International Journal of
Technical and Natural Sciences

IJTNS

ISSN: 2671-3500

VOLUME 5 | ISSUE 1

BALKAN JOURNALS

INTERNATIONAL JOURNAL OF TECHNICAL AND NATURAL SCIENCES

VOLUME 5 | ISSUE 1
e-ISSN: 2671-3519

International Journal of Technical and Natural Sciences

Volume 5 - Issue 1 | June 2024

Address of the Editorial Office
Makedonsko - Kosovska Brigada Str. 1000, Skopje, Republic of North Macedonia

ijtns.ibupress.com

Published by Balkan University Press

Open access, peer-reviewed, bi-annual academic journal | e-ISSN: 2671-3519 | ISSN print: 2671-3500

Content

INTERNATIONAL JOURNAL OF TECHNICAL AND NATURAL SCIENCES

DOI: https://doi.org//10.69648/JEGY2656

International Journal of Natural and Technical Sciences (IJTNS), 2025; 5(1): 1-23

ijtns.ibupress.com

Online ISSN: 2671-3519

Application :03.06.2025

Revision :13.06.2025

Acceptance :13.06.2025

Publication :15.06.2025

Gupta, A., & Singh, N. P. (2025). Analysis of the Acquisition of Health Alliance Group Inc by Reliance Digital Health Limited. International Journal of Natural and Technical Sciences, 5(1), 1-23.

https://doi.org//10.69648/JEGY2656

Analysis of the Acquisition of Health Alliance Group Inc by Reliance Digital Health Limited

Anjali Gupta, N.P Singh

Anjali Gupta[1] and N.P Singh[2]

[1]Research Scholar, School of Business Management & Commerce, MVN University, Palwal-121105, Haryana, India, https://orcid.org/0009-0002-8027-8362

Email: 22ms9002w@mvn.edu.in

[2]Professor, School of Business Management & Commerce, MVN University, Palwal-121105, Haryana, India, https://orcid.org/0000-0002-3006-9522

Email: dr.npsingh@mvn.edu.in

Abstract

Reliance Digital Health Ltd (RDHL), a subsidiary of Reliance Industries Limited (RIL), has acquired a 45% stake in Health Alliance Group Inc. (HAGI) for $10 million in a strategic action to strengthen its position in the global digital healthcare landscape. This acquisition is a major step in increasing RDHL's operational effectiveness and broadening its reach, especially in underdeveloped areas like the US and India. Because of HAGI's superior telemedicine, diagnostic, and preventive care capabilities, RDHL has the chance to incorporate cutting-edge technologies into its service offerings and meet the growing demand for easily accessible, technologically advanced healthcare solutions. The aim of this study is to investigate the strategic significance of the RDHL–HAGI partnership. The research applies an investigative methodology based on secondary data collected from industry reports, corporate announcements, financial disclosures, and media articles. A fusion of qualitative and quantitative techniques, including trend analysis and textual review, is used to evaluate the acquisition's potential impact. The research findings suggest that this strategic alliance could significantly contribute to closing healthcare gaps in access and promoting a more efficient, inclusive, and technology-enabled healthcare ecosystem. Additionally, the partnership is predicted to improve RDHL's competitive positioning in the rapidly changing global digital health sector.

Keywords: Acquisition, Digital Healthcare, Telemedicine, Operational Efficiency, and Market Expansion.

BALKAN JOURNALS

Introduction

The global digital healthcare sector has observed a significant transformation in recent years, driven by rapid advancements in telemedicine, artificial intelligence, data analytics, and remote diagnostics. As healthcare systems across the world confront challenges related to accessibility, affordability, and efficiency - especially in Underprivileged and rural regions - digital health platforms have emerged as essential tools in bridging these gaps. Preventive care as well as virtual consultations grow increasingly popular since they offer scalable and affordable solutions while also lessening the strain on the conventional healthcare system. This digital shift is further supported by rising investments from both public and private sectors, positioning the industry for substantial growth and innovation on a global scale.

In this evolving landscape, Reliance Digital Health Limited (RDHL), a fully-owned subsidiary of Reliance Industries Limited (RIL), has identified itself as a disruptive force. With a December 2024 valuation of US$220 billion, RIL has shown its ability to have an impact on significant sectors like energy, communication, and now healthcare. As part of its strategic expansion, RDHL has announced the acquisition of a 45% stake in Health Alliance Group Inc. (HAGI) for US$10 million (Business Standard, 2024; Aoi Venture, 2024; Bizz Buzz, 2024; CNBC-TV18, 2024; Domain-b, 2024; and Digital Health News, 2024). Established in 1980 in Delaware, HAGI is a prominent leader in digital health platforms, offering innovative telemedicine, diagnostic, and preventative care services primarily to low-income, underrepresented groups. In countries like India, the US, and other nations with healthcare inequality, the acquisition aligns with RDHL's overall goal of creating an all-encompassing and integrated digital healthcare ecosystem that provides services that are both accessible and reasonably priced.

This study's goal is to investigate the strategic significance of RDHL's investment in HAGI and assess how it will impact the company's operations and standing in the global digital healthcare sector. In addition, the study aims to determine how this collaboration could encourage the formation of sustainable and scalable healthcare solutions, especially in areas with insufficient medical infrastructure. By analysing this acquisition, the study is trying to determine if RDHL has the capacity to become a world leader in digital healthcare innovation and to fill important healthcare gaps with technologically advanced solutions.

The overall investigation is determined by ten particular questions that form the conceptual framework of the research. These investigations aim to examine several

aspects of the transaction, like its operational impact, strategic value, long-term vision alignment, innovation integration, and impact on healthcare access. The questions are:

- **Research Question 1:** What are the strategic advantages of RDHL acquiring a 45% stake in HAGI?
- **Research Question 2:** Can this partnership help RDHL address healthcare delivery challenges in underserved global regions?
- **Research Question 3:** How will the integration of HAGI's technological innovations influence RDHL's current healthcare services portfolio?
- **Research Question 4:** Does this acquisition align with RDHL's long-term ambition to become a global leader in digital healthcare solutions?
- **Research Question 5:** How does this transaction contribute to RIL's overarching vision of leveraging technological innovation to drive growth?
- **Research Question 6:** Will this deal strengthen RDHL's competitive position in the global digital health sector?
- **Research Question 7:** What risks might arise from this acquisition, including integration hurdles and competitive pressures in the digital health space?
- **Research Question 8:** How will this collaboration affect RDHL's operational efficiency and market expansion strategies?
- **Research Question 9:** To what extent can this partnership improve healthcare accessibility in India and other underserved markets?
- **Research Question 10:** What implications does this partnership have for RDHL's global healthcare strategy?

Literature Review

The literature on mergers and acquisitions (M&As) in the healthcare industry is thoroughly reviewed in this section, with a focus on digital innovation and cross-border strategic integration. The purpose of this review is to identify theoretical perspectives, patterns, and research gaps that are relevant to the current study by reviewing prior research and industry reports. Positioning this research within the larger academic literature on healthcare M&As and digital transformation, as well as providing a conceptual basis for assessing Reliance Digital Health Ltd.'s (RDHL) acquisition of Health Alliance Group Inc. (HAGI), are the purposes.

The literature focuses on five main concepts: (1) strategic expansion into underdeveloped markets; (2) innovation performance and digital health integration; (3) ethical conduct and employee implications; (4) operational efficiency through M&A; and (5) current research gaps.

Operational Efficiency Through M&As

According to Barkema and Schijven (2008) and Salleh et al. (2013), mergers and acquisitions (M&As) enhance technical efficiency through the integration of technology, infrastructure, and human capital. Reduced expenses, resource optimisation, and the growth of synergy are the means by which these efficiencies are accomplished. According to recent research, strategic acquisitions like RDHL-HAGI can greatly enhance service delivery mechanisms in the context of digital healthcare, especially in domains with limited access (Angel One, 2024; Medical Buyer, 2024).

Digital Health Integration and Innovation

The primary justification for M&A in the high-tech and healthcare sectors is innovation. Innovation may be helped or hampered by the scope and nature of the acquired firm's knowledge base, according to Cloodt et al. (2006). After COVID-19, companies prioritise acquiring digital health capabilities as a competitive need, according to recent industry reports. According to Business Standard (2024), RDHL can enhance its innovation performance by applying HAGI's technology developments in diagnostics and telemedicine.

Ethical Conduct in Healthcare M&As

It is additionally well-known that behavioural integrity has a significant impact on M&A success. Organizational trust and the ethical management of transitions have a strong connection, according to Lin and Wei (2006). Promoting high ethical standards becomes crucial in the healthcare industry, where privacy, regulatory compliance, and personnel security are crucial. In accordance with industry reviews, ethical transparency during integration affects public impression as well as internal efficiency (McKinsey & Company, 2020).

Strategic Expansion Into Underserved Markets

In the opinion of the WHO (2021), digital health is the driving force behind the expansion of healthcare access in underprivileged communities. Following the year 2020, companies have used mergers and acquisitions (M&A) much more frequently as a means of entering underserved and emerging markets strategically. Global

parallels are offered by Singh (2014a, 2014b, 2015a, 2015b, 2017, 2019), where technology companies have acquired assets to scale quickly across geographies. In the case of Lenovo's following acquisitions of Motorola Mobility and IBM's x86 server division, it allowed it to gain a competitive presence in both emerging and Western markets, while Microsoft's acquisition of Nokia made it possible to enter the smartphone market. The above situations illustrate how cross-border acquisitions can be used as strategic instruments to enter untapped client niches and assimilate new technologies. In a comparable manner, RDHL's purchase of a 45 percent share in Health Alliance Group Inc. (HAGI) is part of a bigger plan to target underserved digital populations in Africa, the U.S., and India (Mint, 2024; Economic Times, 2024; MSN, 2024). In domains without solid physical healthcare systems, RDHL is well-positioned to address infrastructure gaps in healthcare delivery by combining Reliance's present infrastructure and market access with HAGI's telemedicine and diagnostic platforms. The adoption of M&A to scale technologies and promote inclusive health outcomes is a global trend that this fits with.

Identified Gap in Literature

Although research on M&A in the healthcare sector increases, only a few percent of studies focus on cross-border digital health acquisitions completed by developing market companies. The vast majority of the literature that is accessible right now focuses on pharmaceutical sector integrations or Western acquisitions. Thus, the RDHL–HAGI agreement is a rare instance of a major Indian company using advanced technology from abroad to fill healthcare gaps both domestically and internationally. A real-time, data-supported investigation of strategic digital healthcare expansion through acquisition is presented by this study, assisting in closing that gap.

Research Methodology

This study explores Reliance Digital Health Ltd.'s (RDHL) strategic acquisition of a 45% share in Health Alliance Group Inc. (HAGI) employing an exploratory and descriptive approach. The methodology aims to employ secondary data sources to examine the acquisition's consequences on strategy, operations, and innovation.

Type of Study

The present study adopts a descriptive and exploratory methodology. Given that RDHL's acquisition of HAGI is a recent development in the area of digital health-

care, the exploratory component is appropriate. The use of descriptive design is particularly important because its purpose is to better understand the acquisition's overall market consequences, technology fit, and strategic goal. An organised understanding of the structures and information at present is made possible by this design. All of these methods are combined to facilitate a more thorough examination of the case's both quantitative and qualitative components.

Data and Data Sources

The study only uses secondary data. These include formal corporate press releases, investor presentations, annual reports, and regulatory filings, all of which offer trustworthy and verified information on the deal. Additionally, news articles and sector reports from verified sources such as *Business Standard*, *Mint*, *CNBC-TV18*, *Medical Buyer*, and *Angel One* are used to supplement and triangulate the data. The selection of these sources is to maintain stability between third-party analysis and the organisation disclosure. For this preliminary strategic evaluation, secondary data is both suitable and economical.

