). The impact of Generative AI on University Students' Learning Experience: A Study on Cognitive and Affective Outcomes. In Journal of Information and Organizational Sciences, vol. 49, no. 2, pp. 329-344

ABSTRACT

Generative Artificial Intelligence (GenAI) is rapidly transforming higher education, yet its impact on learning experiences remains contested. Existing research often isolates either cognitive outcomes (e.g., comprehension, creativity) or affective outcomes (e.g., motivation, engagement), leaving a gap in integrated analyses that also account for heterogeneity across student groups. This study investigates both dimensions simultaneously by examining university students' perceptions of GenAI, focusing on learning, creativity, motivation, and engagement, alongside perceived risks such as overreliance, ethical concerns, and difficulties in verifying accuracy. Data were collected from 93 students and analyzed through Spearman's correlations and unsupervised clustering (k-means) with PCA visualization. Findings indicate low to moderate positive correlations between GenAI usage and learning outcomes, particularly problem-solving and motivation. Cluster analysis reveals diverse usage–perception profiles, including paradoxical cases where frequent users report limited cognitive benefit. These results align with Technology Acceptance Model (TAM) and UTAUT assumptions of perceived usefulness and performance expectancy, while also showing that digital literacy moderates these relationships, especially in critical thinking and responsible use. The study contributes by integrating cognitive and affective outcomes, revealing latent profiles beyond averages, and bridging adoption models with responsible AI frameworks. Practical implications highlight the need for AI literacy training, ethical policies, and instructional design to foster effective and responsible GenAI integration in higher education.

Keywords: Generative Artificial Intelligence, Higher Education, Learning Experience, Technology Acceptance Model, Responsible AI

1. Introduction

In recent years, higher education witnessed huge transformations powered by technological advancements. GenAI comes to the fore with smart tools capable of producing new content such as text, descriptions, and codes, based on the prompts it receives. ChatGPT, Bard, Copilot, and other tools are now an educational reality as university students increasingly adopt such tools to support their self-learning, which includes aiding in academic tasks, concept understanding, idea creation, and much more. However, while these tools seem to promise endless opportunities, the actual impact on the learning experience remains debated. Our study aims to analyze the impact of GenAI tools on the university learning experience in two dimensions, which are the cognitive outcomes, including learning and creativity, and the affective aspects, like motivation and engagement. Moreover, we aim to learn about the students' perceptions of the potential benefits and risks aligning with GenAI. All of this study aims to provide a comprehensive view that could contribute to guiding the responsible use of this technology.

2. Literature review

GenAI is a unique type of artificial intelligence focused on producing new content. It does this by leveraging existing patterns of data to create a statistical model predicting what response it would give based on the prompts it receives. GenAI has gone through the developments that have taken place over the decades, beginning in the mid-20th century when GenAI was very basic, depending on the rules to produce language, and continuing through the late 20th century when we witnessed a massive leap forward with the emergence of neural networks that have made voice recognition and imaging tech as we know it today possible, and finally, what opened up the door for the generation of high-quality text, images, and other content we see today was down to deep learning techniques like Generative adversarial networks (GANs) and variational Autoencoders (VAEs). (Yu and Guo, 2023).

According to Albadarin et al. (2024), ChatGPT serves as a virtual intelligent assistant in its ability to offer rapid responses, immediate feedback, and easy-to-understand explanations.

AI tools have also been utilized in writing and language tasks, assisting in essay writing, paraphrasing, translating text, and grammar checking. In addition, ChatGPT enhances self-directed and personalized learning by helping students understand concepts, clarify assignments, and organize learning plans based on individual needs. (Albadarin et al., 2024).

While this can be seen as an educational opportunity, there are significant risks in using AI-based tools for academic work. The reliability and accuracy of these tools are significant issues, for instance. Students need the requisite skills and competencies to be able to use them effectively and to judge the quality of the generated responses (Albadarin et al., 2024). Albadarin et al. (2024) emphasize that setting clear guidelines regarding the potential risks of AI, such as misinformation, plagiarism, and unequal access to AI tools, is crucial.

2.1. Cognitive Impact of Generative AI

Generative AI tools such as ChatGPT, Google Bard, and Bing AI promote new approaches in educational environments to a better student learning experience (Schei et al., 2024).

One of the cognitive advantages of these tools is that they can be really useful for students in understanding complex concepts. According to Albadarin et al. (2024), ChatGPT is used in idea generation, text translation, and alternative explanations that reinforce students' academic understanding across various disciplines. Bahroun et al.'s (2023) review also supports that AI chatbots assist in breaking down complex topics. Moreover, AI-based tools have also demonstrated improvement in learning efficiency by providing well-organized and structured responses (Albadarin et al., 2024). In the study examining papers on students' perceptions of AI tools, it was identified that AI tools decrease the cognitive load on learners, making the learning process less time-intensive (Schei et al., 2024). Additionally, AI tools can be significant in shaping students' critical thinking skills. A few studies claim that they enhance critical thinking and reasoning abilities by informing students in an interactive process to analyze and interpret responses (Albadarin et al., 2024). In the study investigating papers about students' perceptions and usage of AI chatbots, it was determined that ChatGPT encourages students to consider alternative perspectives and critically evaluate arguments. The study indicates that AI tools such as ChatGPT can improve creative problem-solving, as well (Schei et al., 2024). One study, for instance, found that students who used ChatGPT wrote more elaborate works than students who didn't. Others argue that AI poses a significant risk to reducing deep thinking and independent creativity (Schei et al., 2024). One study found that students worried that ChatGPT gives concise answers and that there is no opportunity to think independently (Schei et al., 2024). Despite this concern, research has shown AI could be a "collaborative creative agent" rather than a substitute for human creativity (Vinchon et al, as cited in Habib et al., 2024). They call this the era of "assisted creativity," in which AI helps and enhances the human creative processes. In the study examining how GenAI affects creativity, undergraduate students performed a creative task with and without the assistance of ChatGPT-3. The results showed that AI enhanced fluency, flexibility, elaboration, and originality in idea generation. But some students suggested that AI made them feel like they were becoming dependent and less confident in their ability to come up with original ideas (Habib et al., 2024). Similarly, Schei et al. (2024) highlighted the potential risk of declining cognitive development and academic depth as students may prefer being given quick answers instead of a deep search.

2.2. Affective Impact of Generative AI

Motivation and engagement are key elements in the learning process as they help learners achieve their learning goals. In an experimental study on computer engineering students, it was revealed that students using AI-supported chatbots in their courses were more motivated than those who did not use chatbots (Neji et al., 2023). In fact, ChatGPT has been shown to dramatically increase student engagement and participation. A review by Schei et al. (2024) found that in general, students find such tools helpful and motivating. A study showed that ChatGPT makes students more motivated to complete reading and writing tasks. Another study showed that ChatGPT increases self-efficacy in academic task completion (Schei et al., 2024). Furthermore, Muñoz et al. (2023) observe that “ChatGPT gave a sense of empowerment and increased engagement” in language learning. They discovered that direct interaction with ChatGPT while exploring a specific concept leads to commendable feedback and aid, which further stimulates a learner's enthusiasm for studying. Similarly, Sandu et al. (2024) found that ChatGPT facilitates student engagement, where they reported it with an engagement score of 3.18 out of 5 through their Australian case study. While some students noted difficulties like a “lack of human interaction”, the authors characterize ChatGPT as providing an “interactive and dynamic learning environment for students”. However, not every study shows a positive influence. Zhu et al. (2024) discussed that non-STEM students were better engaged in scenarios without ChatGPT; therefore, ChatGPT may not facilitate engagement for all students. The study noted that engagement might depend on the instructional design since the students who participated in tasks with a debate style had higher levels of engagement than those in tasks with a fact style. It is noteworthy that AI tools can promote procrastination and surface learning (Schei et al. 2024), which thus can lead to a loss of motivation when, in some cases, students turn into passive learners.

2.3. GenAI challenges and limitations in higher education

The literature also points out to limitations and ethical issues that should be addressed to ensure that AI tools effectively facilitate the learning process while minimizing the risks involved.

- **Academic Integrity and Ethical Concerns:** Research by Schei et al. (2024) argues that plagiarism has become easier and more accessible because chatbots generate output that is both novel and humanlike, which traditional plagiarism detection systems struggle to identify. The employment of these tools could lead to work that is not a student’s own. This raises ethical questions around credit and authorship. If these issues are ignored, they may diminish academic standards, intellectual growth, and disrupt educational process.
- **Accuracy and Reliability of AI Responses:** A study by Zhu et al. (2023) discovered that students usually met with inaccurate responses from AI and had to fact-check that information. Another study mentioned by Mittal et al. 2024 review reported that even advanced models like GPT-4 struggle with accuracy, and that they often approve incorrect code. Thus, AI chatbots are not yet perfect and may misinterpret questions or provide incorrect answers.
- **Limited Contextual Understanding:** AI chatbots can be of benefit, but they still might have issues understanding contexts as we humans do. For instance, Neji et al. (2023) demonstrate that these tools actually rely on patterns and keywords, leading to misinterpretations of questions and superficial answers. Zhu et al. (2023) found that students were dissatisfied with the generic responses provided by ChatGPT, which contained no specific details or insights.
- **Bias in AI-generated Content:** AI models learn from large data sets, and they may retain some biases. A study by Zhu et al. (2023) found that students were worried that ChatGPT might not be accurate. They reported that chat responses might mirror biases in the system and tend to preserve the user’s perception. Given the potential for the creation of misleading content, it is important to be wary of misinformation and to refine AI-generated content through critical thinking.
- **Student Dependency and the Impact on Critical Thinking:** Relying too heavily on AI-written responses can promote surface learning rather than conceptual learning. (Schei et al. 2024). In this experimental study by Zhu et al. (2023), students reported that they allowed the AI to think for them. One student said, “It stops me from thinking." Thus, excessive dependency on AI tools might result in diminished original thinking.

2.3.1. Research Gap and Contribution

Although prior studies document numerous applications of Generative AI in higher education, the literature often remains fragmented. Most investigations focus either on **cognitive outcomes** (e.g., comprehension, problem-solving, creativity) or on **affective outcomes** (e.g., motivation, engagement), but rarely address both perspectives simultaneously. Furthermore, existing work tends to report descriptive findings without exploring **heterogeneity across student groups**. Thus, we still lack an integrated, data-driven perspective on how GenAI shapes both cognitive and affective dimensions of learning and how these dimensions interact across different usage profiles.

Our study addresses this gap by combining correlational analysis with **unsupervised clustering (k-means) and PCA visualization** in a university sample. This approach not only examines direct relationships between GenAI use and learning outcomes, but also reveals distinct clusters of student experiences, including paradoxical profiles such as **high-usage but low cognitive belief**. By capturing such nuances, we offer a comprehensive contribution that extends the existing literature.

2.3.2. Theoretical Framework

To strengthen the interpretation of our findings, we ground the study in established theories of technology adoption and digital literacy.

- **Technology Acceptance Model (TAM).** Davis (1989) proposed that technology acceptance is shaped by **perceived usefulness (PU)** and **perceived ease of use (PEOU).** In our study, students' beliefs that GenAI helps in problem-solving and understanding complex concepts correspond to PU, while their confidence in using GenAI aligns with PEOU. The positive correlations we observe between usage frequency and perceived learning benefits are consistent with TAM's assumptions.

- **Unified Theory of Acceptance and Use of Technology (UTAUT).** According to Venkatesh et al. (2003), adoption is influenced by **performance expectancy, effort expectancy, social influence, and facilitating conditions**. Our results particularly reflect performance and effort expectancy, as students who perceive GenAI as effective tend to show higher motivation and engagement. These insights allow us to situate affective outcomes such as motivation within a well-established theoretical lens.

- **Digital Literacy and Responsible AI Use.** Beyond adoption models, digital literacy frameworks emphasize critical evaluation and responsible use of technology (van Deursen & van Dijk, 2014). In our context, concerns about misinformation, plagiarism, and dependency mirror this perspective. For example, students reporting difficulty in verifying GenAI's outputs illustrate how insufficient digital literacy can weaken the link between technology use and learning gains. Thus, our framework integrates **TAM/UTAUT mechanisms** with digital literacy as a **moderating condition** that explains paradoxical patterns (e.g., frequent use coupled with low cognitive belief).

This theoretical grounding allows our study to contribute not only empirical evidence but also conceptual insights into how GenAI reshapes learning in higher education.

3. Methodology

In this study, the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework is adopted for data analysis. This framework consists of six steps, which are business understanding, data understanding, data preparation, modeling (data analysis), evaluation, and deployment, that aim to guide and facilitate data mining and analysis projects (Shearer, 2000). Based on the study aims, the first four steps are followed to understand the research problem, analyze the data, prepare the collected data for analysis, and perform the statistical analysis.

3.1. Business Understanding

This study aims to evaluate the impact of GenAI tools use on university students' learning experiences, with a focus on variables like comprehension, learning efficiency, critical thinking, creativity, motivation, and engagement. Therefore, a key focus of this study is a comprehensive analysis about GenAI use and its effect on these variables based on demographic characteristics such as gender, age, and department, as well as frequency of use, confidence in using GenAI, and purposes of GenAI use. By determining the impact of GenAI use on both aspects of learning, cognitive and emotional, better-informed decisions can be made about the use of GenAI tools by educators and policymakers.

3.2. Data Understanding

This study is based on survey data obtained from online questionnaires. The questionnaires were shared between 12 March and 12 April 2025, and distributed through online links in WhatsApp, LinkedIn, and face-to-face conversations with students, using a convenience sampling strategy based on participants' availability and willingness to take part in the study. The survey covered a total of 93 participants, including undergraduate and postgraduate students from various disciplines. This survey consists of 21 closed-ended questions, including multiple choices and a 5-point Likert scale to measure agreement from "Strongly Disagree" to "Strongly Agree" as well as frequency scales. Survey topics encompassed of demographics (such as gender, age, department and academic year), GenAI usage, frequency of use, confidence in GenAI, purposes of use, willingness to recommend GenAI for others, perceptions of how GenAI affects problem-solving, learning efficiency, critical thinking, creativity, engagement and motivation as well as challenges associated with GenAI use.

The first variable group is about demographics, the sample has a female proportion of 64.5% whereas male is 35.5%. Distribution based on age is given in a bar chart in four age groups (18-20, 21-23, 24-26, and 27 and above) in **Error! Reference source not found.**. The highest population comes from age group two, which is 21-23 years, and represents 59.1% of the participants.

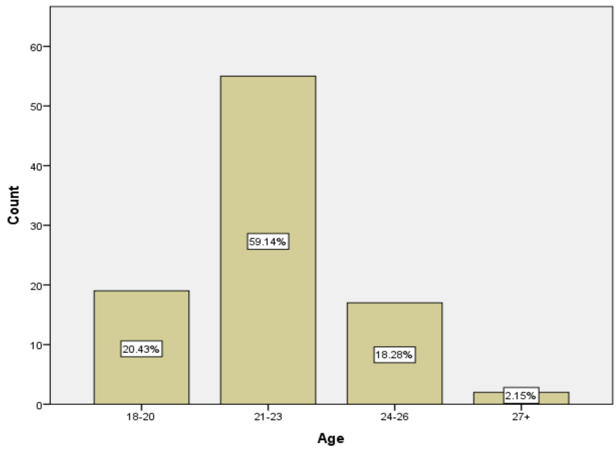

Figure 1. Age distribution

Considering departments, five choices were made (STEM, Social Science and Humanities, Business and Economics, and Arts and Design) along with an "Other" option to specify other majors; however, 20.43% specified their majors in "Other" regardless of that major being listed or not. The highest population is from STEM with 39.78% then Business and Economics with 29.03%.

As for academic year, 15.05% of the respondents are postgraduate students, the biggest population are "fourth year or above" students, and they represent 44.09%, after them, "second year" students represent 20.43% of the total sample.

Other variables were analyzed about GenAI usage, frequency of use, purposes of use, and GenAI associated challenges to measure different aspects to better understand who is affected by GenAI tools in what sense. For "GenAI usage" 95.7% of the participants answered *"Yes"* that they have used GenAI tools for academic purposes, which indicates the popularity of these tools in academia, while 4.3% answered *"No"* they never used them.

Considering "Frequency of use", it is clear that most of the students use them regularly as 22.58% stated that they use them "*Daily*" and 44.09% selected "*A few times a week*" whereas the less frequent usage was those 16.13% who use it *"A few times a month*" along with the 12.9% and 4.3% who chose *"Rarely*" and "*Never*" respectively. (**Error! Reference source not found.**).

Next, in order to analyze how students are impacted by GenAI tools, it was vital to observe what tasks they use them for. Participants were asked to rate the frequency of use for various academic tasks on a 5-point Likert scale from "*Never*" to "*Always*". According to the frequency average (based on the Likert scale), "*Research and Study*" is the most common use of these tools as more than half of the sample (n = 51) selected the level 4 or 5, followed by "*Writing essays/reports", "Generating ideas*", and *"Solving problems"* respectively. The average usage frequency for each purpose is presented in.

When participants were asked about how confident they are to use GenAI for their academic work, generally they showed varied confidence levels as 10.8% noted they were "extremely confident", 37.6% were "very confident", 41% were "slightly confident", and 9.7% were "not confident at all".

These results show that most respondents are confident, although the degree to which they trust it varies.

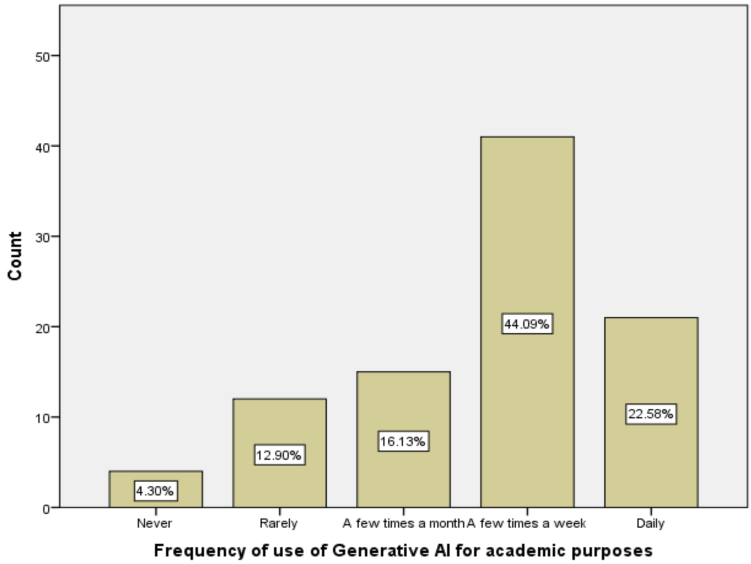

Figure 2. Frequency of use GAI for academic purposes

The second part of the survey questions was regarding the main variables of our research (learning, creativity, motivation, engagement).

We asked participants various questions, such as a Likert-type agreement scale, and some nominal questions (Table 1).

Variable	Question	Values
V1	Generative AI supports my ability to analyze and solve complex problems.	"Strongly Disagree", "Disagree", "Neutral", "Agree", "Strongly Agree"
V2	Compared to traditional learning methods (e.g., textbooks, lectures), how effective do you find Generative AI in understanding complex topics?	"Much less effective than traditional methods", "Slightly less effective", "About the same", "Slightly more effective", "Much more effective"
V3	Do you believe that Generative AI reduces your ability to think critically and independently?	"Strongly Disagree", "Disagree", "Neutral", "Agree", "Strongly Agree"
V4	I feel that Generative AI expands my creative thinking beyond what I would normally generate myself.	"Strongly Disagree", "Disagree", "Neutral", "Agree", "Strongly Agree"
V5	How does generative AI influence your approach to problem-solving?	"Encourages me to explore innovative solutions", "Helps structure my thoughts but does not increase creativity", "Has no impact

Variable	Question	Values
		on my problem-solving process", "Reduces my originality by making me rely on pre-existing ideas", "Other"
V6	How do you think Generative AI affects your creative thinking process?	"Encourages me to develop unique solutions", "Provides structured ideas but does not enhance my creativity"," Limits my creativity by making me depend on existing ideas", "No effect"
V7	Using generative AI increases my motivation to engage with academic content.	"Strongly Disagree", "Disagree", "Neutral", "Agree", "Strongly Agree"
V8	Generative AI encourages me to explore new academic topics beyond what is required.	"Strongly Disagree", "Disagree", "Neutral", "Agree", "Strongly Agree"
V9	Generative AI makes learning more interactive and engaging for me.	"Strongly Disagree", "Disagree", "Neutral", "Agree", "Strongly Agree"
V10	Using generative AI increases my participation in class or group discussions.	"Strongly Disagree", "Disagree", "Neutral", "Agree", "Strongly Agree"
V11	Does the availability of Generative AI reduce your motivation to study independently?	"Strongly Disagree", "Disagree", "Neutral", "Agree", "Strongly Agree"
V12	Would you recommend using Generative AI for studying to other students?	"Strongly Disagree", "Disagree", "Neutral", "Agree", "Strongly Agree"

Table 1. Survey Variables

Our analysis indicates that students' perceptions generally tended to be positive, but with interesting contrasts. For instance, 54.9% of total participants (agree and strongly agree) agreed that GenAI supports their ability to solve complex problems, whereas only 17.2% disagreed with that. Furthermore, 73.1% reported that GenAI was more effective than traditional learning methods in terms of understanding complex problems, indicating a general positive trend among students to view these tools as effective.

Despite this positive attitude, a decent percentage of students reported concerns about the potential negative influence on critical thinking and independent studying. 40.9% noted that GenAI reduced their ability to think critically and independently, which is not a small proportion, indicating the participants' awareness of the potential drawbacks.

Considering creativity, 40.9% agreed on the fact that GenAI expands creative thinking; however, 24.7% expressed their disagreement with that. Notably, when participants were asked how generative AI affects their creative thinking process, 18.3% reported that generative AI limits their creativity by making them depend on existing ideas.

As for motivation and engagement, our data showed that 42% agreed that GenAI increases their motivation, with a decent percentage undecided, 38.7%.

Interestingly, when participants were asked about studying independently, more than one-third (38.7%) of the students expressed that the availability of GenAI tools reduces their motivation to study on their own; however, a similar percentage (34.4%) disagreed with that view. This suggests that students' views of GenAI are diverging.

55.9% of total students revealed that GenAI makes learning more interactive, whereas 17.2% disagreed with that view. It is noteworthy that the impact on class participation was the least, as only 36.5% expressed that it encourages them to participate actively in class.

Notably, 67.7% of total students stated that they are willing to recommend GenAI tools to their peers, reflecting that students generally recognize its benefits and usefulness.

3.2.1. Challenges with Using and Trusting GenAI for Academic Work

When all participants were asked to select which challenges they might face from multiple choices about using GenAI for their academic work, it was found that most students face challenges related to "*Difficulty in verifying the accuracy of information*", with 61.3% (n = 57) of total students opting for it. The second highest frequency was then for "*Ethical concerns*" (n = 34, 36.6%), followed by "*Technical issues*" (n = 28, 30.1%). Other challenges are as follows: "*Over-reliance on AI-generated content*" (n = 25, 26.9%), "*Lack of creativity in AI-generated outputs*" (n = 25, 26.9%), "*Limited understanding of how to use the tools effectively*" (n = 24, 25.8%) (Figure 3).

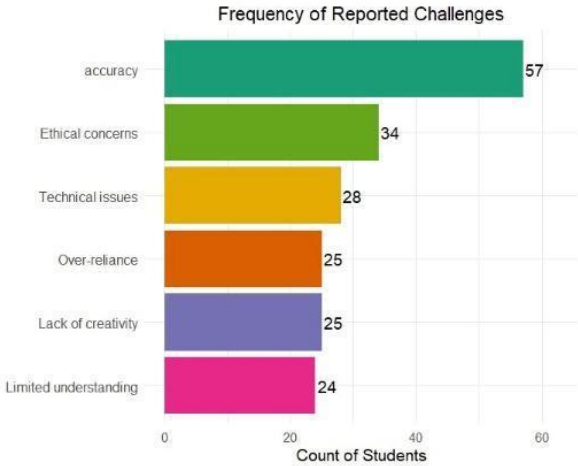

Figure 3. Frequency of Reported Challenges

Thus, our findings show that the major challenges are not merely related to content itself but extend to issues related to ethical and technical aspects. The high percentage of those who have difficulty verifying information points to a serious concern about the accuracy of the generated content, suggesting the importance of developing students' capabilities related to critical evaluation. Also, the high proportion of students mentioning ethical concerns might reflect a growing understanding of the responsible use of these tools. Other challenges that over 25% opted for, like technical issues and the limited understanding of how to use it, might indicate a gap in digital literacy. Overall, Findings reveal that despite the benefits GenAI provides in supporting the learning experience, some actual concerns and challenges that students noted should be addressed to ensure effective and responsible use.

3.3. Data preparation

For the open-ended option "other" in the field of study question, we cleaned the typos and categorized them into the predefined categories. Also, for "other" in the "What challenges have you faced while using generative AI for academic purposes?", there was one "other" specified; we removed it due to its irrelevance. All the responses were coded accordingly, and Statistical Package for the Social Sciences (SPSS) was used to prepare and analyze the dataset.

3.4. Data Analysis

In this study, Spearman's rank correlation coefficient is employed to test the strength and direction of the relationship between GenAI usage and perceived impact on the four determined learning outcomes (learning, creativity, motivation, engagement). Data Analysis was performed using SPSS.

The Spearman's coefficient ranges from -1 to +1; positive signs indicate a positive relationship, whereas a negative sign means an inverse relationship, and zero indicates no correlation (Zar, 2005). The correlation coefficient can be interpreted as follows: "If $0 < r \leq 0.4$, low correlation, If $0.4 \leq r < 0.7$, moderate correlation, and If $0.7 \leq r < 1$, high correlation." (Kafle, 2019).

Additionally, for a better understanding of how GenAI usage shapes their perceptions, k-means clustering was employed to classify students based on their GenAI usage. Before performing K-means clustering, the dataset was standardized using the Z-score standardization method.

The Silhouette method was used to choose the optimal number of clusters (k). “Silhouette is the score of comparing within-cluster distances with between-cluster distances.” (Kim, 2023) The silhouette score ranges from -1 to +1, and the higher score indicates better clustering quality.

After performing k-means, Principal Component Analysis (PCA) was used to reduce the dimensionality to visualize the clusters (Ding and He, 2004). RStudio software was used to conduct the analysis, and R’s functions such as scale(), silhouette(), and k-means() were utilized.

Lastly, in order to draw conclusions about the characteristics of the clusters and for descriptive profiling purposes, we applied T-tests and one-way ANOVA to determine if there are statistically significant differences in the mean value. T-tests and One-way ANOVA were performed using R’s t.test() and aov() functions. For the ANOVA tests that turned out to be significant, we used Scheffe post-hoc tests from RStudio’s DescTools library to determine which pairs of means are statistically significant.

4. Findings

4.1. Correlation method

Throughout Spearman’s correlation analysis, findings reveal the following:

- Variable v1 “*Generative AI supports my ability to analyze and solve complex problems*”.