Time Frame of Analysis

This research utilised data from December 2023 to February 2025. The most important phases of the acquisition that are represented throughout this time period include the announcement, public disclosure, stakeholder reactions, and short-term developments. By analysing this period of time, the researcher can investigate the impacts on strategic direction, stock performance, and market perception that occur instantly and in the initial phases after an acquisition.

Tools and Techniques of Analysis

In the present research, both qualitative and quantitative methods are employed. The overall strategic direction of RDHL and the contribution of the acquisition to the expansion of its healthcare reach can be visualised via trend mapping. The integrated use of these tools offers multiple perspectives on the potential outcomes of the partnership.

Validity and Reliability

The analysis has depended entirely on validated and confirmed sources to ensure their accuracy. Announcements on official company websites, regulatory disclosures, and annual reports are all considered highly reliable. To decrease biases and deception, these were compared to unbiased news reports from trustworthy media

sources. The findings' consistency and authenticity are enhanced by this multifaceted strategy. The incorporation of confirmed and publicly available content further supports the research's potent methodology.

Theoretical Framework

Furthermore, the following previous research findings and theories that focus on the dynamics of mergers and acquisitions in the pharmaceutical and healthcare industries are utilised in the analysis of the RDHL-HGAI agreement.

Technical Efficiency: Engaging in merger and acquisition (M&A) activities is frequently viewed as an approach that increases relative technical efficiency, which ultimately results in higher revenue growth. According to Barkema and Schijven (2008) and Salleh et al. (2013), M&As can motivate companies to increase operational efficiency, which will enhance their competitive edge and market performance through the integration of complementary competencies, economies of scale, and better resource allocation.

Innovative Performance: Cloodt et al. (2006) suggest that non-technological M&As frequently have an adverse impact on the acquiring firm's post-M&A innovation when it comes to innovating performance. The size of the acquired knowledge base, however, plays an important role in technical M&As; if it is substantial in comparison to the acquiring company, it may potentially limit the firm's ability to innovate because of problems with integration and potential knowledge overload.

Ethical Conduct: According to Lin and Wei (2006), there is a significant connection between employee job performance and the ethical conduct that occurs during M&As. According to their research, keeping high ethical standards during the M&A process is essential since it has a direct impact on employee trust, morale, and productivity, all of which have an effect on the acquisition's overall performance.

Growth of RDHL

Reliance Digital Health Limited (RDHL), a subsidiary of Reliance Industries Limited (RIL), is strategically positioned as a leading force in the rapidly advancing digital healthcare sector, with a particular focus on underserved and emerging markets. The growing demand for telemedicine, digital diagnostics, and remote healthcare solutions has opened up significant opportunities for RDHL to expand its global reach. Healthcare technology adoption in regions such as India and the U.S. is

projected to grow by more than 20% annually, with underserved areas presenting the greatest potential for market penetration. By aligning its strategy with these trends, RDHL has crafted a comprehensive growth plan that combines organic expansion through technological innovation with inorganic growth via strategic acquisitions, including its recent investment in Health Alliance Group Inc. (HAGI).

RDHL's strategy to target key segments within the healthcare market is clear through its emphasis on telemedicine, diagnostics, remote care, and health management platforms. As illustrated in Table 1, these segments are poised for substantial growth, especially in underserved rural and urban markets.

Table 1.

Market Segments Targeted by RDHL for Digital Healthcare Growth (December 2024)

Industry	Telemedicine	Diagnostics	Remote Care	Health Management Platforms
Rural Healthcare	40%	30%	25%	5%
Urban Underserved	35%	25%	30%	10%
Established Markets	20%	35%	30%	15%
Global Public Health	15%	25%	40%	20%

Source: Compiled and adapted according to Business Standard (2024), Angel One (2024), CNBC-TV18. (2024), and Digital Health Alliance (2024).

RDHL's growth can also be attributed to a series of strategic milestones and initiatives, which highlight its commitment to becoming a global leader in digital healthcare. These include investments in cutting-edge telemedicine platforms, diagnostics networks, and global partnerships to address critical healthcare gaps.

The growth strategy of RDHL is a complementary approach, with its organic expansion supported by its inorganic growth through acquisitions. For example, the development of telemedicine platforms expands RDHL's reach in rural regions, and the acquisition of HAGI offers it access to cutting-edge digital health technologies that enhance its telemedicine and diagnostics capabilities.

RDHL has the potential to become a global leader in digital healthcare through its combined efforts in technological advancement and strategic acquisitions. The growth promotes Reliance Industries' overall goal for employing technology to encourage social impact and economic growth in further improving RDHL's position in the health care sector. Globally, RDHL is improving health equity and reducing

healthcare gaps due to its strong market focus, innovative collaborations, and solid infrastructure.

Table 2.

Key Milestones in RDHL's Growth Trajectory

Year	Initiative	Objective	Outcome
2024	Acquisition of HAGI (45% Stake)	Leverage HAGI's expertise in telemedicine and diagnostics to expand globally	Entry into underserved regions, targeting India, the U.S., and Africa
2023	Launch of Telemedicine Platforms	Address rural healthcare challenges in India	Delivered over 1 million consultations in the first year
2023–24	Expansion of Digital Diagnostics Network	Establish diagnostics hubs in Tier 2 and Tier 3 cities	Improved access to affordable healthcare diagnostics in underserved areas
2022	Partnerships with AI-Based Healthcare Providers	Deliver personalized, data-driven healthcare solutions	Enhanced service offerings in predictive and preventive care
Source: Compiled and adapted according to Business Standard (2024), Angel One (2024), CNBC-TV18. (2024), Economic Times (2024), Medical Buyer (2024), and Mint (2024).			

Results and Discussion

RDHL Deals of the Past

RDHL's success in the digital health care sector might be related to its growth-oriented policies that promote technological advancement and the smooth integration of acquired business assets, like platforms, products, technology, and skills. The reality that RDHL acquired an important part of Health Alliance by Reliance (HAGI) is evidence of its commitment to employing digital health solutions to expand its presence in areas of disadvantage.

This imperative partnership aligns with RDHL's goal of addressing critical gaps in healthcare access, leveraging the abilities of HAGI to enhance its offerings. By in-

vesting in underserved markets, RDHL is setting up the foundation for a broader ecosystem of telemedicine and diagnostics. As of March 2023, Reliance Digital Health Limited had 1,617.19 million equity shares of ₹10 each, with a total value of ₹18.6 billion. Also, as of March 2024, the company kept 1,617.19 million equity shares of ₹10 each, amounting to ₹18.6 billion (Reliance Industries Limited, 2024).

Table 3 presents a comparative analysis of RDHL's significant strategic deals, presenting their terms, objectives, and results. This approach reflects the growth strategies of global technology giants, which have consistently implemented inorganic growth paths to enhance their market share.

Table 3.

Reliance Digital Health's Investment in Health Alliance Group: Deal Overview

Parameters	Health Alliance by Reliance
Value	$10 million
Year	2024
Payment	Equity acquisition (45% stake in HAGI)
Objective	Strategic move to strengthen presence in underserved regions through digital health solutions
Stage of the Deal	Expected to close within two weeks, contingent on customary closing conditions
Non-listed/Listed	Non-listed
What is in the deal for RIL/RDHL	Expanding digital healthcare offerings, leveraging HAGI's technology for underserved regions

Source: Compiled and adapted according to Business Standard (2024), Angel One (2024), CNBC-TV18. (2024), Economic Times (2024), Medical Buyer (2024), Mint (2024) & Digital Health Alliance (2024).

Reliance Digital Health Ltd (RDHL), a subsidiary of Reliance Industries Ltd, has made a strategic move by acquiring a 45% equity stake in Health Alliance Group Inc. (HAGI) for $10 million (Economic Times, 2024; Angel One, 2024; Medical Buyer, 2024; Mint, 2024; Business Standard, 2024). This acquisition underscores RDHL's commitment to expanding its digital healthcare footprint, with a focus on underserved regions by leveraging HAGI's innovative healthcare technology solutions. HAGI, incorporated in Delaware in December 2023, specializes in addressing critical gaps in healthcare access globally, particularly in the United States and India. The transaction, expected to close within two weeks, subject to customary

conditions, aligns with RDHL's vision of establishing a robust ecosystem for telemedicine and diagnostics (Digital Health Alliance, 2024). This investment represents RDHL's ongoing efforts to bridge healthcare disparities through targeted technology-driven initiatives.

The Deal to Acquire Health Alliance

Reliance Digital Health Ltd (RDHL), a subsidiary of Reliance Industries Ltd, has entered into a strategic contract to acquire a 45% equity stake in Health Alliance Group Inc. (HAGI) for $10 million. This acquisition is part of RDHL's strategy to increase its digital healthcare contributions, particularly in underserved areas (Mint, 2024; Business Standard, 2024; Money Control Hindi, 2024). HAGI, established in Delaware in 2023, emphasizes providing technology-driven healthcare solutions to underserved populations in the United States, India, and other worldwide markets. The partnership is aimed at enhancing RDHL's telemedicine and diagnostic capabilities, contributing to a broader ecosystem for digital healthcare (Angel One, 2024; Digital Health Alliance, 2024). The deal was likely to close within two weeks, contingent on customary closing conditions (Medical Buyer, 2024). This deal was a show of RDHL's commitment to fill critical gaps in healthcare access and leveraging technology to improve healthcare delivery (Economic Times, 2024; Reuters, 2024).

Why RDHL Acquires a Stake in Health Alliance

Reliance Digital Health Ltd. (RDHL), a subsidiary of Reliance Industries Ltd., acquired a 45% stake in Health Alliance Group Inc. (HAGI) for $10 million, with strategic aims. The major motives behind this deal, as stated by experts, professionals, and social media, are as follows:

Expansion into Underserved Markets: RDHL, which already has a significant presence in India, aims to further expand its footprint in underserved healthcare markets, both within and outside India. It could also reach developing markets by way of creating networks of similar healthcare providers. HAGI's technology-driven healthcare solutions, which are in line with filling technological gaps of RDHL, are in sync with RDHL's goals (Mint, 2024; Economic Times, 2024). RDHL is confident that this acquisition will enable it to rapidly scale up its offerings and provide affordable, quality healthcare in regions where access to state healthcare services is limited.

Technological Synergy: RDHL seeks to leverage HAGI's advanced healthcare technology to enhance its telemedicine and diagnostic platforms. The integration

of HAGI's solutions into RDHL's existing infrastructure is expected to improve service delivery, streamline healthcare access, and drive innovation within RDHL's digital healthcare ecosystem (Business Standard, 2024; Digital Health Alliance, 2024). The combination of RDHL's resources and HAGI's cutting-edge solutions will enable both companies to offer integrated, tech-enabled healthcare services more effectively. However, it depends on the quality of integration of different layers, including the integration of abilities and skills of manpower.

Building a Global Healthcare Ecosystem: By acquiring HAGI, RDHL is positioning itself as a leader in the digital healthcare sector. This strategic acquisition will help RDHL build a global healthcare ecosystem that provides telemedicine, diagnostics, and other healthcare services across borders. RDHL's investment in HAGI complements its vision of creating a seamless, integrated global healthcare delivery network that bridges gaps in care for underserved populations (Angel One, 2024).

Diversification of Healthcare Portfolio: With the growing demand for digital healthcare solutions, RDHL's acquisition of HAGI will diversify its healthcare offerings. By broadening its portfolio to include HAGI's technology-driven solutions, RDHL aims to strengthen its competitive position in the fast-evolving digital health market (Medical Buyer, 2024). This expansion is seen as crucial to RDHL's strategy of staying ahead of emerging healthcare trends and tapping into new business opportunities.