A statistically significant relationship between the frequent use of GenAI for academic purposes and the ability to analyze and solve problems. ($r = 0.276$, $p = 0.007$) Results show that the relationship was stronger when using GenAI for the purpose of” research *and studying”,* as the correlation coefficient reached 0.487. A moderate correlation ($r = 0.439$, $p = 0.000$) was found when students it particularly for *“Solving problems”,* indicating that the direct use for problem-solving itself enhances the students’ feeling of their analytical competence. Results also show that there is a weaker correlation with the purpose of *“Generating ideas”* ($r = 0.300$, $p = 0.003$) (**Error! Reference source not found.**).

- Variable v2 *“Compared to traditional learning methods, how effective do you find Generative AI in understanding complex topics?”*

It is apparent from **Error! Reference source not found.** that the strongest correlation is with the more frequent use of GenAI tools (correlation = 0.336, p-value = 0.001), revealing that when students adopt these tools more frequently, they have a higher perception of their ability to understand complex topics, and that GenAI tools are more effective as compared to traditional methods. A statistically significant weak correlations were found when students frequently utilize these tools for “*Research and studying*” (correlation = 0.283, $p = 0.006$) and “*Solving problems*” (correlation = 0.216, $p = 0.038$). The results also indicate that *“Confidence in GenAI*” has a positive correlation with students’ perceived ability of their understanding ($r = 0.239$, $p = 0.021$), suggesting that an increase in the confidence of GenAI tools results in an increase in students’ perceptions of how they can understand complex concepts. (**Error! Reference source not found.**)

- Variable v4 “*I feel that Generative AI expands my creative thinking beyond what I would normally generate myself*”

Finding reveal significant low relationships with the frequent use of GenAI for “research and studying ($r = 0.288$, $p = 0.005$) and for “idea generation and brainstorming” ($r = 0.250$, $p < 0.05$). This means that students who use GenAI tools during the initial stages of thinking and brainstorming are more likely to perceive them as helpful in generating new and creative ideas. (**Error! Reference source not found.**)

- Variable v7 “*Using generative AI increases my motivation to engage with academic content.”*

Analysis revealed weak positive correlations with the frequency of GenAI use for the purposes “*Solving problems*”, “*Research and studying*”, and “*Writing essays/reports*”, with the values: $r = 0.336$, $r = 0.333$, $r = 0.205$, respectively. This can be interpreted as the students who utilize GenAI more frequently for “*solving problems*” and “*research*” are likely to sense higher motivation for learning. And with $r = 0.205$ for “writing essays/reports”, it is likely that students feel more motivated to engage with learning, but a relatively lower association. (**Error! Reference source not found.**)

- Variable v8 “*Generative AI encourages me to explore new academic topics beyond what is required.”*

It was found that using GenAI tools for academic tasks more frequently is relatively correlated with higher motivation in terms of encouraging students to set goals to explore more topics. ($r = 0.219$, $p = 0.035$). Particularly, the higher frequency of utilizing GenAI tools for the purpose of solving problems was found to have a stronger correlation with making students more motivated for exploring academic materials ($r = 0.243$, p = 0.019). And even a stronger relationship when used for (research and studying), as the correlation reached 0.313. (**Error! Reference source not found.**)

- Variable v9 "*Generative AI makes learning more interactive and engaging for me",*

It was found that the sample of students who are leveraging GenAI tools at a higher rate feel a higher sense of engagement and interactivity (correlation = 0.245, p = 0.018).

The results confirm that GenAI could enhance the overall learning experience for students through providing interactive content, supporting comprehension, and encouraging students to learn and explore. Therefore, it can be inferred that leveraging GenAI in academic work has the potential to support both the cognitive and emotional aspects of learning. (**Error! Reference source not found.**)

Notably, Spearman's correlation analysis did not reveal statistically significant relationships between the variables:

- V3: Do you believe that Generative AI reduces your ability to think critically and independently?
- V10: Using generative AI increases my participation in class or group discussions.
- V11: Does the availability of Generative AI reduce your motivation to study independently?

And the independent variables associated with GenAI tools, including usage frequency, frequency of use for particular academic purposes, and confidence in GenAI. The results indicate that perceptions related to the reduction of critical thinking, reduction of motivation for studying independently, as well as class participation, do not correlate with GenAI usage, the purposes of use, or confidence in these tools.

Variable	Frequency of Use	Frequency (Research &Studying)	Frequency (Solving Problems)	Frequency (Generating Ideas)	Frequency (Writing)	Confidence in GenAI
V1	0.276**	0.487**	0.439**	0.300**	0.189	0.158
V2	0.336**	0.283**	0.216*	0.165	0.086	0.239*
V4	0.103	0.288**	0.147	0.250*	0.145	0.134
V7	0.170	0.333**	0.336 **	0.150	0.205*	0.198
V8	0.219*	0.313**	0.243*	0.172	0.137	0.136
V9	0.245*	0.178	0.168	0.183	0.109	0.121

** p < 0.05, ** p < 0.01 V1: "Generative AI supports my ability to analyze and solve complex problems"., V2: "Compared to traditional learning methods, how effective do you find Generative AI in understanding complex topics?" V4: "I feel that Generative AI expands my creative thinking beyond what I would normally generate myself", V7: "Using generative AI increases my motivation to engage with academic content.", V8 "Generative AI encourages me to explore new academic topics beyond what is required.", V9: "Generative AI makes learning more interactive and engaging for me".*

Table 2. Spearman Correlation results

4.2. Clustering method

In this work, two models were employed to classify participants based on their perceptions of GenAI's impact on their learning experiences. The first model focuses on the cognitive outcomes, using the following key variables: GenAI usage frequency, support for problem-solving (v1), effectiveness compared to traditional learning (v2), concerns about critical thinking loss (v3), and support for creativity (v4).

The second model aims to classify students based on how they perceive GenAI to influence their affective outcomes, and the variables: GenAI usage frequency, motivation (v7), encouragement to explore topics (v8), interactivity (v9), and class participation (v10) were used.

In the clustering based on cognitive outcomes, the dataset was standardized to z-scores, and with the R language function *silhouette()*, average silhouette scores were calculated to choose the right number of clusters. K = 4 was the value that gave the highest score (**Error! Reference source not found.**), and thus four clusters were used when applying the *k-means* function. Our results show that participants were classified into the four clusters as follows: 18, 20, 20, and 35, respectively. Figure 5 shows the PCA clusters. The results show four different patterns in students' interactivity with GenAI tools (Table).

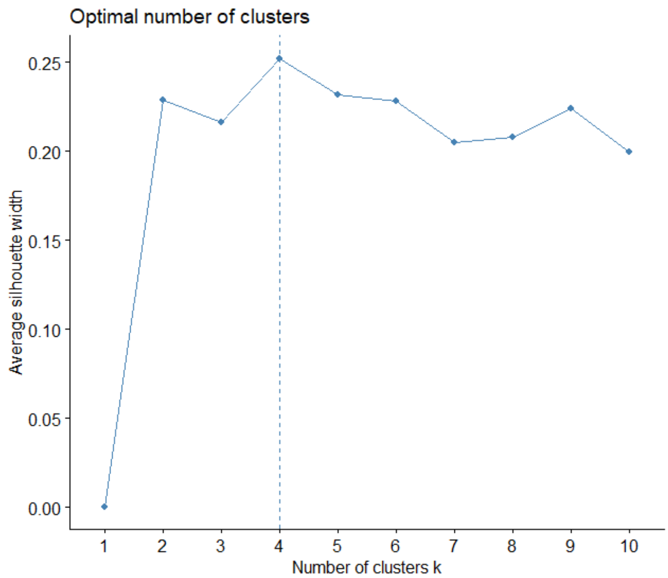

Figure 4. Optimal number of cognitive outcomes clusters

These clusters differ in terms of usage frequency, concerns about negative effects, and cognitive impact perceptions.

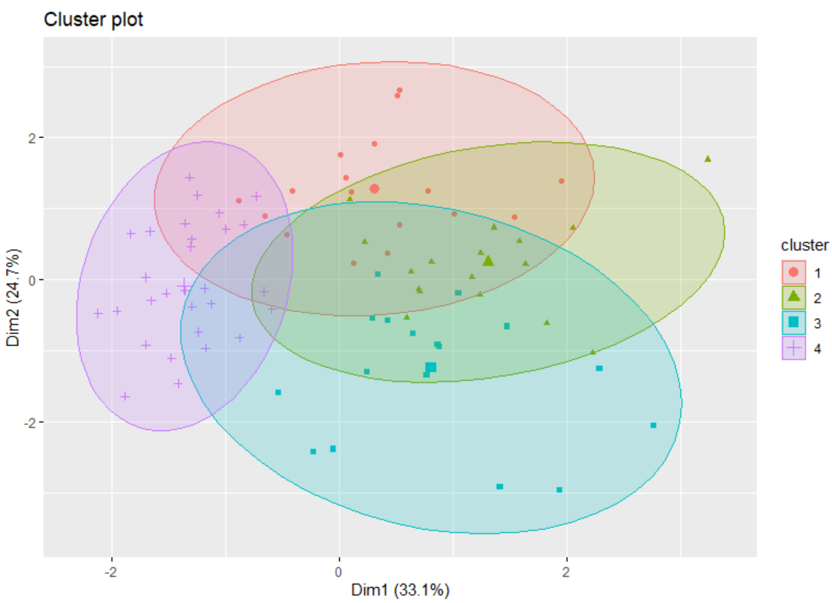

Figure 5. PCA clusters

Variables / Clusters	Cluster 1: Moderate but Cognitively Concerned	Cluster 2: Neutral and Low Engagement	Cluster 3: High Usage, Low Cognitive Belief	Cluster 4: Active and Cognitively Positive
GenAI Usage Frequency	3.88	2.10	4.05	4.26
Support Problem-Solving	3.44	3.05	2.45	4.43
Effectiveness Compared to Traditional Learning	2.55	3.60	4.35	4.80
Concern about Loss of Critical Thinking	4.00	3.15	2.50	3.46
Support for Creativity	3.56	2.85	2.05	3.80

Table 3. means of the clusters

GenAI usage frequency: ANOVA analysis revealed a significant difference in the usage level of GenAI tools among the four clusters. ($F (3, 89) = 42.72$, $p < 0.001$) Post-hoc Scheffe tests indicated that cluster 2 has a significantly lower usage level than cluster 1. Both clusters 3 and 4 have significantly higher GenAI usage than cluster 2, indicating various usage patterns among the students.

Support for problem solving: Findings of ANOVA revealed significant differences among the clusters in terms of their agreement that GenAI supports in their problem-solving ($F (3, 89) = 30.42$, $p < 0.001$). Scheffe tests indicated that cluster 3 has a significantly lower level of agreement than cluster 1. Cluster 4 significantly demonstrated stronger agreement that GenAI aids in solving problems compared to clusters 2 and 3.

Effectiveness compared to traditional learning: The analysis shows significant differences in the perceptions of the four clusters in terms of how effective GenAI is compared to traditional learning ($F (3, 89) = 58.27$, $p < 0.001$). It is apparent that cluster 4 has the highest perceived effectiveness of GenAI tools, significantly more than clusters 1 and 2, followed by cluster 3 and then cluster 2. The differences between cluster 1 and the other clusters were also significant, suggesting that cluster 1 has the lowest level of agreement among all.

Concerns about the loss of critical thinking: ANOVA analysis for the concerns about the loss of critical thinking due to relying on GenAI tools revealed significant differences between the clusters ($F (3, 89) = 6.65$, $p < 0.001$). Cluster 1 showed the highest level of concerns about GenAI reducing the ability of critical thinking. Notably, cluster 3 has the lowest level of concern and was significantly different from clusters 1 and 4. Cluster 4 has the second highest level of concern, just after Cluster 1.

Support for creativity: The ANOVA analysis revealed significance in the mean value between the clusters for their perceptions on how it supports their creativity. Scheffe post hoc tests show that cluster 4 has the strongest agreement that GenAI tools support creativity, whereas cluster 3 reported the least agreement. It is noteworthy that there are many significant pairs of means, particularly between clusters 1 and 3, clusters 3 and 4, and clusters 2 and 4.

Based on the significant differences observed in the data, cluster 1 can be labeled as “moderate but cognitively concerned” as they represent the moderate usage group but the most cognitively concerned. Cluster 2 labeled as “neutral and low usage” as they are significantly low frequent users who are in a neutral position about how GenAI impacts cognitive outcomes. As for clusters 3 and 4, they can be labeled as “high usage and low cognitive believe” and “active and cognitively positive”, respectively. This is grounded by how cluster 3 recorded the second highest usage frequency, and at the same time, it showed a limited cognitive impact, and that cluster 4 is the most frequent user and optimistic towards GenAI cognitive impact.

Overall, the findings confirm substantial differences and unexpected patterns in students' attitudes about GenAI, particularly in cluster 3, where they were frequent users (second highest frequency rate) and had the lowest levels of agreement that GenAI supports creativity or problem solving, but also the lowest level of agreement that the reliance on these tools could reduce critical thinking. Notably, cluster 4 demonstrated the highest frequency users and they had the most optimistic perception for effectiveness, support for problem solving and creativity, yet they were the most concerned group about the loss of critical thinking. These various patterns suggest that the more frequent use of GenAI does not always translate to more optimistic perceptions. Our results suggest that students’ attitudes toward GenAI are different, and their perceptions are formed in a complex way between the perceived benefits and potential concerns.

Variable	**F**	**df (between, within)**	**p-value**
GenAI usage frequency	42.72	(3, 89)	< 0.001***
Support for problem-solving	30.42	(3, 89)	< 0.001***
Effectiveness Compared to Traditional Learning	58.27	(3, 89)	< 0.001***
Concerns about critical thinking	6.65	(3, 89)	< 0.001***
Support for creativity	21.77	(3, 89)	< 0.001***

*** p < 0.001.

Table 4. Anova Test

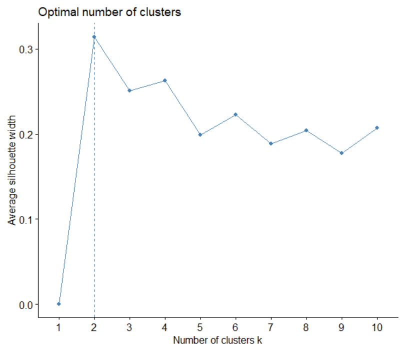

Figure 6. Optimal number of affective outcomes clusters

Considering the clustering based on the affective outcomes, the data was first rescaled into z-scores, and using the R *silhouette()* function, it was determined that k = 2 gives the largest silhouette score (Figure 6). Therefore, two clusters were used in the application of the *k-means* function. The results show that 42 and 51 participants were classified in cluster 1 and cluster 2, respectively.

The means (M) of cluster 2 among all variables (GenAI usage frequency, motivation, encouragement to explore topics, interactivity, class participation) were higher than those in Table 5.

Figure 7 shows the PCA visualization for the clusters. In order to compare the affective perceptions among the two clusters, independent sample t-tests were performed, and their results are shown in Table 5.

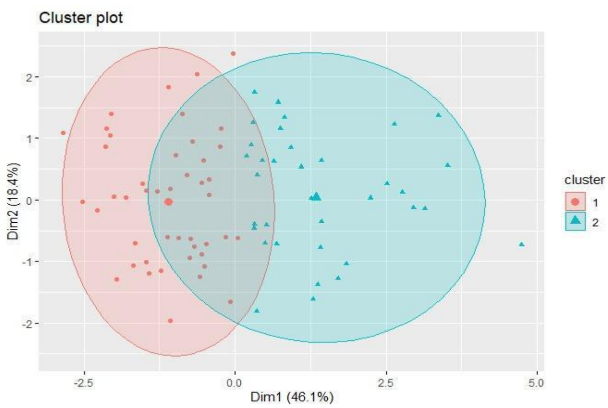

Figure 7. PCA visualization

The results revealed statistically significant differences ($p < 0.05$) in all five variables between the two clusters. Cluster 2 students reported significantly higher levels of motivation, curiosity to explore new topics, interactivity, and class participation, compared to students of cluster 1.

Thus, cluster 1 is labeled as “Low-Motivation and Low-Engagement Users,” while cluster 2 as “Highly Motivated and Engaged Users”. Findings of this study support that higher frequency usage of GenAI tools is associated with higher levels of motivation and engagement in learning contexts. It is noteworthy that the largest mean difference between the two clusters was in the variable related to encouragement to explore new topics.

Variable	Cluster 1 (n = 42)		Cluster 2 (n = 51)		t	P
	M	***SD***	***M***	***SD***		
GenAI usage frequency	**3.19**	**1.254**	**4.08**	**0.744**	-4.040***	0.000
Motivation	**2.52**	**0.833**	**3.80**	**0.664**	-8.067***	0.000
Encouragement to explore topics	**2.33**	**0.954**	**4.02**	**0.620**	-9.388***	0.000
Interactivity	**2.67**	**1.052**	**4.24**	**0.651**	-8.429***	0.000
Class participation	**2.26**	**1.061**	**3.51**	**0.987**	-5.825***	0.000

*** $p < 0.001$.

Table 5. Independent Samples t-Tests

5. Discussion and conclusion

The analysis shows that students are increasingly adopting GenAI tools in their academic tasks for different purposes, indicating that these tools play an important role in the lives of higher education students. Thus, understanding and adopting these tools is an inevitable reality. In our findings, students view GenAI tools positively in general, and particularly in facilitating comprehension, enhancing efficiency, and increasing enthusiasm towards learning. However, at the same time, a decent number of students reported some concerns and risks related to the adoption of GenAI tools, especially in regard to the difficulty of validating the accuracy of the content generated by these tools which aligns with what Zhu et al. (2023) stated that accuracy was a problem noticed by students, the overreliance on them and the potential drawback of reducing critical thinking, that was similar to Zhu et al. (2023) findings that many students think that AI might affect them badly by thinking on behalf of them. In addition, challenges related to ethical aspects and the necessary skills and competencies to use these tools were posed.

Furthermore, Spearman’s correlation analysis revealed that students with higher usage rates feel an increase in problem solving, creativity, motivation, and engagement in their learning processes. However, the clustering analysis findings emphasize a large variety in the evaluation of GenAI’s cognitive impact, indicating that a higher usage rate does not necessarily lead to increased positive perception of GenAI’s cognitive impact.

As for the affective aspect, the cluster analysis took similar patterns with correlations. The clusters were a group of those who use it more frequently with higher levels of engagement and motivation in their learning process, which aligns with part of Liang et al. (2023)’s findings, as they stated that the students who use it more have higher self-efficacy as well as a group with less frequent usage and lower levels of motivation and engagement.

This study highlights that GenAI use enhances both motivation and learning efficiency, yet overreliance may hinder critical thinking. Future work should explore these paradoxes with larger and more diverse samples. Our findings call for a balance between innovation and responsibility in AI-integrated education.

5.1. Linking Findings to Theory and Contributions

The results reinforce the mechanisms of **Technology Acceptance Models (TAM/UTAUT)** by showing that perceived usefulness and performance expectancy explain why higher use correlates with stronger learning outcomes and motivation. At the same time, the “high-use but low-belief” cluster illustrates that **digital literacy moderates** these pathways, highlighting that critical evaluation skills are essential for GenAI adoption to yield positive cognitive gains. Scientifically, the study contributes by (i) integrating cognitive and affective outcomes in one framework, (ii) moving beyond descriptive analysis through clustering, and (iii) bridging adoption models with responsible-use perspectives.

5.2. Practical and Policy Implications

For universities, these findings suggest that the success of GenAI integration depends not only on access but also on **instructional design and governance**. Institutions should:

- Develop **clear policies** on responsible use, plagiarism, and verification.
- Incorporate **AI literacy modules** into curricula to train students in prompt design, fact-checking, and ethical considerations.
- Support faculty in creating assessments that balance AI-assisted creativity with independent reasoning. Such practices can maximize motivation and engagement while safeguarding critical thinking and academic integrity.

5.3. Limitations and Future Research

This study is limited by its sample size (n = 93) and its reliance on self-reported, cross-sectional data, which reduce generalizability and preclude causal inference. Because the participants were recruited from a single institutional context, disciplinary and cultural variations in GenAI use could not be captured, further constraining external validity. Moreover, certain constructs such as creativity and participation were measured with single items. While this approach minimizes survey length, it may restrict construct reliability and fail to capture the multidimensional nature of creativity and engagement. Future studies should adopt validated multi-item scales to strengthen psychometric robustness.

The policy implications discussed in this study are exploratory and may vary across institutional contexts depending on infrastructure, digital literacy, and governance frameworks. Therefore, the recommended actions (e.g., responsible-use policies, AI-literacy modules, faculty training) should be adapted to local needs rather than applied uniformly.

Future research should use larger, more diverse and multi-institutional samples, apply longitudinal designs to capture changes over time, and combine surveys with qualitative methods (e.g., interviews, focus groups) to uncover deeper insights. Testing moderated models (e.g., SEM) could also clarify how digital literacy and institutional conditions shape the adoption and outcomes of GenAI.

References

Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. Discover Education, 3(1), 60. https://doi.org/10.1007/s44217-024-00060-0

Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. Sustainability, 15(17), 12983. https://doi.org/10.3390/su151712983

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Ding, C., & He, X. (2004, July). K-means clustering via principal component analysis. In Proceedings of the 21st International Conference on Machine Learning (pp. 29–36). ACM. https://doi.org/10.1145/1015330.1015408

Habib, S., Vogel, T., Anli, X., & Thorne, E. (2024). How does generative artificial intelligence impact student creativity? Journal of Creativity, 34(1), 100072. https://doi.org/10.1016/j.yjoc.2024.100072

Kafle, S. C. (2019). Correlation and regression analysis using SPSS. The OCEM Journal of Management, Technology, and Social Sciences, 1(1), 125–132. https://doi.org/10.3126/ojmts.v1i1.29287

Kim, S., Cho, S., Kim, J. Y., & Kim, D. J. (2023). Statistical assessment on student engagement in asynchronous online learning using the k-means clustering algorithm. Sustainability, 15(3), 2049. https://doi.org/10.3390/su15032049

Liang, J., Wang, L., Luo, J., Yan, Y., & Fan, C. (2023). The relationship between student interaction with generative artificial intelligence and learning achievement: Serial mediating roles of self-efficacy and cognitive engagement. Frontiers in Psychology, 14, 1285392. https://doi.org/10.3389/fpsyg.2023.1285392

Mittal, U., Sai, S., & Chamola, V. (2024). A comprehensive review on generative AI for education. IEEE Access, 12, 10345–10360. https://doi.org/10.1109/ACCESS.2024.3355555

Muñoz, S. A. S., Gayoso, G. G., Huambo, A. C., Tapia, R. D. C., Incaluque, J. L., Aguila, O. E. P., & Arias-Gonzáles, J. L. (2023). Examining the impacts of ChatGPT on student motivation and engagement. Social Space, 23(1), 1–27.

Neji, W., Boughattas, N., & Ziadi, F. (2023). Exploring new AI-based technologies to enhance students' motivation. Issues in Informing Science & Information Technology, 20, 123–136. https://doi.org/10.28945/5135

Sandu, R., Gide, E., & Elkhodr, M. (2024). The role and impact of ChatGPT in educational practices: Insights from an Australian higher education case study. Discover Education, 3(1), 71. https://doi.org/10.1007/s44217-024-00071-x

Schei, O. M., Møgelvang, A., & Ludvigsen, K. (2024). Perceptions and use of AI chatbots among students in higher education: A scoping review of empirical studies. Education Sciences, 14(8), 922. https://doi.org/10.3390/educsci14080922

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. Journal of Data Warehousing, 5(4), 13–22.

Van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2014). The digital divide shifts to differences in digital skills: A model and critical review. *New Media & Society, 16*(3), 507–526. https://doi.org/10.1177/1461444813487959

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Yu, H., & Guo, Y. (2023, June). Generative artificial intelligence empowers educational reform: Current status, issues, and prospects. Frontiers in Education, 8, 1183162. https://doi.org/10.3389/feduc.2023.1183162

Zar, J. H. (2005). Spearman rank correlation. In B. Everitt & D. Howell (Eds.), Encyclopedia of statistics in behavioral science (pp. 4973–4975). Wiley. https://doi.org/10.1002/0470013192.bsa598

Zhu, G., Fan, X., Hou, C., Zhong, T., Seow, P., Shen-Hsing, A. C., & Poh, T. L. (2023). Embrace opportunities and face challenges: Using ChatGPT in undergraduate students' collaborative interdisciplinary learning. arXiv preprint arXiv:2305.18616. https://doi.org/10.48550/arXiv.2305.18616

Journal of Information and Organizational Sciences
Volume 49, Number 2 (2025)
Journal homepage: jios.foi.hr

JIOS

DOI: 10.31341/jios.49.2.11
Open Access

UDC 004.056:004.8:519.6:61
Original Scientific Paper

A Hybrid IG-PCA and Machine Learning Approach for Accurate Intrusion Detection in IoMT with Imbalanced Data

Willy Riyadi[1*], Kurniabudi[1], Jasmir[1], Yudi Novianto[1], Desi Kisbianty[1] and Xaverius Sika[1]

[1]Faculty of Computer Science, Universitas Dinamika Bangsa, Jambi, Indonesia
*Correspondence: willyriyadi@unama.ac.id

PAPER INFO

Paper history:
Received 22 August 2025
Accepted 02 December 2025

Citation:
Riyadi, W., Kurniabudi, Jasmir, Novianto, Y., Kisbianty, D. & Sika, X. (2025). A Hybrid IG-PCA and Machine Learning Approach for Accurate Intrusion Detection in IoMT with Imbalanced Data. In Journal of Information and Organizational Sciences, vol. 49, no. 2, pp. 345-359

ABSTRACT

The rapid growth of the Internet of Medical Things (IoMT) has introduced critical cybersecurity challenges, highlighting the need for robust and accurate intrusion detection systems (IDS). This study presents a hybrid machine learning (ML) framework to strengthen intrusion detection in IoMT networks using the CIC-IoMT2024 dataset. The framework combines Information Gain (IG) and Principal Component Analysis (PCA) for feature selection and dimensionality reduction, while SMOTEENN and SMOTETomek are applied to address severe class imbalance. The processed data are classified using Random Forest (RF), K-Nearest Neighbors (KNN), XGBoost (XGB), Multi-Layer Perceptron (MLPC), and Logistic Regression (LR), with hyperparameters optimized through Bayesian Optimization. Performance is evaluated using Accuracy, Precision, Recall, F1-Score, and AUC. Experimental results reveal that the optimized XGB classifier with SMOTEENN achieves a peak accuracy of 99.811%. This top-tier performance surpasses several existing benchmarks, validating the effectiveness of integrating IG-PCA with advanced resampling and optimization strategies. This work contributes a lightweight, scalable, and highly accurate IDS, offering a practical and efficient solution for enhancing security in resource-constrained, next-generation medical IoT systems.