Strategic Long-term Growth: In the opinion of RDHL, this acquisition is in line with its broader strategy of inorganic as well as organic growth, parallel to the approach followed by global technology companies. Similar to the strategies of large companies like Microsoft, Google, and Verizon, RDHL realizes that HAGI's innovative platform is a means to accelerate its growth path in the digital healthcare space (Digital Health Alliance, 2024). By making an investment in HAGI, RDHL goals to not only expand its market reach but also develop more capabilities and an expanded business model in the healthcare industry.

Impact of the deal on RDHL and its Operations

The acquisition of a 45% stake in Health Alliance Group Inc. (HAGI) by Reliance Digital Health Ltd (RDHL) is expected to have a significant influence on RDHL's operations, aligning it as a stronger player in the global digital healthcare market. This acquisition is in line with RDHL's long-term strategy of expanding its presence in underserved healthcare markets and improving its digital health solutions. The impact of this contract can be analysed through several important magnitudes, in-

cluding operational synergies, market expansion, technological enhancement, and long-term growth potential.

Operational Synergies and Integration

The acquisition is well-positioned to create substantial operational synergies between RDHL and HAGI, primarily in the combination of HAGI's innovative healthcare technologies with RDHL's existing infrastructure. By merging their resources and expertise, RDHL can rationalize healthcare delivery, increase operational efficiency, and leverage HAGI's platform to enhance service availability. As known by experts, the acquisition will increase RDHL's telemedicine and diagnostic services, which are central to its business model (Business Standard, 2024). These synergies are expected to accelerate RDHL's ability to provide scalable, tech-driven healthcare solutions across different geographies, especially in regions where access to quality healthcare is limited (Digital Health Alliance, 2024).

This quote supports the finding that such integration will streamline healthcare delivery and increase RDHL's efficiency in underserved regions.

Market Expansion and Geographic Reach

The strategic shift to acquire a stake in HAGI will significantly increase RDHL's market expansion abilities, primarily in the United States and other universal markets. By Leveraging into HAGI's developing networks and customer base, RDHL can boost its entry into these markets and expand its influence in the digital healthcare sector. RDHL aims to leverage HAGI's expertise in expressing healthcare access gaps, which aligns with its broader objective of enhancing healthcare accessibility in underserved regions (Mint, (2024). This environmental diversification is likely to stabilize RDHL's competitive positioning, supporting it to create a healthier presence in the global healthcare market. MSN (2024) and Business Upturn (2024) emphasized that this deal is part of RDHL's broader effort to enhance its global reach by leveraging innovative telemedicine and diagnostics platforms. As reported by Business Upturn, this move enables RDHL to align with the fast digitization of healthcare and target geographic regions where healthcare access remnants a challenge, thus strengthening its occurrence in both established and emerging markets.

This reinforces RDHL's strategic intent to grow its geographic footprint through partnerships with regionally established players.

Technological Enhancement and Innovation

Technological integration is another crucial driver of this acquisition's impact on RDHL. HAGI's advanced healthcare technology solutions, including telemedicine and diagnostics, are expected to complement RDHL's available offerings and increase its digital health ecosystem (Angel One, 2024). The acquisition will enable RDHL to integrate cutting-edge technology into its operations, making it quicker and more responsive to changing healthcare demands. This technological development aligns with RDHL's vision of becoming a leader in digital healthcare, permitting it to develop innovative solutions that present gaps in healthcare access and improve patient results (Medical Buyer, 2024).

This reflects how RDHL is enhancing its innovation capability and patient service standards through this deal.

Long-term Growth and Strategic Positioning

After a strategic perception, this acquisition strengthens RDHL's position in the digital healthcare sector, providing it with a platform for long-term growth. Parallel to other global technology companies, RDHL has followed a non-organic growth strategy by acquiring strategic assets, like HAGI, to expand its portfolio and increase its capabilities (Economic Times, 2024). This deal is expected to not only improve RDHL's digital healthcare contributions but also help it stay ahead of competitors in the fast-evolving market. With the combination of HAGI's solutions, RDHL is well-positioned to capitalize on the rising demand for digital health services, which are expected to play a progressively important role in global healthcare delivery in the coming years (Business Standard, 2024).

This finding indicates that RDHL is setting itself up for long-term positioning in the global healthcare value chain.

Stock Market Response to the RDHL-HAGI Acquisition: Impact on RIL's Share Price

A significant impact of this acquisition has been felt in the market response to Reliance Industries Ltd (RIL), RDHL's parent company. Subsequent to the announcement, RIL's share price experienced a rise of over 1%, signifying strong investor confidence. Table 4 demonstrates the variations in RIL's share price together with market directories on NSE and BSE during this period:

Table 4.

Stock Market Reaction to RDHL-HAGI Acquisition: Impact on RIL's Share Performance

Date	Share Price	Market Index on NSE	Market Index on BSE
Dec 03, 2024	1,323.30	24,367.50	80,529.20
Dec 04, 2024	1,308.95	24,488.75	81,036.22
Dec 05, 2024	1,322.05	24,539.15	81,182.74
Dec 06, 2024	1,311.55	24,729.45	81,575.96
Dec 09, 2024	1,295.15	24,633.90	81,602.58
Dec 10, 2024	1,284.85	24,652.65	81,575.96
Dec 11, 2024	1,278.20	24,620.50	81,568.39
Dec 12, 2024	1,262.90	24,604.45	81,476.76
Dec 13, 2024	1,272.85	24,498.35	81,212.45
Dec 16, 2024	1,268.30	24,753.40	82,000.31
Dec 17, 2024	1,245.30	24,584.80	81,511.81
Dec 18, 2024	1,253.25	24,297.95	80,666.26
Dec 19, 2024	1,230.45	23,877.15	79,029.03
Dec 20, 2024	1,205.30	23,960.70	79,335.48
Dec 23, 2024	1,222.30	23,738.20	78,488.64
Dec 24, 2024	1,222.75	23,769.10	78,707.37
Dec 26, 2024	1,216.55	23,775.80	78,557.28
Dec 27, 2024	1,221.05	23,801.40	78,607.62
Dec 30, 2024	1,210.70	23,796.90	78,637.58
Dec 31, 2024	1,215.45	23,560.60	77,982.57

Source: Compiled and adapted according to Investing.com, (2024), National Stock Exchange of India (NSE), (2024). Bombay Stock Exchange (BSE), (2024).

The stock market response to the RDHL-HAGI acquisition, especially after the stake acquisition on December 20, 2024, showed mixed reactions. Initially, following the acquisition announcement, RIL's share price increased to ₹1,323.30 on December 3, 2024. However, after December 6, 2024, the stock began to decline, reaching ₹1,230.45 by December 19, 2024. On December 20, 2024, when the stake in RDHL was acquired, RIL's share price continued to fall, dropping to ₹1,205.30. This suggests that despite the positive initial reaction, the acquisition didn't lead to a significant boost in RIL's stock price. Instead, investor sentiment appeared more

cautious, possibly due to broader market factors or profit-taking. As of December 31, 2024, RIL's share price was ₹1,215.45, reflecting a slight recovery post-acquisition. The market Index on NSE and BSE also tracked a similar pattern, rising on December 16, 2024, before progressively decreasing along with RIL's stock price. Although the announcement and the acquisition set off some initial optimism, the decline in RIL's share price suggests that investors become more cautious over time, possibly reviewing the long-term effects of the deal. This shift in sentiment affects how market reactions to acquisitions can be complex and influenced by several factors.

Analysis of Data & Interpretation With Respect to Research Questions

Based on the event analysis, content analysis, and trend analysis presented in earlier sections, the possible answers to the 10 research questions are summarized as follows.

Research Question 1: What are the strategic advantages of RDHL acquiring a 45% stake in HAGI?

Digital Health Alliance (2024) said the acquisition allows RDHL to leverage HAGI's technological innovations in digital healthcare, enhancing its telemedicine and diagnostics services, and improving healthcare access in underserved regions (Finance Saathi, 2024; M&A Critique, 2024; Navbharat Times, 2024; ICaps Sri Lanka, 2024; Digital Health Alliance, 2024). This shows clearly how RDHL benefits from HAGI's innovative platforms both in terms of technological depth and market positioning, particularly in the areas of diagnostics and AI-based telehealth solutions. The partnership allows scalable healthcare delivery by increasing RDHL's product capability.

Research Question 2: Can this partnership help RDHL address healthcare delivery challenges in underserved global regions?

Yes, the partnership is aimed at addressing healthcare delivery challenges by utilizing HAGI's expertise to expand RDHL's reach in underserved markets, thereby improving healthcare accessibility and outcomes (Business Standard, 2024; Republic Biz, 2024; Digital Health Alliance, 2024; India.com, 2024). The agreement gives RDHL the ability to provide affordable digital healthcare services in inaccessible regions, particularly by utilising HAGI's established preventive care methods. These address barriers based on geography and income.

Research Question 3: How will the integration of HAGI's technological innovations influence RDHL's current healthcare services portfolio?

The integration is expected to enhance RDHL's existing services, particularly in digital diagnostics and telehealth, allowing for more comprehensive healthcare solutions and improving RDHL's competitive position (Mint, 2024; Business Standard, 2024; Digital Health Alliance, 2024; Money Control Hindi, 2024). This reveals how RDHL's present offerings will not only increase but also become more complex and efficient. The use of artificial intelligence-powered diagnostics will improve clinical accuracy and service delivery speed.

Research Question 4: Does this acquisition align with RDHL's long-term ambition to become a global leader in digital healthcare solutions?

Yes, the acquisition aligns with RDHL's long-term vision by expanding its service portfolio and reinforcing its commitment to innovation in the global digital healthcare landscape (Medical Buyer, 2024; CNBC, 2024; Business Standard, 2024; Sarthe, A. 2024). This shows how the acquisition can serve as an introduction to international markets and the future leadership in the sector.

Research Question 5: How does this transaction contribute to RIL's overarching vision of leveraging technological innovation to drive growth?

The transaction supports RIL's vision of growth through technological innovation by enhancing RDHL's digital healthcare capabilities, allowing it to tap into the growing digital health market (Economic Times 2024; Angel One 2024; Chola Securities, 2024; Medical Buyer 2024; Mint 2024). This explains how RDHL enhances RIL's wider journey into technology-related ventures.

Research Question 6: Will this deal strengthen RDHL's competitive position in the global digital health sector?

The acquisition is likely to strengthen RDHL's competitive position by expanding its technological capabilities and market reach, enabling it to compete more effectively in the digital health sector (Business Standard, 2024; Digital Health Alliance, 2024; Business Standard, 2024). This proves RDHL can now set itself apart in international healthcare technology marketplaces.

Research Question 7: What risks might arise from this acquisition, including integration hurdles and competitive pressures in the digital health space?

The potential risks include integration challenges and competitive pressures from other digital health providers, as well as the need for RDHL to effectively leverage HAGI's technologies (Medical Buyer, 2024; Business Standard, 2024; CNBC, 2024). It also implies that a seamless system and human capital integration are crucial for post-deal success.

Research Question 8: How will this collaboration affect RDHL's operational efficiency and market expansion strategies?

The collaboration is expected to enhance operational efficiency by streamlining processes and integrating advanced technologies, facilitating market expansion strategies in underserved regions. (Business Standard, 2024; The Startups News, 2024; Angel One, 2024). This represents how RDHL can now effectively enter new markets, generating more revenue.

Research Question 9: To what extent can this partnership improve healthcare accessibility in India and other underserved markets?

The partnership has the potential to significantly improve healthcare accessibility by providing affordable and scalable healthcare solutions in India and other underserved markets (Medical Buyer, 2024; Economic Times, 2024; CNBC, 2024). The observation indicates actual desires to use HAGI's scalable technology to reach nationally.

Research Question 10: What implications does this partnership have for RDHL's global healthcare strategy?