Keywords: Internet of Medical Things, Intrusion Detection Accuracy, IG-PCA, Machine Learning, Bayesian Optimization

1. Introduction

The proliferation of the Internet of Medical Things (IoMT) has catalyzed a paradigm shift in healthcare, enabling innovations such as remote patient monitoring, automated diagnostics, and interconnected smart health services (Alturki et al., 2025; Razdan & Sharma, 2022; Wang et al., 2021). This digital transformation holds immense potential for improving patient outcomes and operational efficiency. However, the increasing connectivity and inherent heterogeneity of IoMT devices have also introduced significant security vulnerabilities (Ahmed et al., 2024; Binbusayyis et al., 2022; Mathkor et al., 2024). Operating in resource-constrained environments and often lacking standardized security protocols, these devices are prime targets for cyberattacks, making data privacy, secure communication, and real-time threat detection paramount concerns (Papaioannou et al., 2022; Zachos et al., 2021).

To mitigate these threats, the development of robust Intrusion Detection Systems (IDS) specifically designed for IoMT ecosystems has become a critical area of research (Alalhareth & Hong, 2024; Ibrahim & Al-Wadi, 2024). m (ML)-based IDS have demonstrated considerable promise by learning network behavior to intelligently classify legitimate and malicious traffic (Berguiga et al., 2025; Chaganti et al., 2022). Nonetheless, their practical implementation is hindered by several persistent challenges, including the 'curse of dimensionality' from high-dimensional data, severe class imbalance where attack data is sparse, and the stringent requirement for lightweight models that can operate in real-time on devices with limited computational power (Husain et al., 2025).

In response, the literature presents several strategies. To manage data complexity, hybrid methods combining Information Gain (IG) for feature selection and Principal Component Analysis (PCA) for dimensionality reduction have been explored to enhance classification accuracy while minimizing computational load (Kumar et al., 2022; Nasir et al., 2022; Odhiambo Omuya et al., 2021; R.M. et al., 2020). To counteract class imbalance, over-sampling and hybrid sampling techniques like SMOTEENN and SMOTETomek are widely used to improve model sensitivity towards minority attack classes (Alsharaiah et al., 2025; Bouke et al., 2024; Sarkar et al., 2024; Talukder et al., 2024). Recent advanced models have achieved high accuracy; for instance, systems integrating deep learning with blockchain have reported accuracies of 99.7% across 18 attack types (Abdiwi, 2024). while federated learning approaches have reached 97.31%. Similarly, combining SMOTE with classical ML models has yielded accuracies up to 99.2% (Mohsin & Jony, 2024).

Beyond the technical vulnerabilities, security breaches in IoMT ecosystems pose direct threats to organizational stability and patient safety. A successful cyberattack could lead to the compromise of sensitive patient health information (PHI), disruption of critical clinical workflows, or even manipulation of life-sustaining medical devices, causing irreparable harm to patients and catastrophic reputational damage to healthcare providers (Chuma & Ngoepe, 2022). Despite these advancements, a significant research gap persists. Many existing solutions focus on maximizing a single metric like accuracy, often at the cost of computational efficiency, model interpretability, or scalability—qualities that are non-negotiable for real-world IoMT deployment. A holistic framework that systematically integrates and optimizes preprocessing, feature engineering, and model tuning to create a balanced, lightweight, and effective IDS is still needed. This study addresses this gap by proposing a comprehensive ML-based framework to optimize intrusion detection, validated on the contemporary CIC-IoMT 2024 dataset (Dadkhah et al., 2024). Our approach prioritizes a pragmatic balance between high detection performance and computational feasibility, ensuring its suitability for resource-constrained IoMT infrastructures.

The major contributions of this research are as follows:

1. Development of an Optimized IDS Framework: We propose and evaluate an end-to-end intrusion detection framework that synergistically integrates data preprocessing, feature engineering (using IG and PCA), and a comparative analysis of advanced data balancing techniques (SMOTEENN and SMOTETomek). This creates a robust and computationally efficient system specifically tailored for IoMT environments.
2. Systematic Hyperparameter Optimization: We employ Bayesian Optimization (BO) for automated and efficient hyperparameter tuning of multiple ML classifiers. This approach systematically enhances model performance and reduces the manual effort and computational cost associated with traditional grid-search or random-search methods.
3. Comprehensive Performance Validation: The proposed framework is rigorously validated on the recent and highly relevant CIC-IoMT 2024 dataset. This provides a robust and up-to-date benchmark of its effectiveness against modern cyber threats, assessing performance across multiple metrics, including accuracy, precision, recall, F1-score, and Area Under Curve (AUC).

We present a comprehensive performance evaluation of our proposed method against baseline models and other optimization approaches from the literature. This demonstrates that our framework achieves an enhanced balance between high detection accuracy and practical computational efficiency.

2. Related work

The growing complexity and interconnectivity of IoMT systems have prompted extensive research into developing secure, efficient, and intelligent IDS. A significant body of literature has focused on leveraging ML to identify malicious behavior in healthcare-related IoT environments. Traditional signature-based IDS approaches are often ineffective in IoMT contexts due to their inability to detect novel or evolving threats. Consequently, ML-based IDS have gained traction due to their adaptive learning capabilities and ability to generalize from data. Studies such as those by (Chaganti et al., 2022) and (Ibrahim & Al-Wadi, 2024) have demonstrated the feasibility of ML models like Decision Trees, Random Forests, and Support Vector Machines in IoMT-specific datasets. These models achieved high detection accuracy but often required further optimization for real-time deployment due to their computational overhead.

Efficient feature engineering plays a pivotal role in enhancing IDS performance, particularly when handling high-dimensional IoMT data. To improve detection speed and simplify model complexity, several studies have adopted hybrid feature reduction techniques. For instance, (Sayed et al., 2023) and (Kumari & Jain, 2024), combined Information Gain (IG) and Principal Component Analysis (PCA), achieving higher classification accuracy and reduced training time. Similarly, (Odhiambo Omuya et al., 2021) developed a hybrid PCA-IG model to address the challenges of high-dimensional datasets. By minimizing redundant and irrelevant attributes, the PCA-IG model significantly improved classification accuracy and computational efficiency. This approach achieved superior performance in accuracy, precision, recall, and training time compared to conventional feature selection methods and classifiers such as Naïve Bayes. In addition to feature optimization, researchers have explored various data resampling techniques to mitigate class imbalance. Methods such as SMOTE (Synthetic Minority Over-sampling Technique), SMOTEENN, and SMOTETomek have been widely employed to enhance model sensitivity toward minority classes. (Husain et al., 2025) and (Alsharaiah et al., 2025) reported substantial improvements in minority class detection using these balancing strategies, while (Mohsin & Jony, 2024) integrated SMOTE-based balancing with classical classifiers, achieving accuracy gains exceeding 99% across multiple attack categories.

Furthermore, (Doménech et al., 2025) demonstrated the significant impact of preprocessing and balancing techniques, such as Random Undersampling and SMOTE, on the CIC-IoMT2024 dataset, improving accuracy by 26.35% and the F1-score by 29.40% over baseline models. However, they also observed that SMOTE occasionally reduced the detection rate for minority classes, such as "ARP Spoofing," due to the generation of ambiguous synthetic samples. Complementarily, (Torre et al., 2025) employed Federated Learning integrated with Differential Privacy and Homomorphic Encryption to ensure decentralized training with strong privacy protection, albeit at the expense of higher computational complexity. Overall, while the existing literature presents numerous innovations in IDS for IoMT, several limitations persist. Many models are designed for general IoT scenarios and do not account for the stringent latency and power requirements in healthcare settings. Additionally, model interpretability, essential for healthcare providers, is often overlooked. Few studies offer scalable solutions that combine lightweight processing with high detection accuracy and balanced class performance.

3. Proposed methodology

The methodology of this study follows a structured ML pipeline designed to construct and evaluate an effective IDS for IoMT networks. The entire workflow, from data preparation to performance evaluation, is illustrated in Figure 1 to validate their effectiveness and suitability for deployment in resource-constrained IoMT environments.

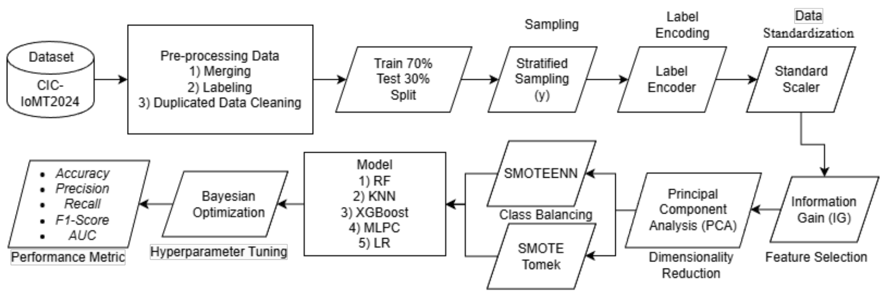

Figure 1. Proposed Method

3.1. Datasets Description

This research is grounded in the CIC-IoMT 2024 dataset (Dadkhah et al., 2024), provided by the Canadian Institute for Cybersecurity. The dataset encompasses diverse and representative network traffic patterns commonly observed in IoMT environments, including both normal operations and various forms of malicious activity such as DDoS, infiltration, and other significant cyber threats.

3.2. Data Preprocessing and Sampling

The initial stage of this study involved a systematic preprocessing pipeline applied to the CIC-IoMT2024 dataset to ensure high data quality and efficient computational handling. Initially, all raw data files were merged into a unified dataframe. A deduplication procedure was then performed to eliminate redundancy, resulting in the removal of 5,119 duplicate entries. This yielded a clean dataset consisting of 7,155,712 unique records, each described by 46 attributes. Each record in the dataset was categorized using the 'Label' feature as either 'Benign' or one of 18 distinct malicious attack types, as illustrated in Figure 2. To prepare the dataset for ML, all categorical variables, including the target 'Label', were converted into numerical form using Label Encoding.

The dataset was subsequently divided into training and testing subsets using a 70/30 split ratio, 5,008,998 samples for training (X_train, y_train) and 2,146,714 samples for testing (X_test, y_test). Given the large dataset size, training models directly on the full data would lead to substantial computational overhead. To address this, stratified random sampling was employed to maintain the proportional representation of all classes within both subsets, ensuring consistency with the original class distribution and minimizing sampling bias. A maximum threshold of 15,000 samples per class was established for downsampling the majority classes. Classes with counts exceeding this threshold (e.g., TCP_IP-DDoS-UDP, MQTT-DDoS-Connect_Flood) were randomly reduced to 15,000 samples, while minority classes with fewer instances (e.g., Recon-Ping_Sweep, Recon-VulScan) were retained entirely. This hybrid downsampling strategy effectively reduced the size of the training data while preserving the diversity across all 19 classes. The resulting sampled training dataset comprised 232,087 records, with class distribution summarized as follows:
Benign (15,000), TCP_IP-DoS-UDP (15,000), TCP_IP-DDoS-TCP (15,000), TCP_IP-DDoS-SYN (15,000), TCP_IP-DDoS-ICMP (15,000), MQTT-DoS-Publish_Flood (15,000), MQTT-DDoS-Connect_Flood (15,000), MQTT-DDoS-Publish_Flood (15,000), TCP_IP-DoS-ICMP (15,000), TCP_IP-DoS-SYN (15,000), Recon-Port_Scan (15,000), TCP_IP-DoS-TCP (15,000), TCP_IP-DDoS-UDP (15,000), Recon-OS_Scan (11,314), ARP_Spoofing (11,233), MQTT-DoS-Connect_Flood (8,941), MQTT-Malformed_Data (3,591), Recon-VulScan (1,490), and Recon-Ping_Sweep (518).

To enhance the performance and convergence of the learning algorithms, feature normalization was applied using StandardScaler, which transformed each feature to have a mean of zero ($\mu = 0$) and a standard deviation of one ($\sigma = 1$), as expressed in Equation (1).

$$z = (x - \mu) / \sigma \tag{1}$$

Standardization is a critical step to ensure that all features are on a comparable scale, preventing features with larger magnitudes from disproportionately influencing the model.

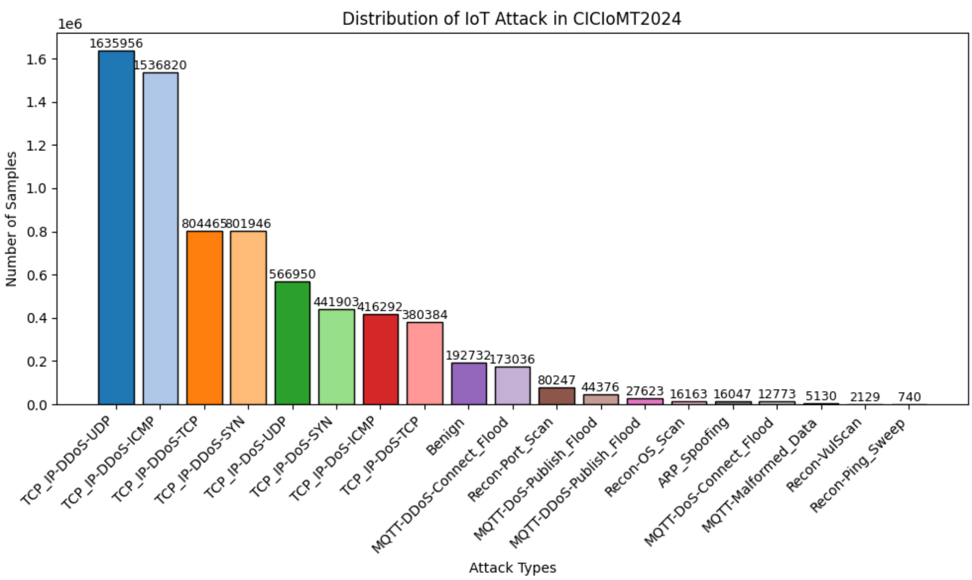

Figure 2. Distribution of cyberattacks

3.3. Feature Selection

IG is employed to select features with the highest predictive power by measuring their mutual information with the target class. It is calculated as the reduction in entropy after splitting the dataset based on an attribute. Formally, IG for an attribute A relative to a dataset S as shown in Equation (2):

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (2)$$

Where, $Entropy(S)$ is the entropy of the entire dataset, Values A denotes the set of possible values for attribute A, S_v is the subset of S where attribute A has value v, $\frac{|S_v|}{|S|}$ represents the proportion of the dataset falling into subset S_v. A higher IG value indicates a stronger dependency between the feature and the target variable, suggesting greater predictive relevance. The IG scores were computed for all 42 standardized features using the training dataset. Based on the distribution of IG values, an empirical threshold of IG $>$ 0.001 was established to differentiate high-impact from low-impact features. Consequently, 40 features with IG scores above this threshold were retained for further model training and evaluation. The ranked list of features is presented in Table 1.

Rank	Feature Name	IG Value
1	IAT	2.644043
2	Tot sum	1.742246
3	Tot size	1.711671
4	Min	1.648812
5	Max	1.637822
6	AVG	1.609903
7	Magnitue	1.603614
8	Header_Length	1.537972
9	Srate	1.422746
10	Rate	1.422336
…	…	…
31	ARP	0.028318
32	IPv	0.027613
33	LLC	0.025614
34	DNS	0.009515
35	HTTP	0.007317
36	IRC	0.006482
37	cwr_flag_number	0.004747
38	Telnet	0.003675
39	SMTP	0.001834
40	ece_flag_number	0.001489
41	SSH	0.000744

Table 1. IG result

The remaining three features, namely Drate, DHCP, and IGMP, exhibited IG scores of 0.000000, indicating negligible contribution to class discrimination. Therefore, these features were excluded from further analysis to reduce noise and enhance model efficiency.

3.4. Dimensionality Reduction

To further refine the feature space and mitigate potential multicollinearity among the 42 features previously selected by Information Gain (IG), Principal Component Analysis (PCA) was applied. PCA is an unsupervised linear transformation technique that projects the data into a new subspace defined by orthogonal principal components (PCs) that capture the maximum variance within the original feature set. Considering the high-dimensional and multi-class characteristics of the CIC-IoMT2024 dataset, the number of principal components was empirically determined to be 31. This configuration was selected to ensure that both dominant and subtle variance patterns were retained—critical for distinguishing among the 19 different traffic classes, including those with closely related characteristics. The transformation process is grounded in the covariance relationship between features, mathematically expressed in Equation (3):

$$cov_{x,y} = \frac{\sum(x_i - \underline{x})\left(y_i - \underline{y}\right)}{N-1} \tag{3}$$

Where x_i and y_i are individual sample values, $\underline{x}$ and $\underline{y}$ are the mean values of variables x and y, and N denotes the total number of observations. The PCA transformation demonstrated that 31 components collectively explained 99.99% of the total variance, confirming that the reduced feature space preserved nearly all relevant information from the IG-selected features while significantly minimizing redundancy. This dimensionality reduction enhanced computational efficiency and reduced the potential risk of overfitting during model training. The transformed training and testing datasets exhibited shapes of (232,087, 31) and (2,146,714, 31), respectively, indicating successful feature compression while maintaining discriminative power. Table 2 summarizes the proportion of variance explained by each principal component. The first few components—PC1 (21.20%), PC2 (9.67%), PC3 (7.58%), PC4 (7.44%), and PC5 (6.57%)—collectively accounted for over 52% of the total variance, demonstrating that most of the variability in the data was captured by a relatively small number of components.

Principal Component (PC)	Individual Variance Explained (%)	Cumulative Variance Explained (%)
PC1	21.20	21.20
PC2	9.67	30.87
PC3	7.58	38.45
PC4	7.44	45.90
PC5	6.57	52.47
PC6	5.20	57.67
PC7	4.43	62.10
PC8	4.25	66.35
PC9	4.15	70.50
PC10	4.01	74.51
…	…	…
PC26	0.33	99.21
PC27	0.28	99.50
PC28	0.23	99.73
PC29	0.13	99.85
PC30	0.08	99.94
PC31	0.05	99.99

Table 2. PCA result

3.5. Class Balancing

A persistent challenge in most cybersecurity datasets, including the CIC-IoMT2024 dataset used in this study, is class imbalance. This issue arises when certain attack types or benign traffic dominate the dataset, causing machine learning (ML) models to become biased toward majority classes and perform poorly in detecting minority or rare attacks. To address this problem, two robust hybrid resampling techniques were applied to the PCA-transformed training data (comprising 31 components): SMOTEENN and SMOTETomek. Both methods combine synthetic over-sampling of minority classes with under-sampling or cleaning mechanisms to achieve a more balanced and representative dataset.

1. **SMOTEENN:** The Synthetic Minority Over-sampling Technique combined with Edited Nearest Neighbors (SMOTEENN) first generates synthetic samples for underrepresented attack classes to increase their presence in the dataset. Subsequently, the ENN cleaning process removes ambiguous or noisy samples from both majority and minority classes to refine the decision boundaries. Applying this approach produced a training dataset consisting of 227,025 samples across all 19 classes. The post-resampling distribution showed improved balance, with class counts ranging approximately between 5,692 and 14,909 instances per class.

2. **SMOTETomek:** The SMOTE combined with Tomek Links (SMOTETomek) method similarly begins by synthesizing minority samples using SMOTE and then identifies Tomek links—pairs of samples from different classes that are nearest neighbors. The majority-class instances involved in these links are removed to better delineate class boundaries. This method resulted in a slightly larger, more uniform dataset containing 271,764 samples, with class counts ranging approximately between 12,013 and 14,983 instances per class.

Both methods successfully mitigated the class imbalance observed in the original dataset (where class counts ranged from 518 to 15,000). While SMOTEENN introduced a moderately balanced structure emphasizing data cleanliness, SMOTETomek achieved a near-uniform class distribution and clearer decision margins. These balanced datasets were subsequently used to independently train and evaluate the machine learning models to assess their robustness and generalization across balanced conditions..

3.6. Model

Four ML models were chosen due to their established effectiveness in intrusion detection tasks:

1) Random Forest (RF) builds a large collection of decision trees and then aggregates their individual predictions. This aggregation process boosts the model's overall accuracy and reduces the risk of overfitting to the training data (Al-Abadi et al. (2023)).
2) K-Nearest Neighbors (KNN) is a non-parametric algorithm used for classification by labeling a data point according to the most common class among its k nearest neighbors within the feature space (Sun & Chen (2021)).
3) XGBoost (XGB) is an advanced gradient boosting algorithm. It develops a predictive model by creating a series of decision trees one after another, with each new tree aiming to rectify the inaccuracies of the preceding models in the sequence (Salehpour et al. (2024)).
4) Logistic Regression (LR) is a fundamental statistical algorithm used for binary classification tasks. Despite its name, it is a classification model that calculates the probability of a specific outcome. It employs the logistic (sigmoid) function to transform the output of a linear equation into a probability score between 0 and 1, making it highly interpretable and computationally efficient (Chalichalamala et al. (2023)).
5) Multi-Layer Perceptron Classifier (MLPC) is a supervised artificial neural network model consisting of interconnected layers of nodes: input, hidden, and output. It applies nonlinear activation functions to capture complex patterns in data and employs backpropagation for weight optimization, enabling robust classification performance across diverse domains, particularly in high-dimensional and nonlinear problem spaces (Zhao et al. (2025)).

3.7. Hyperparameter Optimization

To fine-tune each model and maximize its predictive performance, BO was employed. BO efficiently explores the hyperparameter search space by building a probabilistic surrogate model that estimates the objective function based on prior evaluations. In this study, the objective function $f(x)$ represents the mean 5-fold cross-validation accuracy for a given set of hyperparameters x. The optimization iteratively updates its belief about $f(x)$ using a Gaussian Process (GP) and selects the next candidate point x^* by maximizing an acquisition function, typically the Expected Improvement (EI). Mathematically, the BO process can be expressed as (4).

$$x^* = \arg\max_{x \in X} \alpha(x \mid D_t), \quad \text{where} \quad \alpha(x \mid D_t) = \mathbb{E}[max(f(x) - f(x^+), 0)] \tag{4}$$

Here, $D_t = \{(x_i, f(x_i))\}_{i=1}^{t}$ denotes the set of observed evaluations, $f(x^+)$ is the best performance obtained so far, and $\alpha(x \mid D_t)$ represents the acquisition function guiding the exploration–exploitation trade-off. This automated and probabilistically guided tuning process ensures that each classifier—trained on both SMOTEENN and SMOTETomek balanced datasets—is optimized toward its most effective configuration before final evaluation, thereby enhancing the robustness, generalizability, and reliability of our findings.

3.8. Evaluation

To validate the dependability and stability of the proposed models, we evaluated their performance on intrusion detection within the IoMT environment using a diverse range of evaluation metrics. These metrics include:

1) $Accuracy = \frac{TN+TP}{TN+FP+TP+FN}$, measuring overall correctness
2) $Precision = \frac{TP}{TP+FP}$, measures the fraction of true positive cases among all instances classified as positive, indicating how accurately the model identifies positive outcomes.
3) $Recall = \frac{TP}{TP+FN}$, reflecting the model's sensitivity to actual positives.
4) $F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$, balancing precision and recall.
5) $AUC = TPR * \frac{TP}{TP + FN}$, evaluates a model's capability to differentiate between classes in classification tasks, providing insight into its overall discriminative power.

This evaluation framework facilitates the effective development of a smart IDS tailored for the intricate IoMT landscape.

4. Result and Analysis

This research presents a thorough evaluation of the ML models developed for intrusion detection within IoMT environments. A systematic performance analysis is conducted for each model using datasets balanced through SMOTEENN and SMOTETomek techniques. The results are critically interpreted and positioned within the context of existing scholarly works to underscore the novel contributions of this study. The implementation was carried out on Google Colab Pro, leveraging high-performance CPUs and substantial RAM resources to ensure efficient execution. This computationally robust configuration facilitates the handling of memory-intensive operations and accelerates processing, thereby enhancing the overall experimental workflow.

4.1. Experimental Design

The discussion section interprets these results, compares them with existing literature, and highlights the contributions of this research. The evaluation was conducted using five state-of-the-art ML classifiers: RF, KNN, XGB, MLPC and LR. Each model was trained and tested on two distinct datasets that were processed using the IG-PCA feature engineering pipeline and balanced with either SMOTEENN or SMOTETomek. Model performance was assessed using standard evaluation metrics: Accuracy, Precision, Recall, F1-Score and AUC.

To ensure each model achieved its maximum potential, all classifiers were fine-tuned using the BO process. A predefined search space, outlining the range of possible values for each hyperparameter, was established to guide the optimization. The specific search spaces for each classifier are detailed in Tables 3 through 7.

Hyperparameter	Type	Search Space
'n_neighbors'	Integer	[100, 1000]
'max_depth'	Integer	[5, 100]
'min_samples_split'	Integer	[2, 20]
'min_samples_leaf'	Integer	[1, 10]
'max_features'	Categorical	[sqrt, log2, none]
'bootstrap'	Categorical	[True, False]
'criterion'	Categorical	['gini', 'entropy']

Table 3. RF Search Space Perimeter

Hyperparamete r	Type	Search Space
'n_neighbors'	Integer	[1, 90]
'weights'	Categorical	['uniform', 'distance']
'metric'	Categorical	['euclidean', 'manhattan', 'minkowski']
'algorithm'	Categorical	['auto', 'ball_tree', 'kd_tree', 'brute']
'leaf_size'	Integer	[10, 900]
'p'	Integer	[1, 2]

Table 4. KNN Search Space Perimeter

Perimeter name	Type	Search Space
'n_estimators’	Integer	[50, 500]
'max_depth’	Integer	[3, 90]
'learning_rate’	Float	[0.01, 0.3]
'subsample’	Float	[0.5, 1.0]
'colsample_bytree’	Float	[0.5, 1.0]
'gamma’	Float	[0, 5]
'min_child_weight’	Integer	[1, 10]
'reg_alpha’	Float	[0, 5]
'reg_lambda’	Float	[0, 5]
‘scale_pos_weight’	Float	[0.5, 5]

Table 5. XGB Search Space Perimeter

Hyperparameter	Type	Search Space
'Penalty'	Categorical	['l1', 'l2', 'elasticnet', 'none']
'C'	Real	[1e-3, 1e3, prior='log-uniform']
'solver'	Categorical	['lbfgs', 'liblinear', 'saga', 'newton-cg']
'max_iter'	Integer	[100, 2000]
'l1_ratio'	Real	[0, 1, prior='uniform']

Table 6. LR Search Space Perimeter

Hyperparameter	Type	Search Space
'n_layers'	Categorical	[1, 2, 3]
'hl1'	Integer	32 – 512
'hl2'	Integer	16 – 512
'hl3'	Integer	8 – 512
'activation'	Categorical	['relu', 'tanh', 'logistic']
'solver'	Categorical	['adam', 'sgd', 'lbfgs']
'alpha'	Real	1e-5 – 1e-1 (log-uniform)
'learning_rate'	Categorical	['constant', 'invscaling', 'adaptive']
'learning_rate_init'	Real	1e-4 – 1e-1 (log-uniform)
'max_iter'	Integer	200 – 2000
'batch_size'	Integer	32 – 512
'early_stopping'	Categorical	[True, False]

Table 7. MLPC Search Space Perimeter

4.2. Performance on the SMOTEENN-Balanced Dataset

Models trained on the dataset balanced with the SMOTEENN technique demonstrated exceptional classification performance. Notably, ensemble-based methods yielded the most competitive results. Specifically, RF and XGB achieved near-perfect scores across all evaluation metrics. This highlights their robustness in mitigating class imbalance and their capacity for modeling complex decision boundaries.