This partnership is an important component of RDHL's global healthcare strategy, allowing it to expand its mark in the digital health sector and address healthcare challenges on a global scale (Startups News, 2024; Medical Buyer 2024; CNBC 2024). This identifies that the acquisition of HAGI is necessary for RDHL's strategy of international expansion.

Conclusion

Reliance Digital Health Ltd.'s (RDHL) strategic acquisition of a 45% share in Health Alliance Group Inc. (HAGI) represents an influential turning point in RDHL's growth as a major player in the global digital healthcare market. The agreement is more than simply a financial investment; it is an innovative initiative meant to close important gaps in healthcare access, particularly in impoverished parts of India, the US, and developing countries. Through incorporating HAGI's inno-

vative diagnostic and telemedicine healthcare solutions, powered by artificial intelligence, RDHL reiterated its dedication to building an innovative, inclusive, and scalable healthcare ecosystem.

The agreement is in perfect alignment with Reliance Industries Limited's (RIL) overarching mission of employing technology innovation to encourage social impact and sustainable growth. It improves RDHL's operational effectiveness and puts the business in a stronger position to address the growing demand for digital healthcare solutions around the globe. In addition, through offering unique services and enhancing accessibility, affordability, and care quality, the acquisition helps RDHL to improve its competitive position and support goals related to public health.

The improvement in RIL's share price as a result of the initial investor reaction shows that the market is highly confident in the strategic intent and possibilities of this move. With the development of digital diagnostics and virtual care platforms that get beyond geographical and infrastructural hurdles, the alliance is anticipated to make it easier for RDHL to reach health-deficient areas.

Nevertheless, there are some challenges with the acquisition. Difficulty of integration, problems with cultural alignment, and increasing competition from both existing and developing digital health providers are potential threats. To fully reap the rewards of this partnership, RDHL needs to overcome these barriers with thorough strategic planning, strong change management, and a patient-centric perspective.

In conclusion, the RDHL–HAGI acquisition, in brief, is a courageous and appropriate move of strategy that not only enhances RDHL's business strengths but also indicates its wider ambition to redefine healthcare delivery on a global scale. By using modern technology, providing underprivileged people with preference, and promoting RIL's purpose of innovation-led impact, RDHL is well-positioned to change digital healthcare in India as well as globally. Future studies should look at the partnership's long-term operational results, post-integration effectiveness, and broader impact on healthcare accessibility in low-resource environments.

References

Angel One. (2024). Cloodt, *December 23*. Retrieved on December 30, 2024, from https://www.angelone.in/news/ril-subsidiary-to-acquire-stake-in-health-alliance-group

Aoi Ventures. (2024). *Reliance Digital Health's acquisition announcement, December 21*. Retrieved on December 23, 2024, from https://x.com/aoiventures/status/1870924935599972631

Barkema, H. G., & Schijven, M. (2008). Toward unlocking the full potential of acquisitions: The role of organizational restructuring. *Academy of Management Journal*, *51*(4), 696-722. https://doi.org/10.5465/amj.2008.33665279

Bizz Buzz. (2024). Reliance Digital acquires 45% stake in Health Alliance Group for $10 million. *Bizz Buzz, December 21*. Retrieved December 25, 2024, from https://www.bizzbuzz.news/markets/reliance-digital-acquires-45-stake-in-health-alliance-group-for-10-million-1346350

Bombay Stock Exchange (BSE). (2024). *Index archive data*. Retrieved on February 5, 2025, from https://www.bseindia.com/Indices/IndexArchiveData.html

Business Standard. (2024). *Reliance Industries shares gain after subsidiary acquires stake in HAGI, December 23*. Retrieved on December 30, 2024, from https://www.business-standard.com/markets/news/reliance-industries-shares-surge-after-subsidiary-acquires-stake-in-hagi-124122300372_1.html

Business Standard. (2024). Reliance Industries shares surge after subsidiary acquires stake in HAGI, December 23. Retrieved on February 4, 2025, from https://www.business-standard.com/markets/news/reliance-industries-shares-surge-after-subsidiary-acquires-stake-in-hagi-124122300372_1.html

Business Standard. (2024, December 23). Reliance Industries shares gain after subsidiary acquires stake in HAGI, *December 23*. Retrieved on February 7, 2025, from https://www.business-standard.com/markets/news/reliance-industries-shares-surge-after-subsidiary-acquires-stake-in-hagi-124122300372_1.html

Business Upturn. (2024). Reliance Digital Health to acquire 45% stake in US-based Health Alliance Group for Rs 84 crore, *December 21*. Retrieved December 23, 2024, from https://www.businessupturn.com/business/corporates/reliance-digital-health-to-acquire-45-stake-in-us-based-health-alliance-group-for-rs-84-crore/.

Capital Market. (2024). *Reliance Digital Health to acquire 45% stake in Health Alliance Group Inc, December 21*. Retrieved on December 23, 2024, from https://www.capitalmarket.com/markets/news/corporate-news/Reliance-Digital-Health-to-acquire-45percent-stake-in-Health-Alliance-Group-Inc-/1568458

Chola Securities. (2024). *Reliance Industries arm to acquire 45% stake in Health Alliance Group, December 21*. Retrieved on February 5, 2025, from https://www.cholasecurities.com/news/hot-pursuit/reliance-inds-arm-to-acquire-45-stake-in-health-alliance-group/1568466

Cloodt, M., Hagedoorn, J., & Van Kranenburg, H. (2006). Mergers and acquisitions: Their effect on the innovative performance of companies in high-tech industries. *Research policy*, *35*(5), 642-654.

CNBC-TV18. (2024). Reliance Digital to acquire 45% stake in Health Alliance Group for $10 million, December 25. Retrieved on February 3, 2025, from https://www.cnbctv18.com/market/stocks/reliance-industries-arm-reliance-digital-to-acquire-45-stake-in-health-alliance-group-for-10-million-19527794.htm

Domain-b. (2024). *Reliance Digital Health to acquire 45% stake in Health Alliance Group for $10 million, December 23*. Retrieved on December 30, 2024, from https://www.domain-b.com/management/m-a/reliance-digital-health-to-acquire-45-stake-in-health-alliance-group-for-10-million

Economic Times. (2024). *Reliance Industries shares in focus after 45% stake acquisition in Health Alliance, December 23*. Retrieved on December 28, 2024, from https://economictimes.indiatimes.com/markets/stocks/news/reliance-industries-shares-in-focus-after-45-stake-acquisition-in-health-alliance-for-10-million/articleshow/116582735.cms

Economic Times. (2024). Reliance Industries shares in focus after HAGI acquisition, December 23. Retrieved on February 5, 2025, from https://economictimes.indiatimes.com/markets/stocks/news/reliance-industries-shares-in-focus-after-45-stake-acquisition-in-health-alliance-for-10-million/articleshow/116582735.cms

Finance Saathi. (2024). Reliance Digital Health to acquire 45% stake in Health Alliance Group, *December 21*. Retrieved on December 23, 2024, from https://financesaathi.com/news_details/reliance-digital-health-to-acquire-45-stake-in-health-alliance-group-for-10-million

Frost & Sullivan. (2022). Global telehealth growth opportunities, 2020–2026, June 30. Retrieved on February 8, 2025, from https://store.frost.com/global-telehealth-growth-opportunities.html

ICaps Sri Lanka. (2024). Reliance Industries arm to acquire 45% stake in Health Alliance Group for $10 million, *December 23*. Retrieved on December 23, 2024, from https://icapsrilanka.com/reliance-industries-arm-to-acquire-45-stake-in-health-alliance-group-for-10-million-company-business-news/

India.com. (2024). Mukesh Ambani makes big move: Buys 45% stake in Health Alliance Group Inc. (HAGI) in a $10 million deal, *December 21*. Retrieved on February 25, 2025, from https://www.india.com/business/mukesh-ambani-makes-big-move-buys-45-stake-in-health-alliance-group-inc-hagi-deal-is-worth-10-million-usd-reliance-industries-reliance-digital-health-7483450/

Investing.com. (2024). *Reliance Industries Ltd. stock overview*. Retrieved on February 5, 2025, from https://in.investing.com/equities/reliance-industries

Lin, C. Y. Y., & Wei, Y. C. (2006). The role of business ethics in merger and acquisition success: An empirical study. *Journal of Business Ethics*, *69*, 95-109.

M&A Critique. (2024). Reliance Industries arm to acquire 45% stake in Health Alliance Group, *December 23*. Retrieved on December 30, 2024, from https://mnacritique.mergersindia.com/news/reliance-inds-arm-to-acquire-45-stake-in-health-alliance-group/

McKinsey & Company. (2020). The future of healthcare: A digital revolution. Retrieved on February 5, 2025, from https://www.mckinsey.com/industries/healthcare/our-insights/the-future-of-healthcare

Medical Buyer. (2024). *Reliance Digital Health to acquire 45% stake in Health Alliance Group Inc, December 23*. Retrieved on December 26, 2024, from https://medicalbuyer.co.in/reliance-digital-health-to-acquire-45-stake-in-hagi-for-usd-10m/

Mint. (2024). *Reliance Industries' healthcare investments to boost digital health reach globally, December 23*. Retrieved on December 30, 2024, from https://www.livemint.com/market/stock-market-news/reliance-industries-healthcare-investments-boost-digital-health-portfolio

Mint. (2024). Reliance's healthcare investments to boost digital health globally, December 23. Retrieved on February 12, 2025, from https://www.livemint.com/companies/news/reliance-industries-to-acquire-45-stake-in-health-alliance-group-for-10-million-11734780139890.html

Moneycontrol. (2024, December 20). Reliance Digital Health Health Alliance Group 45% , ₹85 . Retrieved on February 12, 2025, from https://hindi.moneycontrol.com/news/markets/reliance-industries-subsidiary-reliance-digital-health-buys-a-45-percent-stake-in-health-alliance-group-1945332.html

MSN. (2024). Reliance Industries arm to acquire 45% stake in Health Alliance Group for $10 million, *December 23*. Retrieved January 28, 202, from https://www.msn.com/en-in/money/general/reliance-industries-arm-to-acquire-45-stake-in-health-alliance-group-for-10-million/ar-AA1wgTTj.

National Stock Exchange of India (NSE). (2024). *Reports and indices historical index data*. Retrieved on February 5, 2025, from https://www.nseindia.com/reports-indices-historical-index-data

Navbharat Times. (2024). *Reliance arm to acquire 45% stake in Health Alliance Group: See details, December 21*. Retrieved on February 5, 2025, from https://navbharattimes.indiatimes.com/business/business-news/reliance-arm-to-acquire-45-percent-stake-in-health-alliance-group-see-detail/articleshow/116541512.cms

Pragativadi. (2024). *Reliance Industries' arm to acquire 45% stake in Health Alliance Group for $10 million, December 23.* Retrieved on February 5, 2025, from https://pragativadi.com/reliance-industries-arm-to-acquire-45-stake-in-health-alliance-group-for-10-million/

Reliance Industries Limited (2024). *Integrated annual report 2023–24*. https://www.ril.com/ar2023-24/pdf/RIL_IAR_2024.pdf

Republic Biz. (2024). *Reliance Industries Limited acquisition update: Subsidiary of Mukesh Ambani's company to take 45% stake in details, December 23.* Retrieved on February 5, 2025, from https://www.republicbiz.com/companies/reliance-industries-limited-acquisition-update-subsidiary-of-mukesh-ambani-s-company-to-take-45-stake-in-details

Reuters. (2024). *Reliance Industries announces major healthcare expansion with acquisition of Health Alliance Group stake, December 23*. Retrieved on December 23, 2024, from https://www.reuters.com/markets/reliance-industries-announces-major-healthcare-expansion

Salleh, W. A., Wan Mahmood, W. M., Sufian, F., Mohd Jamarudi, E. M., Sekharan Nair, G. K., & Ahmad, S. (2013). The efficiency of mergers and acquisitions in Malaysia based telecommunication companies. *Asian Social Science*, *9*(2), 12-23.