The detailed performance metrics following BO are presented in Table 8, with a corresponding graphical representation in Figure 3. These results underscore the superior performance of the ensemble models. XGB attained the highest scores across all metrics, with an accuracy, precision, recall, and F1-score of 99.811%. RF followed closely with highly comparable performance metrics. Both models achieved a near-perfect Area Under the Curve (AUC) of 99.999%, indicating exceptional discriminative capability.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
RF	99.795	99.795	99.795	99.795	99.999
XGB	99.811	99.811	99.811	99.811	99.999
KNN	93.923	93.943	93.923	93.901	99.773
LR	78.047	83.261	78.047	76.071	99.117
MLPC	95.815	95.961	95.815	95.808	99.817

Table 8. The BO result for SMOTEENN

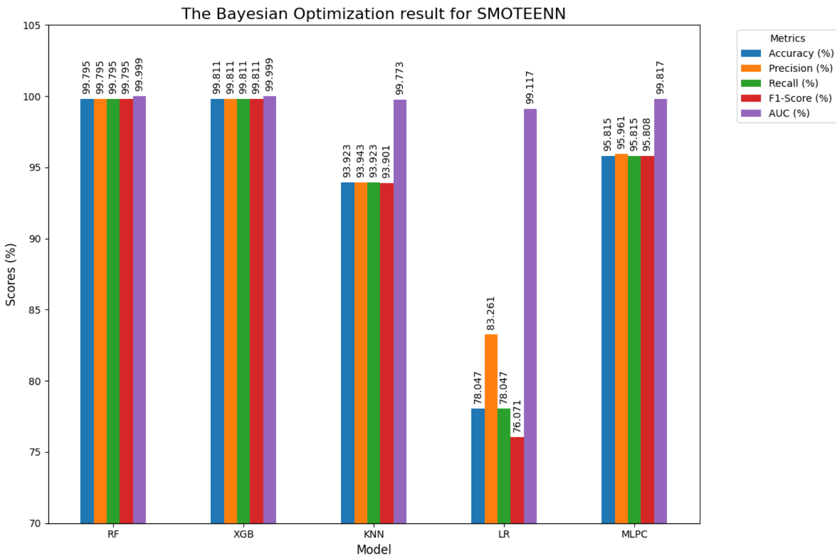

Figure 3. The BO result for SMOTEENN

BO identified the optimal configuration for RF as comprising 224 estimators with a maximum depth of 18. The use of the entropy criterion and max_features = 'log2' facilitated the development of diverse and minimally correlated trees, thereby enhancing the ensemble's robustness. Furthermore, setting min_samples_leaf = 3 and min_samples_split = 10 effectively mitigated overfitting by ensuring adequate sample representation in each terminal node and controlling the tree splitting process. This configuration enabled the RF model to achieve a weighted F1-score of 0.9977, reflecting strong predictive performance across all classes.

For XGB, the optimal setup consisted of 224 estimators, a maximum depth of 25, and a learning rate of 0.1357, achieving a balance between model complexity and convergence speed. Parameters such as colsample_bytree = 0.7224 and subsample = 0.8517 promoted ensemble diversity by limiting feature and sample usage per tree. The regularization terms reg_alpha = 0.7516 and reg_lambda = 3.2294 were instrumental in preventing overfitting by constraining model complexity. Additionally, gamma = 0.0524 and min_child_weight = 4.5367 regulated partitions to avoid excessive granularity. The scale_pos_weight = 1.4732 parameter addressed any residual class imbalance. This comprehensive tuning resulted in a weighted F1-score of 0.9981, marginally surpassing RF and demonstrating its outstanding classification power.

In contrast, the non-ensemble models demonstrated lower, albeit still effective, performance. The KNN classifier achieved a respectable F1-score of 93.901% and an AUC of 99.773%. LR yielded more moderate results, with an F1-score of 76.071% and an AUC of 99.117%, indicating reasonable but comparatively lower effectiveness. In summary, the findings confirm that for IoMT intrusion detection, ensemble methods, particularly XGB and RF, when coupled with the SMOTEENN balancing technique, deliver the most accurate, robust, and generalizable classification performance.

4.3. Performance on the SMOTETomek-Balanced Dataset

Models trained on the dataset balanced using the SMOTETomek algorithm demonstrated consistently high classification performance. The results affirmed the superiority of ensemble-based approaches, which delivered the most competitive outcomes. Both RF and XGB achieved near-perfect scores across all evaluation metrics, validating their robustness in handling class imbalance and their capability to model intricate decision boundaries.

The detailed outcomes of the BO for hyperparameter tuning are cataloged in Table 9 and visualized in Figure 4. Consistent with the findings from the SMOTEENN dataset, ensemble methods emerged as the top performers. XGB was the leading model, achieving the highest weighted F1-score of 98.892%, marginally surpassing RF, which secured an F1-score of 98.752%. Both models also attained exceptional AUC values, indicating robust discriminative power.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
RF	98.754	98.784	98.754	98.752	99.975
XGB	98.897	98.899	98.897	98.892	99.983
KNN	91.959	91.943	91.959	91.901	99.647
LR	76.719	81.287	76.719	74.773	98.861
MLPC	94.212	94.985	94.212	94.057	99.694

Table 9. The BO result for SMOTETomek

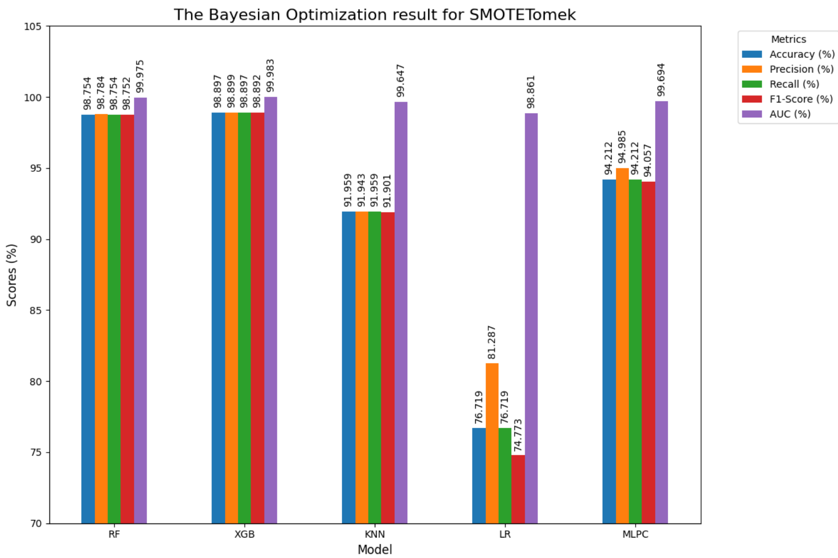

Figure 4. The BO result for SMOTETomek

BO identified the optimal configuration for RF vs 224 estimators with a maximum depth of 18. The selection of the entropy criterion alongside max_features = 'log2' was instrumental in fostering the growth of diverse and minimally correlated trees, which enhanced the overall robustness of the ensemble. To prevent overfitting, min_samples_leaf = 3 and min_samples_split = 10 were implemented, ensuring that each terminal node was supported by a sufficient number of samples while controlling the frequency of splits. This carefully tuned configuration enabled RF to achieve strong predictive accuracy and balanced performance.

For XGB, the optimal hyperparameter set included 224 estimators, a maximum depth of 25, and a learning rate of 0.1357, a combination designed to strike a balance between the model's predictive power and its convergence stability. Overfitting was further mitigated by setting colsample_bytree = 0.7224 and subsample = 0.8517, which promoted ensemble diversity by restricting the proportion of features and samples used for each tree. The regularization terms (reg_alpha = 0.7516, reg_lambda = 3.2294) and split-control parameters (gamma = 0.0524, min_child_weight = 4.5367) effectively constrained model complexity. Finally, scale_pos_weight = 1.4732 was applied to address any residual class imbalance. This optimized setup confirmed XGB's potent and well-balanced classification capability.

The other models also yielded noteworthy results. KNN and the MLPC delivered solid performance, with F1-scores of 91.901% and 94.057%, respectively. In contrast, LR showed comparatively lower predictive accuracy, suggesting a greater sensitivity to the data structure produced by the SMOTETomek process. In conclusion, these findings reinforce that for IoMT intrusion detection, ensemble methods—particularly XGB and RF —when paired with the SMOTETomek balancing technique, provide the most accurate, robust, and generalizable classification performance.

4.4. Comparative Analysis of proposed techniques

The integration of IG for feature selection and PCA for dimensionality reduction proved highly effective in refining the dataset. By prioritizing the most informative features and eliminating redundant or irrelevant attributes, this preprocessing strategy substantially reduced data complexity while preserving critical discriminative information. This step not only enhanced computational efficiency but also contributed to improving model generalization capability. In addition, the application of the SMOTEENN and SMOTETomek balancing methods successfully mitigated the issue of class distribution imbalance. These techniques ensured that the models remained unbiased toward majority classes, thereby improving their ability to accurately detect minority-class attack instances—a crucial requirement for intrusion detection in IoMT environments. As shown in Table 10, our proposed XGB model achieved an accuracy of 99.811% on the CIC-IoMT 2024 dataset. This performance surpasses several benchmarks, including the RF model from Dadkhah et al. (73.3%) and the XGB model from Mohsin & Jony (99.20%). Notably, our model's accuracy is highly comparable to other state-of-the-art methods, such as those by Lucia Hernandez-Jaimes et al. (99.83%) and Ghourabi & Alkhalil (99.85%), validating the effectiveness of the optimized XGB pipeline proposed in this study.

Model	Accuracy
RF with 19 Class (Dadkhah et al., 2024)	73.3%
RF with 2 Class + RandomOverSampler (Abdiwi, 2024)	99.7%
XGB with 19 Class (Mohsin & Jony, 2024)	99.20%
XGB with 6 Class + SMOTE (Doménech et al., 2025)	99.83%
XGB with 2 Class + Federated Learning + BO (Ghourabi & Alkhalil, 2025)	99.85%
Proposed XGB with 19 Class + IG-PCA + SMOTEENN + BO (this paper)	99.811%

Table 10. Comparison best accuracy result for CIC-IoMT 2024 dataset

4.5. Organizational Implications and Practical Adoption

Beyond its technical accuracy, the practical viability of the proposed IDS is determined by its integration within a healthcare organization's existing operational, governance, and strategic frameworks. This section discusses the key organizational implications for its successful adoption.

For adoption and implementation, the proposed framework is designed as a software-based solution that can be deployed on a dedicated server or virtual machine within the hospital's data center. Integration into the existing infrastructure would involve configuring the system to monitor network traffic from core switches or gateways, a task manageable by the organization's network and cybersecurity teams. While a data scientist may be beneficial for initial model fine-tuning, the lightweight nature of the IG-PCA pipeline minimizes the need for specialized, high-cost hardware, making it accessible for institutions with limited resources. The primary requirement is a moderately provisioned server and collaboration between existing IT personnel for setup and maintenance, ensuring a low barrier to entry.

Effective governance is crucial for translating the system's alerts into actionable responses. Responsibility for the IDS would typically reside with the Information Security team, under the direct oversight of the Chief Information Security Officer (CISO). A clear incident response protocol must be established: upon detecting a potential threat, the IDS would generate an alert, ideally feeding into a central Security Information and Event

Management (SIEM) system (González-Granadillo et al., 2021). Tier-1 security analysts would then be responsible for initial validation. Confirmed critical threats would be escalated to a dedicated incident response team, with the CISO holding the authority to approve decisive actions, such as isolating compromised medical devices to contain the threat and protect the wider network.

The strategic benefits for the healthcare organization extend far beyond technical threat detection. By providing robust security, the IDS directly enhances patient trust in the hospital's digital services, which is critical for the adoption of telehealth and online patient portals. Furthermore, implementing such an advanced security measure helps the organization demonstrate due diligence and achieve compliance with stringent data protection regulations and cybersecurity accreditation standards. Proactively identifying threats reduces the risk of operational downtime in critical clinical systems, preventing significant financial losses and ensuring continuity of patient care. Ultimately, the system serves as a strategic asset, providing the CISO with valuable threat intelligence to inform risk management policies and justify security investments (Ramezan, 2025).

5. Conclusion

This study proposed and validated a comprehensive framework for intrusion detection tailored to the unique challenges of the IoMT, including high-dimensional feature spaces and pronounced class imbalance. The methodology integrated a multi-stage data preprocessing pipeline consisting of IG for feature selection, PCA for dimensionality reduction, and hybrid resampling through SMOTEENN and SMOTETomek to achieve data balance. This approach effectively enhanced data quality and improved overall model robustness. Furthermore, BO was systematically employed to fine-tune the hyperparameters of five classifiers—RF, KNN, XGB, LR, and MLPC—resulting in consistent performance gains across all models. Notably, the XGB classifier combined with the SMOTEENN technique achieved a peak accuracy of 99.811% on the CIC-IoMT2024 dataset. This performance is highly competitive with state-of-the-art approaches and surpasses several existing benchmarks. Although the proposed XGB model did not achieve the absolute highest accuracy, the difference (0.039%) is marginal and statistically insignificant given the 19-class complexity of the dataset. Importantly, the proposed framework achieved this result with lower computational overhead and improved class balance, underscoring its scalability and deployment potential in real-world IoMT intrusion detection systems.