Sarthe, A. (2024). *Reliance Digital Health to acquire 45% stake in Health Alliance Group, December 23.* Retrieved on December 23, 2024, from https://www.digitalhealthnews.com/reliance-digital-health-to-acquire-45-stake-in-health-alliance-group

Singh, N. P. (2014a). Microsoft acquired Nokia in unipolar operating system market. *Independent Journal of Management & Production*, *5*(3), 598-622.

Singh, N. P. (2014b). Verizon acquired Vodafone-Analysis of Market Reaction. *Industrija*, *42*(3).

Singh, N. P. (2015a). Google bought and sold Motorola Mobility-What it means. *International Journal of Business Policy and Strategy Management*, *1*, 13-24.

Singh, N. P. (2015b). Lenovo Acquires IBM's x86 Low-end Server Business. *Industrija*, *43*(3).

Singh, N. P. (2017). Lenovo Acquired Motorola Mobility in a Bipolar World of Mobile phone makers. *Industrija*, *45*(2).

Singh, N. P. (2019). Oracle acquired AI startup DataFox. *Industrija*, *47*(2).

The Startups News. (2024). *Reliance Industries' arm Reliance Digital to acquire 45% stake in Health Alliance Group for $10 million, December 24.* Retrieved on February 5, 2025, from https://thestartupsnews.com/reliance-industries-arm-reliance-digital-to-acquire-45-stake-in-health-alliance-group-for-10-million/

World Health Organization (WHO). (2021). *Global strategy on digital health 2020–2025*. https://apps.who.int/iris/handle/10665/344249

World Health Organization (WHO). (2021). Global strategy on digital health 2020–2025. Retrieved on February 12, 2025, from https://apps.who.int/iris/handle/10665/344249

INTERNATIONAL JOURNAL OF TECHNICAL AND NATURAL SCIENCES

DOI: https://doi.org//10.69648/NDDD5147

International Journal of Natural and Technical Sciences (IJTNS), 2025; 5(1): 25-41.

ijtns.ibupress.com

Online ISSN: 2671-3519

Application :01.06.2025

Revision :15.06.2025

Acceptance :13.06.2025

Publication :15.06.2025

Dina, K., Velkovski, V., & Samet, A. (2025). Determination of Hematological and Biochemical Parameters in Blood Samples from Oncological Patients in Kosovo. International Journal of Natural and Technical Sciences, 5(1), 25-41.

https://doi.org//10.69648/NDDD5147

Kushtrim Dina[1], Valentina Velkovski[2], Aksu Samet[3]

[1] Faculty of Technology and Technical Sciences, Veles, University "St. Kliment Ohridski" Bitola, Republic of North Macedonia, https://orcid.org/0009-0000-3515-8324

[2] Faculty of Technology and Technical Sciences, Veles, University "St. Kliment Ohridski" Bitola, Republic of North Macedonia, https://orcid.org/0000-0003-4978-5710

[3] Faculty of Dental Medicine, International Balkan University, Skopje, North Macedonia, https://orcid.org/0000-0001-7922-5058

Correspondence concerning this article should be addressed to Valentina Velkovski.

Email: valentina.velkovski@uklo.edu.mk

BALKAN JOURNALS

Determination of Hematological and Biochemical Parameters in Blood Samples from Oncological Patients in Kosovo

Kushtrim Dina, Valentina Velkovski, Aksu Samet

Abstract

Cancer is a complex disease characterized by the uncontrolled proliferation of abnormal cells, which possess the potential to invade surrounding tissues and metastasize to distant organs. Its development is influenced by a variety of factors, including genetic predisposition, environmental exposures, and individual biological characteristics. The aim of this research is to analyze the blood of 100 patients diagnosed with cancer seeking care at the University Clinical Center of Kosovo for Cancer Care, Oncology Clinic. The patients range in age from 29 to 85 years, with half being men and half being women. The blood parameters survey was conducted between January and June 2024 and was voluntary. Blood samples were tested for a total of ten hematological and biochemical parameters: hemoglobin, erythrocytes, platelets, leukocytes, iron, sedimentation, AST, ALT, urea, and creatinine. Statistical analysis of the data was conducted using the Statistical Package for Social Sciences (SPSS). Non-parametric tests were employed, including the median equality test based on the Chi-square distribution and Spearman's rank correlation coefficient. There is a deviation from the reference values of the measured parameters in sedimentation, where the average measured value for all patients is above the upper reference limit. Namely, the sedimentation in cancer patients covered by the analysis ranges between 5 mm/h and 40 mm/h, while the average is 14.3 mm/h. Differences in median values for almost all blood parameters are also observed between patients who have metastasized disease and those in whom the disease has not metastasized. Nutritional status is a dynamic rather than a static condition. Consequently, it should be regularly evaluated throughout the various stages of the treatment process to ensure consistent and timely access to appropriate nutritional interventions.

Keywords: Cancer, Analysis, Blood, Hematology, Biochemistry

INTRODUCTION

Impact of Chemotherapy on Hematological and Biochemical Profiles in Cancer Patients

Cancer is a leading cause of death globally, with complex causes including genetic, environmental, and structural factors (Sobin & Wittekind, 2009; Lodish et al., 2000). Treatment options such as surgery, chemotherapy, and radiotherapy vary by cancer type, stage, and location (Shahid, 2016). However, chemotherapy often causes significant side effects that may affect treatment success (Sheikh, 2014). Chemotherapy disrupts cell growth by damaging DNA, leading to declines in hemoglobin, platelets, and white blood cells, thus weakening the immune system and increasing infection risk (Devi et al., 2006; Ramasamy et al., 2016). It also affects liver and kidney function. Many chemotherapeutic drugs are metabolized in the liver, contributing to hepatocellular damage and elevated liver enzymes such as ALT, AST, ALP, and LDH (Steele & Narendran, 2012; Uo, 2006). Kidney toxicity may present as increased urea and creatinine due to tubular cell damage (Moore, 2016). Close monitoring of hematological and biochemical profiles is essential to managing side effects and improving outcomes in cancer patients receiving chemotherapy.

The Role of Nutrition in Cancer Care

Cancer is a complex disease requiring integrated, multimodal treatment approaches. Among supportive therapies, nutrition plays a critical role in improving treatment outcomes and quality of life. Evidence shows that professional nutritional intervention enhances nutrient intake, physical function, and the overall prognosis by preventing and managing cancer-related malnutrition (Caccialanza et al., 2016). Malnutrition is a common yet often overlooked complication in cancer care, affecting up to 87% of patients, with 15–40% experiencing weight loss at diagnosis (Wigmore et al., 1997). Despite its impact, there is no universally accepted gold standard for diagnosing malnutrition in cancer. Inadequate nutritional support is linked to reduced treatment options, longer hospital stays, increased toxicity, and lower treatment response (van Cutsem et al., 2005). Alarmingly, up to 20% of cancer-related deaths may result from malnutrition rather than the cancer itself (Muscaritoli et al., 2021). Improving access to nutritional care should be prioritized across healthcare systems to enhance cancer treatment effectiveness and patient outcomes.

LITERATURE REVIEW

Hematological and Biochemical Markers in Breast Cancer Progression

Numerous hematological and biochemical parameters—such as hemoglobin, total leukocyte count, differential leukocyte count, random blood sugar (RBS), liver enzymes (SGPT, SGOT), bilirubin, ALP, total protein, creatinine, and urea—have been studied in breast cancer patients (Shreya et al., 2023). Among these, blood sugar, urea, hemoglobin, and alkaline phosphatase (ALP) showed significant variations between early- and late-stage breast cancer. For instance, elevated RBS levels (normal: 80–140 mg/dL) were observed in 20 out of 56 patients, with 85% in late-stage and 15% in early-stage cancer. High RBS is linked to increased tumor grade, potentially due to enhanced cell proliferation driven by hyperglycemia and hyperinsulinemia (Shander et al., 2004). ALP, a marker of liver and bone activity (normal: 40–125 mg/dL), was elevated in 31 patients, with 59% in advanced-stage disease. This supports ALP as a potential indicator of metastasis, particularly in bone and liver. Anemia was present in 30 patients, with 84% of cases in late-stage breast cancer, suggesting disease progression correlates with declining hemoglobin levels. Likewise, elevated urea levels (normal: 20–40 mg/dL) were found in 70% of late-stage and 29.6% of early-stage patients, pointing to renal dysfunction or systemic catabolism. Finally, 75% of advanced-stage patients showed elevated total protein, correlating with disease severity and systemic involvement. These findings highlight the prognostic value of routine biochemical markers in staging and monitoring breast cancer progression.

Impact of Chemotherapy on Hematological and Biochemical Profiles in Cancer Patients

In a retrospective cohort study by Wondimneh et al. (2019), involving 376 cancer patients at Ayder Comprehensive Specialist Hospital, Ethiopia, significant changes in hematological and biochemical parameters were observed pre- and post-chemotherapy. The study included 228 women and 147 men. Hematological findings showed statistically significant decreases in white blood cells (WBC), red blood cells (RBC), hemoglobin (Hb), hematocrit (HCT), platelets (PLT), and neutrophils (NUT) after chemotherapy ($p < 0.05$), except for lymphocytes, which did not show a significant change. The WBC decrease was attributed to the suppression of hematopoietic stem cells by chemotherapy, increasing susceptibility to infections. Similarly, reduced RBC, Hb, and HCT levels suggested impaired erythropoiesis and possible chemotherapy-induced nephrotoxicity, affecting erythropoietin produc-

tion. Biochemical parameters—including urea, creatinine (CRT), alanine transaminase (ALT), and aspartate transaminase (AST)—showed no statistically significant changes post-treatment ($p > 0.05$), although minor trends were noted. ALT and AST levels showed a slight, non-significant increase, while urea and CRT decreased marginally. The study concludes that chemotherapy significantly affects hematological profiles, contributing to anemia and immunosuppression, while changes in biochemical parameters were minimal. These findings underscore the need for regular monitoring to manage treatment-related toxicities effectively.

The Role of Nutritional Intervention in Cancer Treatment

Nutritional intervention plays a vital role in comprehensive cancer care. Clinical evidence supports its effectiveness in reducing hospitalization duration, minimizing treatment-related toxicity, and enhancing nutrient intake, physical function, and overall quality of life. Importantly, nutritional support can significantly improve treatment outcomes by preventing or mitigating malnutrition—a common and often life-threatening condition among cancer patients. In fact, nearly 25% of cancer-related deaths may be attributed to malnutrition rather than the malignancy itself (Vitaloni et al., 2022). Patients with gastrointestinal cancers are particularly vulnerable due to disease-related digestive tract impairment, yet many lack adequate access to nutritional care. To address this, interdisciplinary and multiprofessional care models are essential. Such models, successfully implemented in a few European countries like the Netherlands and Sweden, ensure timely, effective, and safe nutritional interventions. However, across Europe, these remain exceptions rather than standard practice (Caccialanza et al., 2020).

METHODS AND MATERIALS

The subject research was conducted on a sample of 100 patients diagnosed with cancer seeking care at the University Clinical Center of Kosovo for Cancer Care, Oncology Clinic. All samples were taken after approval from the institutional ethics committee and the personal consent of the patients. Blood samples were collected from patients in heparinized tubes. The patients range in age from 29 to 85 years, with half being men and half being women. The blood parameters survey was conducted between January and June 2024 and was voluntary. The patients included in the research have various demographic and socioeconomic characteristics.

In order to test the set goals, the patients covered by the research were given a survey regarding the use of nutritional supplements. For the purposes of the survey, a survey questionnaire was prepared, which contained open and closed-type questions, as well as multiple-choice questions that allow for the selection of multiple answers. In addition to demographic and socioeconomic questions about the patients, the survey questionnaire also included questions about the use of dietary supplements, changes in dietary habits, including preferences or avoidance of certain foods since the cancer diagnosis, and questions about the disease itself., that is, diagnosis, tumor stage, treatment history, and comorbidities (Dina et al, 2024). The patients diagnosed with cancer also had a blood count, that is, the measurement of ten hematological and biochemical parameters in the blood, shown in Table 1, together with their reference values. Through the analysis of blood samples, the values of: hemoglobin, erythrocytes, platelets, leukocytes, iron, sedimentation, AST, ALT, urea, and creatinine. The erythrocyte sedimentation rate (ESR) was determined using the Westergren method, a simple, cost-effective, and widely utilized laboratory technique. Statistical analysis of the collected data was conducted using the Statistical Package for Social Sciences (SPSS). Non-parametric statistical tests were applied, including the test for equality of medians based on the Chi-square distribution, as well as Spearman's rank correlation coefficient to assess the relationships between variables.

Table 1

Blood parameters measured through blood count of patients.

Parameter	Reference values
Hemoglobin	11.0-16.3 g/dl
Erythrocytes	3.8-5.8 $x10^6 mm^3$
Platelets	150-390 $x10^6 mm^3$
Leukocytes	3.5-10 $x10^6 mm^3$
Iron	60-180 mcg/dl
Sedimentation	3-10 mm/h
ACT	2-37 U/L
ALT	3-41 U/L
Urea	1.7-8.3 mm/l
Creatinine	53-115 mm/l

Understanding the Importance of ESR Levels in Cancer Patients

The erythrocyte sedimentation rate (ESR) is a nonspecific laboratory marker that measures the rate at which erythrocytes (red blood cells) settle at the bottom of a test tube over a specified period. This process is influenced by the density differences among the components of blood, where heavier, iron-rich red blood cells settle beneath white blood cells and plasma, forming a gradient. The rate of sedimentation reflects the presence and intensity of inflammation within the body (MediSearch, 2023). Elevated ESR values can serve as important clinical indicators in oncology. High ESR levels have been associated with undiagnosed malignancies and are often correlated with poorer prognosis. This parameter may help clinicians assess absolute cancer risk, identify common sites of tumor involvement, and evaluate overall disease progression. Notably, ESR has shown utility in distinguishing between aggressive and non-aggressive tumor types, such as in prostate cancer. Additionally, elevated ESR may indicate metastatic spread and has been linked to reduced survival rates in patients with cutaneous melanoma (MediSearch, 2023).

RESULTS AND DISCUSSION

Blood Analysis in Oncology Patients

According to the results obtained from the analysis of certain blood parameters in the cancer patients covered by this research (Table 2), it can be concluded that with the exception of sedimentation, the average measured values for all other parameters are within the reference values.

Namely, from the blood analysis, it can be seen that hemoglobin in cancer patients ranges from 8 to 15.5 g/dl, with the average measured value being 12.4 g/dl, with a standard deviation of about 2 g/dl. The results show that 24% of the cancer patients covered by the research have hemoglobin below the reference level of 11 g/dl, while the rest have hemoglobin within the reference values.

The situation is similar to erythrocytes, where 26% of patients have slightly lower values than normal. Namely, the values of this parameter range from a minimum of 2.86 to a maximum of 5.12 $x10^6 mm^3$, while the average measured value for all patients is 4.1 $x10^6 mm^3$, and the standard deviation is equal to 0.6 $x10^6 mm^3$.

Table 2

Descriptive statistics of blood count results

Indicator	Hgb.	Erith.	Sluggi	Leuk.	Iron	Sedim.	AST	ALT	Urea	Creat
Reference values	11-16,3	3,8-5,8	150-390	3,5-10	60-180	3-10	2-37	3-41	1,7-8,3	53-115
Average	12.4	4.1	222.1	8.6	80.2	14.3	30.7	35.2	8.0	91.8
Median	12.2	4.1	201.2	8.5	70.9	10.5	30.1	32.2	7.4	80.8
Minimum	8	2.86	92.2	4.6	30.3	5	15.5	15.32	5	50.35
Maximum	15.5	5.12	401	15.5	150.2	40	60.6	90.5	22.2	200.2
Stand. Dev.	2.0	0.6	69.4	2.1	23.8	9.8	11.1	15.1	2.4	29.0
Asymmetric.	-0.15	-0.29	0.79	0.78	1.27	1.33	0.72	1.05	3.68	1.34
Flatness	1.98	2.19	3.15	3.73	4.71	3.91	2.80	4.02	18.88	4.66
Zhark-Bera	4.63	4.10	10.5	12.35	39.3	32.9	8.7	22.7	1276.5	41.6
P-value	0.099	0.129	0.005	0.002	0.000	0.000	0.013	0.000	0.000	0.000

According to the reference values, platelets in the blood should be between 150 and 390 $x10^6mm^3$. In the patients analyzed, platelets ranged between 92.2 $x10^6mm^3$ and 401 $x10^6mm^3$, with a mean of 222.1 $x10^6mm^3$ and a standard deviation of 69.4 $x10^6mm^3$. Only 11% of the total number of patients have platelets outside the reference values.

Regarding leukocytes, 18% of patients have leukocytes above the upper limit of 10 $x10^6mm^3$, while in the rest of the cancer patients, the value of this parameter is within the normal range. The minimum value of this parameter is 4.6 $x10^6mm^3$, and the maximum is 15.5 $x10^6mm^3$, while the average for all patients is 8.6 $x10^6mm^3$, with a standard deviation of 2.1 $x10^6mm^3$.

For iron, cancer patients have values of this blood parameter between 30.3 and 150.2 mcg/dl. The average value of the parameter for all patients is 80.2 mcg/dl, with a standard deviation of 23.8 mcg/dl. A total of 8% of patients have blood iron

values below the lower reference limit of 60 mcg/dl, while the rest have iron within the reference values.

The only deviation from the reference values of the measured parameters is the sedimentation, where the average measured value for all patients is above the upper reference limit. Namely, the sedimentation in cancer patients covered by the analysis ranges between 5 mm/h and 40 mm/h, while the average is 14.3 mm/h. The reference values for this parameter are between 3 and 10 mm/h, while exactly half of the patients have elevated sedimentation.

Regarding AST and ALT parameters, 28% and 30% of patients have elevated values. For AST, measured values in cancer patients range between 15.5 and 60.6 U/L, while for ALT, the range of measured values ranges from 15.32 to 90.5 U/L. The mean AST for all cancer patients was 30.7 U/L, with a standard deviation of 11.1 U/L, while the mean ALT for all patients was 35.2 U/L, with a standard deviation of 15.1 U /L.

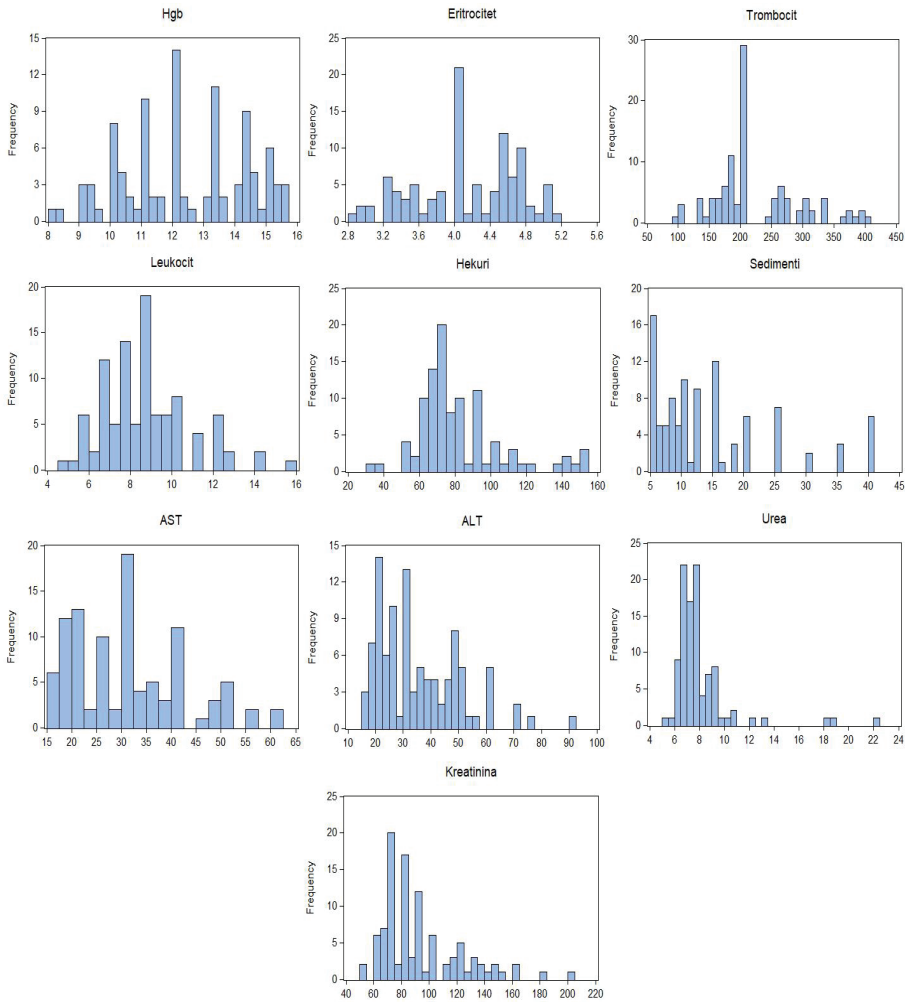
Regarding urea, reference values are between 1.7 and 8.3 mm/l, while the range of measured values in cancer patients is between 5 and 22.2 mm/l. The average value of urea for all cancer patients is 8 mm/l, with a standard deviation of 2.4 mm/l, while 24% of the patients covered by the research have elevated values of this parameter.

Finally, regarding creatinine, 2% of cancer patients have creatinine below the lower reference limit of 53 mm/l, while 21% of patients have elevated creatinine, that is, creatinine above the upper limit of 115 mm/l. The range of values of this parameter ranges from 50.35 mm/l to 200.2 mm/l, with the average value for all patients being 91.8 mm/l, with a standard deviation of 29 mm/l.

Figure 1 below shows the distribution of values for blood parameters. As can be seen from what is shown, it can be concluded that the data do not follow a normal distribution, which indicates that in further analysis it is necessary to use non-parametric tests, such as the test for the equality of the median, which follows a Chi-square distribution, or Spearman's rank correlation coefficient.

Figure 1

Distribution chart of blood count parameters



From the perspective of the relationship between the age of patients and the values of blood parameters, it can be concluded that there is a moderate positive relationship (correlation) between age and: leukocytes (0.22), sedimentation (0.37), AST (0.28), ALT (0.26), urea (0.37), and creatinine (0.43). Additionally, we have a somewhat weaker, but negative relationship between age and iron, with a correlation coefficient of -0.18. In other words, an increase in age is often followed by an increase in the values of the above parameters, that is, a decrease in the value of iron.

In contrast to this, we have the absence of any linear relationship between the age of patients and hemoglobin, erythrocytes, and platelets, where the correlation coefficients assessed are statistically insignificant at the 0.05 significance level. It is important to point out that the rest of the coefficients are statistically significant at the 0.05 significance level, with the exception of the one between patients' age and iron, which is statistically significant at the 0.1 significance level.

Table 3

Correlation coefficients between age The blood count.