Overall, the findings demonstrate the efficacy and practical viability of the proposed framework in developing accurate, efficient, and scalable intrusion detection systems suitable for resource-constrained IoMT environments. This research contributes a practical tool that enhances the cybersecurity posture of healthcare institutions. By enabling reliable and efficient threat detection, the framework strengthens operational resilience, protects critical medical data, and supports governance structures essential for secure digital transformation in the healthcare sector. Future work will focus on two key directions. First, the scalability and robustness of the framework will be validated using larger and more heterogeneous IoMT datasets. Second, lightweight deep learning architectures will be explored to further reduce detection latency while adhering to the strict energy efficiency and performance constraints inherent in IoMT devices.

Acknowledgement

This research is a part of Fundamental Research funded under contract number 012/MOU/LPPM/UNAMA/1V /2025, and supported by Yayasan Dinamika Bangsa.

References

Abdiwi, F. G. (2024). Hybrid Machine Learning and Blockchain Technology for Early Detection of Cyberattacks in Healthcare Systems. International Journal of Safety and Security Engineering, 14(6), 1883–1893. https://doi.org/10.18280/ijsse.140622

Ahmed, S. F., Alam, M. S. Bin, Afrin, S., Rafa, S. J., Rafa, N., & Gandomi, A. H. (2024). Insights into Internet of Medical Things (IoMT): Data fusion, security issues and potential solutions. Information Fusion, 102. https://doi.org/10.1016/j.inffus.2023.102060

Al-Abadi, A. A. J., Mohamed, M. B., & Fakhfakh, A. (2023). Enhanced Random Forest Classifier with K-Means Clustering (ERF-KMC) for Detecting and Preventing Distributed-Denial-of-Service and Man-in-the-Middle Attacks in Internet-of-Medical-Things Networks. Computers, 12(12), 262. https://doi.org/10.3390/computers12120262

Alalhareth, M., & Hong, S. C. (2024). Enhancing the Internet of Medical Things (IoMT) Security with Meta-Learning: A Performance-Driven Approach for Ensemble Intrusion Detection Systems. Sensors, 24(11). https://doi.org/10.3390/s24113519

Alsharaiah, M. A., Almaiah, M. A., Shehab, R., Obeidat, M., El-Qirem, F. A., & Aldhyani, T. (2025). An explainable AI-driven transformer model for spoofing attack detection in Internet of Medical Things (IoMT) networks. Discover Applied Sciences, 7(5), 488. https://doi.org/10.1007/s42452-025-07071-5

Alturki, B., Abu Al-Haija, Q., Alsemmeari, R. A., Alsulami, A. A., Alqahtani, A., Alghamdi, B. M., Bakhsh, S. T., & Shaikh, R. A. (2025). IoMT landscape: navigating current challenges and pioneering future research trends. In Discover Applied Sciences (Vol. 7, Issue 1). Springer Nature. https://doi.org/10.1007/s42452-024-06351-w

Berguiga, A., Harchay, A., & Massaoudi, A. (2025). HIDS-IoMT: A Deep Learning-Based Intelligent Intrusion Detection System for the Internet of Medical Things. IEEE Access, 13, 32863–32882. https://doi.org/10.1109/ACCESS.2025.3543127

Binbusayyis, A., Alaskar, H., Vaiyapuri, T., & Dinesh, M. (2022). An investigation and comparison of machine learning approaches for intrusion detection in IoMT network. The Journal of Supercomputing, 78(15), 17403–17422. https://doi.org/10.1007/s11227-022-04568-3

Bouke, M. A., El Atigh, H., & Abdullah, A. (2024). Towards robust and efficient intrusion detection in IoMT: a deep learning approach addressing data leakage and enhancing model generalizability. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-024-19916-z

Chaganti, R., Mourade, A., Ravi, V., Vemprala, N., Dua, A., & Bhushan, B. (2022). A Particle Swarm Optimization and Deep Learning Approach for Intrusion Detection System in Internet of Medical Things. Sustainability, 14(19), 12828. https://doi.org/10.3390/su141912828

Chalichalamala, S., Govindan, N., & Kasarapu, R. (2023). Logistic Regression Ensemble Classifier for Intrusion Detection System in Internet of Things. Sensors, 23(23), 9583. https://doi.org/10.3390/s23239583

Chuma, K. G., & Ngoepe, M. (2022). Security of electronic personal health information in a public hospital in South Africa. Information Security Journal: A Global Perspective, 31(2), 179–195. https://doi.org/10.1080/19393555.2021.1893410

Dadkhah, S., Neto, E. C. P., Ferreira, R., Molokwu, R. C., Sadeghi, S., & Ghorbani, A. A. (2024). CICIoMT2024: A benchmark dataset for multi-protocol security assessment in IoMT. Internet of Things, 28, 101351. https://doi.org/10.1016/j.iot.2024.101351

Doménech, J., León, O., Siddiqui, M. S., & Pegueroles, J. (2025). Evaluating and enhancing intrusion detection systems in IoMT: The importance of domain-specific datasets. Internet of Things, 32, 101631. https://doi.org/10.1016/j.iot.2025.101631

Ghourabi, A., & Alkhalil, A. (2025). A Federated Learning Model for Detecting Cyberattacks in Internet of Medical Things Networks. IEEE Access, 13, 123018–123030. https://doi.org/10.1109/ACCESS.2025.3588808

González-Granadillo, G., González-Zarzosa, S., & Diaz, R. (2021). Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures. Sensors, 21(14), 4759. https://doi.org/10.3390/s21144759

Husain, G., Nasef, D., Jose, R., Mayer, J., Bekbolatova, M., Devine, T., & Toma, M. (2025). SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models. Algorithms, 18(1), 37. https://doi.org/10.3390/a18010037

Ibrahim, M., & Al-Wadi, A. (2024). Enhancing IoMT network security using ensemble learning-based intrusion detection systems. Journal of Engineering Research (Kuwait). https://doi.org/10.1016/j.jer.2024.12.003

Kumar, A., Sangwan, S. R., Arora, A., & Menon, V. G. (2022). Depress-DCNF: A deep convolutional neuro-fuzzy model for detection of depression episodes using IoMT. Applied Soft Computing, 122, 108863. https://doi.org/10.1016/j.asoc.2022.108863

Kumari, P., & Jain, A. K. (2024). Timely detection of DDoS attacks in IoT with dimensionality reduction. Cluster Computing, 27(6), 7869–7887. https://doi.org/10.1007/s10586-024-04392-9

Mathkor, D. M., Mathkor, N., Bassfar, Z., Bantun, F., Slama, P., Ahmad, F., & Haque, S. (2024). Multirole of the internet of medical things (IoMT) in biomedical systems for managing smart healthcare systems: An overview of current and future innovative trends. Journal of Infection and Public Health, 17(4), 559–572. https://doi.org/10.1016/j.jiph.2024.01.013

Mohsin, M., & Jony, A. I. (2024). A Comparative Analysis of Medical IoT Device Attacks Using Machine Learning Models. Malaysian Journal of Science and Advanced Technology, 429–439. https://doi.org/10.56532/mjsat.v4i4.318

Nasir, M., Javed, A. R., Tariq, M. A., Asim, M., & Baker, T. (2022). Feature engineering and deep learning-based intrusion detection framework for securing edge IoT. The Journal of Supercomputing, 78(6), 8852–8866. https://doi.org/10.1007/s11227-021-04250-0

Odhiambo Omuya, E., Onyango Okeyo, G., & Waema Kimwele, M. (2021). Feature Selection for Classification using Principal Component Analysis and Information Gain. Expert Systems with Applications, 174, 114765. https://doi.org/10.1016/j.eswa.2021.114765

Papaioannou, M., Karageorgou, M., Mantas, G., Sucasas, V., Essop, I., Rodriguez, J., & Lymberopoulos, D. (2022). A Survey on Security Threats and Countermeasures in Internet of Medical Things (IoMT). Transactions on Emerging Telecommunications Technologies, 33(6). https://doi.org/10.1002/ett.4049

Ramezan, C. A. (2025). Understanding the chief information security officer: Qualifications and responsibilities for cybersecurity leadership. Computers & Security, 152, 104363. https://doi.org/10.1016/j.cose.2025.104363

Razdan, S., & Sharma, S. (2022). Internet of Medical Things (IoMT): Overview, Emerging Technologies, and Case Studies. In IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India) (Vol. 39, Issue 4, pp. 775–788). Taylor and Francis Ltd. https://doi.org/10.1080/02564602.2021.1927863

R.M., S. P., Maddikunta, P. K. R., M., P., Koppu, S., Gadekallu, T. R., Chowdhary, C. L., & Alazab, M. (2020). An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture. Computer Communications, 160, 139–149. https://doi.org/10.1016/j.comcom.2020.05.048

Salehpour, A., Norouzi, M., Balafar, M. A., & SamadZamini, K. (2024). A cloud-based hybrid intrusion detection framework using XGBoost and ADASYN-Augmented random forest for IoMT. IET Communications, 18(19), 1371–1390. https://doi.org/10.1049/cmu2.12833

Sarkar, M., Lee, T.-H., & Sahoo, P. K. (2024). Smart Healthcare: Exploring the Internet of Medical Things with Ambient Intelligence. Electronics, 13(12), 2309. https://doi.org/10.3390/electronics13122309

Sayed, N., Shoaib, M., Ahmed, W., Noman Qasem, S., M. Albarrak, A., & Saeed, F. (2023). Augmenting IoT Intrusion Detection System Performance Using Deep Neural Network. Computers, Materials & Continua, 74(1), 1351–1374. https://doi.org/10.32604/cmc.2023.030831

Sun, B., & Chen, H. (2021). A Survey of k Nearest Neighbor Algorithms for Solving the Class Imbalanced Problem. Wireless Communications and Mobile Computing, 2021(1). https://doi.org/10.1155/2021/5520990

Talukder, Md. A., Sharmin, S., Uddin, M. A., Islam, M. M., & Aryal, S. (2024). MLSTL-WSN: machine learning-based intrusion detection using SMOTETomek in WSNs. International Journal of Information Security, 23(3), 2139–2158. https://doi.org/10.1007/s10207-024-00833-z

Torre, D., Chennamaneni, A., Jo, J., Vyas, G., & Sabrsula, B. (2025). Toward Enhancing Privacy Preservation of a Federated Learning CNN Intrusion Detection System in IoT: Method and Empirical Study. ACM Transactions on Software Engineering and Methodology, 34(2), 1–48. https://doi.org/10.1145/3695998

Wang, J., Lim, M. K., Wang, C., & Tseng, M.-L. (2021). The evolution of the Internet of Things (IoT) over the past 20 years. Computers & Industrial Engineering, 155, 107174. https://doi.org/10.1016/j.cie.2021.107174

Zachos, G., Essop, I., Mantas, G., Porfyrakis, K., Ribeiro, J. C., & Rodriguez, J. (2021). An anomaly-based intrusion detection system for internet of medical things networks. Electronics (Switzerland), 10(21). https://doi.org/10.3390/electronics10212562

Zhao, Q., Wang, F., Wang, W., Zhang, T., Wu, H., & Ning, W. (2025). Research on intrusion detection model based on improved MLP algorithm. Scientific Reports, 15(1), 5159. https://doi.org/10.1038/s41598-025-89798-0

Contact

Igor Balaban
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, HR-42000 Varaždin, Croatia

Tel.: +385 42 390 832
Fax: +385 42 213 413
E-mail: igor.balaban@foi.hr
E-mail: jios@foi.hr (Corespondence Mail)
jios-pub@foi.hr (Publishing Board)
URL: www.jios.foi.hr

JIOS is published with financial support of the Ministry of Science, Education and Youth of the Republic of Croatia.

www.jios.foi.hr

Proofreading

Authors

Copyeditor and Computer Type Set

Mladen Konecki

Cover and Layout Design

Josipa Bađari
Mladen Konecki

JIOS is indexed in the following repositories and portals

Academic Journals Database
Academic Resource Index
Advanced Science Index
BASE
Central and Eastern European Online Library (CEEOL)
CORE
Crossref
CWTS Journal Indicators
Dimensions
Directory of Open Access Scholarly Resources (ROAD)
Directory of Open Access Journals (DOAJ)
Directory of Research Journals Indexing (DRJI)
Elektronische Zeitschriftenbibliothek
ERIH PLUS
EuroPub – Directory of Academic and Scientific Journals
Genamics JournalSeek
Google Scholar
HRČAK - Central portal of Croatian scientific journals
Index Copernicus
Internet Archive
Information Matrix for the Analysis of Journals (MIAR)
International Journal Impact Factor (IJIFACTOR)
J-Gate
JournalTOCs
LENS.ORG
Microsoft Academic
OAIster
Open AIRE
Open Academic Journals Index (OAJI.net)
PRADEC
Scientific Indexing Services (SIS)
Scilit
Sherpa Romeo
Ulrich's Periodicals Directory
VINITI RAN
WorldCat
World Journal Clout Index Report (WJCI)

JIOS is indexed in the following databases

SCOPUS
Emerging Sources Citation Index (ESCI) - WoS
Academic Search Complete (EBSCO)
Academic Search Elite (EBSCO)
Academic Search Premier (EBSCO)
Academic Search Ultimate (EBSCO)
Aerospace Database - ProQuest
Civil Engineering Abstracts - ProQuest
Central & Eastern European Academic Source (CEEAS) - EBSCO
Communication Abstracts - ProQuest
EBSCO Essentials
INSPEC
Library and Information Science Abstracts (LISA) - ProQuest
Library & Information Science Source - EBSCO
Library, Information Science & Technology Abstracts (LISTA) - EBSCO
Library Literature & Information Science - EBSCO
Metadex – ProQuest

Frequency

JIOS is published twice per year.

Access to JIOS is free of charge through following sources:

Journal of Information and Organizational Sciences
Hrčak - central portal of Croatian scientific journals
Directory of Open Access Journals (DOAJ)
OAIster

JIOS

CONTENTS