Parameter	Age
Hemoglobin	-0.10
ritrocite	-0.13
Platelets	0.15
Leukocytes	0.22
Iron	-0.18
Sedimenti	0.37
S	0.28
L	0.26
Urea	0.37
Creatinin	0.43

From the perspective of different characteristics of the patients, an analysis was made to see if there are certain systematic differences in the median values of the measured blood parameters between separate groups of patients. In this context, the results of the analysis showed that older patients generally have higher median values of urea and creatinine, with these values tending to increase with increasing age. So, for example, people up to the age of 40 have a median urea value of 7 mm/l, while those over 70 years of age have a median urea of 7.91 mm/l.

The situation is similar to creatinine. Namely, people under the age of 50 have a median value of creatinine of 70.5 mm/l, while the median value of creatinine for those over 70 years old is 95.45 mm/l.

In terms of gender, the analysis showed that men generally have higher values than women for leukocytes (8.8 x$10^6$$mm^3$ in men vs. 7.7 x$10^6$$mm^3$ in women), sedimentation (15 mm/h in men vs. 8 mm/h in women), urea (7.7 mm/l in men vs. 7.01 in women) and creatinine (90.25 mm/l in men vs. 70.7 mm/l in women).

Finally, differences in median values for almost all blood parameters were also observed between patients with metastatic disease and those without metastatic disease. Namely, patients who have metastasis of the disease have generally higher values of leukocytes (9.05 vs. 7.7), sedimentation (15 vs. 8.5), AST and ALT (34.3, respectively 40.3 vs. 22.75, that is, 25.25), urea (7.79 vs. 7), and creatinine (90.3 vs. 72.2). On the other hand, patients with metastasis of the disease have generally lower values of hemoglobin (11.26 vs. 13.3), erythrocytes (4 vs. 4.43), and iron (70.2 vs. 80.3).

Regarding the other independent variables that were the subject of analysis, the results showed the absence of statistically significant differences in the median value of blood parameters between patients with different places of residence, different body mass index, cigarette consumption, presence of a history of disease in the family, as well as between patients receiving different treatments of the disease.

In addition, the graphs below (Fig. 2, Fig. 3, Fig. 4) graphically show the identified differences.

Figure 2

Graphs showing the differences in the median values for specific parameters from the blood count, according to the age of the patients

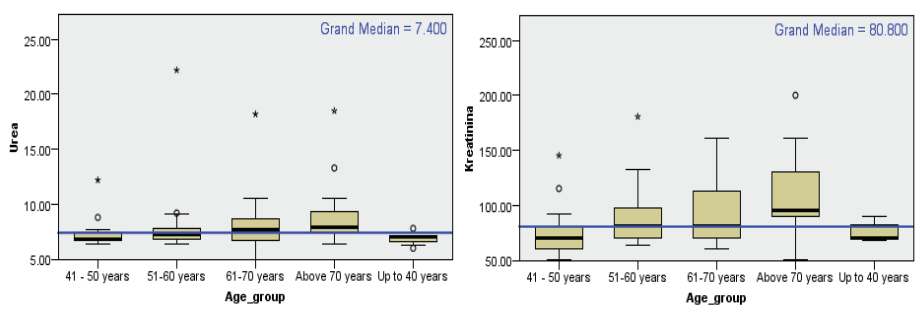


Figure 3

Graphs showing the differences in the median values for certain parameters from the blood count, according to the gender of the patients.

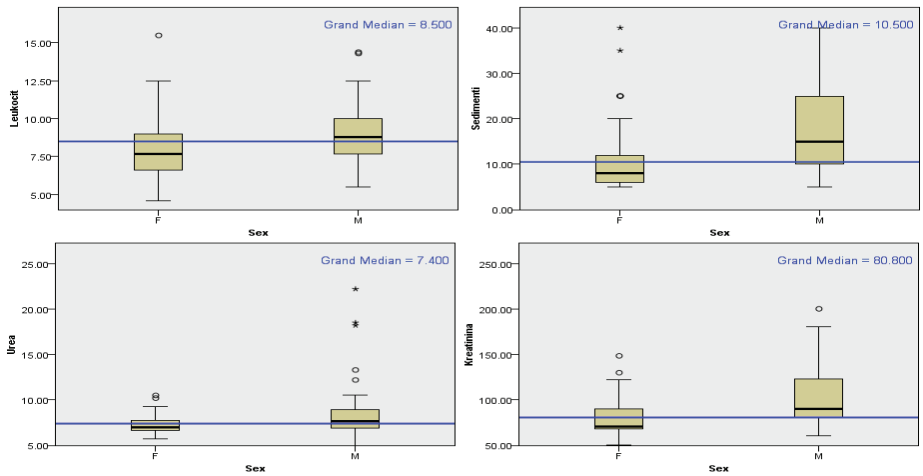
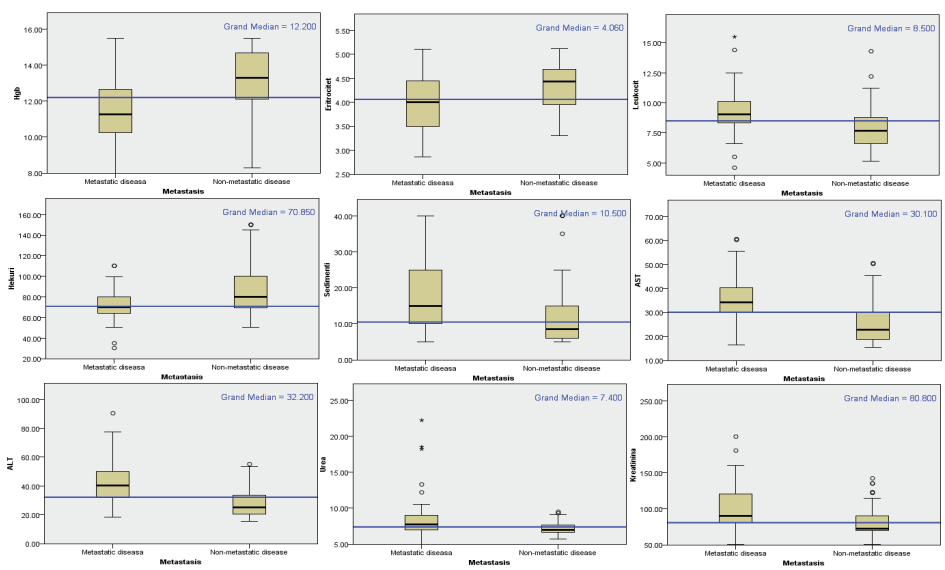


Figure 4

Graphs showing the differences in median values for certain blood count parameters, according to whether the disease has metastasized or not



Finally, the results of the conducted equality tests are shown in Table 4 below.

Table 4

Tests for equality of median values of blood count parameters for different groups of patients according to age group, sex, and whether the disease is metastatic or not.

Parameter	Age-Group		Gender		Metastasis	
	Stat. on a test.	P-value.	Stat. on a test.	P-value.	Stat. on a test.	P-value.
Hemoglobin	6.639	0.156	0.361	0.548	17.832	0.000
Eritrociti	2.081	0.721	0.040	0.841	4.848	0.028
Platelets	8.066	0.089	1.000	0.317	0.040	0.841
Leokocit	3.174	0.529	5.762	0.016	12.928	0.000
Iron	0.611	0.962	0.400	0.841	9.014	0.003
Sedimenti	7.287	0.121	14.440	0.000	17.668	0.000
ACT	5.627	0.229	0.640	0.424	15.968	0.000
AJIT	6.139	0.189	1.973	0.160	17.512	0.000
Urea	14.760	0.005	4.840	0.028	9.014	0.003
Creatinin	12.784	0.012	11.579	0.001	11.488	0.001

Integrating Nutrition into Oncology: Clinical Benefits and Future Directions

Nutrition plays a vital role in cancer care, influencing disease progression, treatment tolerance, and quality of life. Nutritional interventions—including counseling, supplements, and enteral or parenteral support—have been shown to improve body weight, energy, protein intake, and reduce treatment-related side effects (Prado et al., 2022). These benefits lead to better treatment adherence, survival, shorter hospital stays, lower costs, and improved physical function (Caccialanza et al., 2020; Senesse et al., 2008). Beyond physical health, nutrition supports psychosocial well-being by reducing anxiety, preserving social bonds around meals, and promoting healthy habits. Combined with exercise, it helps maintain muscle mass and independence. Patients receiving nutritional counseling often report improved quality of life (Sonneborn-Papakostopoulos et al., 2021; Marin Caro et al., 2007). Early, routine nutrition screening and a personalized, multimodal assessment are essential to optimize patient care.

Future Directions

Despite evidence, nutrition remains under-integrated in oncology due to limited awareness and dialogue. Future efforts should focus on:

- Embedding nutrition specialists in oncology teams
- Standardizing malnutrition screening tools
- Developing cancer- and treatment-specific nutritional protocols
- Investing in research on nutritional genomics, microbiome interactions, and immunonutrition
- Ensuring universal access to nutrition services regardless of socioeconomic status

Integrating nutrition fully into cancer care will improve clinical outcomes and enhance survivorship and quality of life.

Conclusion

The results of the blood analysis show that 24% of the 100 cancer patients covered by the research have hemoglobin below the reference level of 11 g/dl, while the rest have hemoglobin within the reference values. In 26% of patients, the values of erythrocytes are somewhat lower than normal values. Only 11% of the total number of patients have platelets outside the reference values. 18% of patients have leukocyte values above the upper limit of 10 $x10^{6}mm^{3}$. A total of 8% of patients have blood iron values below the lower reference limit of 60 mcg/dl. Sedimentation in cancer patients covered by the analysis ranges between 5 mm/h and 40 mm/h, while the average is 14.3 mm/h. The reference values for this parameter are between 3 and 10 mm/h, while exactly half of the patients have elevated sedimentation. Regarding AST and ALT parameters, 28% and 30% of patients have elevated values. The average value of urea for all cancer patients is 8 mm/l, with a standard deviation of 2.4 mm/l, while 24% of the patients covered by the research have elevated values of this parameter. Regarding creatinine, 2% of cancer patients have creatinine below the lower reference limit of 53 mm/l, while 21% of patients have elevated creatinine, that is, creatinine above the upper limit of 115 mm/l.

The increase in age is followed by an increase in the values of the parameters leukocytes, sedimentation, AST, ALT, urea, and creatinine, that is, a decrease in the value of iron. The results of the analysis showed that older patients generally had higher

median urea and creatinine values, with these values tending to increase with increasing age. Differences in the median values for almost all blood parameters are also observed between patients who have metastasized disease and those in whom the disease has not metastasized. Namely, patients who have metastasis of the disease have generally higher values of leukocytes (9.05 vs. 7.7), sedimentation (15 vs. 8.5), AST and ALT (34.3, respectively 40.3 vs. 22.75, that is, 25.25), urea (7.79 vs. 7), and creatinine (90.3 vs. 72.2). On the other hand, patients with metastasis of the disease have generally lower values of hemoglobin (11.26 vs. 13.3), erythrocytes (4 vs. 4.43) and iron (70.2 vs. 80.3). Evaluating the nutritional status of every cancer patient is essential, as nutrition influences disease progression, symptoms, treatment response, recovery, and overall prognosis. Health systems must invest in qualified nutrition experts to support cancer care teams. Currently, insufficient staffing and limited recognition of nutrition's importance among oncologists and policymakers hinder patients from accessing vital nutritional support. There is a critical need for enhanced education for both healthcare professionals and patients, especially regarding nutrition in digestive system cancers. Developing targeted educational and scientific resources will help raise awareness and enable effective nutritional interventions tailored to patient needs.

References

Aziz, H. A., & Habeeb, J. M. (2019). Study the effect of chemotherapy on some hematological and biochemical parameters of cancer patients in AL-Muthanna Province. Indian J PublicHealth Res Dev., 10(2), 813–816.

Caccialanza, R., Pedrazzoli, P., Cereda, E., Gavazzi, C., Pinto, C., Paccagnella, A., Beretta, G., Nardi, M., Laviano, A ., Zagonel, V . (2016). Nutritional support in cancer patients: a position paper from the Italian Society of Medical Oncology (AIOM) and the Italian Society of Artificial Nutrition and Metabolism (SINPE). J Cancer, 7, 131–135.

Caccialanza, R., Goldwasser, F., Marschal, O., Ottery, F., Schiefke, I., Tilleul, P., Zalcman, G., & Pedrazzoli, P. (2020). Unmet needs in clinical nutrition in oncology: A multinational analysis of real-world evidence, Ther Adv Med Oncol., 12, 1-10.

Devi, L.I., Ralte, L., & Ma, A. (2006). Serum biochemical profile of breast cancer patients. Eur J Pharm Med Res., 2(6), 210–214.

Dina, K., Pavlova, V., Tahiri, A., Kovaci, I., Blazhevska, T., Knights, V., Pavlovska Dimkovski, M., & Srbinoska, M. (2024). A survey on the motivation, attitudes and frequency of use of dietary supplements in cancer patients, Journal of Hygienic Engineering and Design, 47, 121-126.

Lodish, H., Arnold, B.S., Lawrence, Z., Paul, M., & David, B.J.E. (2000). Molecular Cell Biology. 4th ed. NCBI.

Marin Caro, M.M., Laviano, A., & Pichard, C. (2007). Impact of nutrition on quality of life during cancer. Curr Opin Clin Nutr Metab Care, 10, 480–487.

MediSearch (2023). https://medisearch.io/blog/what-is-esr-level-in-cancer-patients

Moore, D.C. (2016). Drug-induced neutropenia. Pharmacovigil Forum., 41 (12), 765–768.

Mughal, T. I. (2004). Current and future use of hematopoietic growth factors in cancer medicine. Hematol Oncol., 22, 3,121–134.

Muscaritoli, M., Arends, J., Bachmann, P., Baracos, V., Barthelemy, N., Bertz, H., Bozzetti, F., Hutterer, E., Isenring, E., Kaasa, S., Krznaric, Z., Laird, B., Larsson, M., Laviano, A., Muhlebach, S., Oldervoll, L., Ravasco, P., Solheim, S. T., Strasser, F., de van der Schueren, M., Preiser, J.C., & Bischoff, S.C., (2021). ESPEN practical guideline: Clinical Nutrition in cancer, Clin Nutr, 40, 2898–2913.

Network NCC. (2018). Cancer- and Induced Anemia. In: NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®), 1–51.

Prado, C.M., Laviano, A., Gillis, C., Sung, A., Gardner, M., Yalcin, S., Dixon, S., Newman, S., Bastasch, M., Sauer, A., Hegazi, R., & Chasen, M. (2022). Examining guidelines and new evidence in oncology nutrition: a position paper on gaps and opportunities in multimodal approaches to improve patient care, Support Care Cancer, 30, 3073–3083.

Ramasamy, P., & Sundaresan, S. (2016). Haematological and biochemical changes in pre and post treatment of buccal mucosa carcinoma patients, International journal of pharmaceutical science and research, 7 (7), 1000-1005.

Senesse, P., Assenat, E., & Schneider, S. (2008). Nutritional support during oncologic treatment of patients with gastrointestinal cancer: who could benefit? Cancer Treat Rev, 34, 568–575.

Shahid, S. (2016). Critical reviews in oncology/hematology review of hematological indices of

cancer patients receiving combined chemotherapy & radiotherapy or receiving radiotherapy alone. Crit Rev Oncol Hematol. 105, 145–155.

Shander, A., Knight, K., Thurer, R., Adamson, J., & Spence, R. (2004). Prevalence and outcomes of anemia in surgery: a systematic review of the literature, Am. J. Med., 116, 7A, 58S-69S.

Sheikh, N., & Masood, M.N.N. (2014). Hematological and serological changes. J Soc Dev New Net Env. 5, 6.

Shreya, S., Shekher, A., Puneet, P., Prasad, S. B., & Buddhi Prakash Jain B. P. (2023). Haematological and biochemical analysis of blood samples from early and late stage breast cancer patients in India, Bioinformation, 19, (7), 806-809.

Sobin, L.H., & Wittekind, C. (2009). TNM Classification of Malignant Tumors. 7th ed. In Sobin LH, Gospodarowicz MK, Wittekind C, eds. Wiley-Blackwell, 1–333.

Sonneborn-Papakostopoulos, M., Dubois, C., Mathies, V., Heß, M., Erickson, N., Ernst, T., & Huebner, J. (2021). Quality of life, symptoms and dietary habits in oncology outpatients with malnutrition: a cross-sectional study, Med Oncol., 38, 1-20.

Steele, M, & Narendran, A. (2012). Mechanisms of defective erythropoiesis and anemia in pediatric acute lymphoblastic leukemia (ALL). Ann Hematol. 91(10), 1513–1518.

Uo, H. (2006). Science GS of NA and. Disaster Nursing in a Ubiquitous Society Care Package for Cancer Patients In Times of Handbook of Self-Managing. Side Effects of Chemotherapy in Times of Disaster. In: Handbook of Self-Managing Side Effects of Chemotherapy in Times of Disaster; 1–26.

Van Cutsem, E., & Arends, J. (2005). The causes and consequences of cancer-associated malnutrition, Eur J Oncol Nurs, 9, S51–S63.

Vitaloni, M., Caccialanza, R., Ravasco, P, Carrato, A., Kapala, A., van der Schueren, M., Constantinides, D., Backman, E., Chuter, D., Santangelo, C., & Maravic, Z. (2022). The impact of nutrition on the lives of patients with digestive cancers: a position paper, Supportive Care in Cancer, 30, 7991–7996.

Wigmore, S.J., Plester, C.E., Ross, J.A., & Fearon, K.C.H. (1997). Contribution of anorexia and hypermetabolism to weight loss in anicteric patients with pancreatic cancer, Br J Surg., 84, 2, 196-197.

Wondimneh, B., Setty, S.A.D., Asfeha, G.G., Belay, E., Gebremeskel, G., & Baye, G. (2019). Comparison of Hematological and Biochemical Profile Changes in Pre- and Post-Chemotherapy Treatment of Cancer Patients Attended at Ayder Comprehensive Specialized Hospital, Mekelle, Northern Ethiopia 2019: A Retrospective Cohort Study, Cancer Management and Research, 13, 625–632.

INTERNATIONAL JOURNAL OF TECHNICAL AND NATURAL SCIENCES

DOI: https://doi.org/10.69648/NVAX8158

International Journal of Natural and Technical Sciences (IJTNS), 2025; 5(1): 43-75

ijtns.ibupress.com

Online ISSN: 2671-3519

Application: 02.06.2025

Revision: 16.06.2025

Acceptance: 26.06.2025

Publication: 30.06.2025

Saxena, M., & Singh, N. P. (2025). Role of Hybrid Feature Selection Algorithms in Foreseeing Student Performance. International Journal of Natural and Technical Sciences, 5(1), 43-75. https://doi.org/10.69648/NVAX8158

Mamta Saxena, School of Engineering and Technology, MVN University Palwal, Haryana, India, https://orcid.org/0009-0007-8476-2836. Email: 17ca9001w@mvn.edu.in

Netra Pal Singh, School of Engineering and Technology, MVN University Palwal, Haryana, India, https://orcid.org/0000-0002-3006-9522. Email: deanset@mvn.edu.in

BALKAN JOURNALS

Role of Hybrid Feature Selection Algorithms in Foreseeing Student Performance

Mamta Saxena, Netra Pal Singh

Abstract

The interest of analysts in applying analytical algorithms to investigate students' performance data with a view to enhancing their knowledge is growing among data miners. The key factor of this trend is ample multimedia data generated by educational institutions with the usage of technologies, tools of e-learning, and other digital platforms for conducting online courses. Educators could utilize these data to examine and understand students' learning behaviors by using data mining techniques to forecast student achievement, among other things. In data mining models, the difficult task is to select effective strategies that will generate satisfactory forecast accuracy. To achieve such goals, this article uses six hybrid feature selection (FS) algorithms such as hybrid PSO, hybrid GA, filter wrap, wrapper embedded, ensemble based, and filter embedded and five algorithms such as "Random Forest (RF)", "Decision Tree (DT)", "Logistic Regression (LR)", "K-Nearest Neighbor (KNN)", and "Support Vector Machine" (SVM) to forecast the performance of students and to make comparison of performance of five classification processes in the perspective of six FS models. Based on the analysis of the given data set, it is concluded that the performance of classification models improved when 20 important features were part of the data in place of all 30 attributes. Further reduction in the number of attributes did not result in further improvement in the performance of classification models in predicting students' success or failure.

Keywords: 'Hybrid Feature Selection', 'Random Forest', 'Decision Tree', 'Logistic Regression', 'KNN', 'SVM', 'EDM'

INTRODUCTION

Every educational institution desires that students who are enrolled with them must achieve academic success to a high standard because it helps in creating brand value for the institutions. Students, on the other hand, value their academic performance since it is central to achieving success, including getting admitted to a prominent university or institution. It will finally fetch a remunerative job with better growth in their career. It will ultimately raise their lifestyle and living. The question is "how does data mining come into the picture?" To mention, researchers from the University of Miami claim that Grade Point Average (GPA) in high school can predict the category of college a student will attend. In addition, GPA in high school can also predict who will complete their degree program and who will not, along with predictions about their future earnings (Marte, 2014). Therefore, it is becoming essential to develop better data mining techniques to estimate the educational accomplishment of the students beforehand, making use of easily accessible data. This new set of improved data mining algorithms will assist analysts in finding masked patterns, classifying the data into predefined categories, and predicting students' academic accomplishment from students' databases at universities.

It is a well-documented fact that for extracting valuable insights from volumes of data that are being captured by all sectors of the economy, data analytical algorithms are used. Some of the key contribution of the researchers of data mining in different sectors are Koh and Tan (2005), Sodhro and Zahid (2021) in healthcare sector, Vazan et al (2017) in manufacturing and engineering, Awoyemi et al (2017) in fraud detection, Lan et al (2018) in bioinformatics, Farzaneh and Fadlalla (2017) in business organizations, Gosh and Chaudhuri (2021) in stock markets, Shah (2022) in remote sensing. In addition, researchers have applied data mining algorithms in many other fields. The educational sector is no exception to these trends. Large volumes of data are collected by the educational institutions as these institutions have invested in the implementation of technologies in the classroom, i.e., usage of smart boards, computers, smartphones, tablets, iPads, learning management systems, virtual reality systems, etc. (Zauzmer, 2020; Winstead, 2025; Jackson, 2013; Velazquez, 2023). These investigations necessitate scraping information from these raw databases and finding patterns or valuable information with educational data mining (EDM) tools.

EDM tools are very useful for making desirable and accurate predictions, whether students will pass or fail, by extracting hidden information from the educational

databases available with educational institutions, rather than waiting for the final results of the examination. Hasan et al (2020) discovered that the success or failure of students could be predicted with 88.3% accuracy with the help of a random forest classification algorithm.

Data pertaining to students is of interest to management of institutions because it can help in predicting performance of students well in advance that can be used in devising strategies for taking corrective measures if needed (Nagahi et al., 2020; Okubo et al.,2017; BYung-Hak, 2018); predicting dropouts (Wang et al., 2017; Whitehill et al., 2017; Xing & Dongping, 2019), identifying undesirable behavior of students (off-task behaviors) (Bohong et al., 2020; Baker et al., 2016; Goldberg et al., 2021, Cetintas et al., 2010, Zaletelj & Kosir, 2017), and tracking psychological state of student(s) in real time with wearable technology and sensors (Antoniou et al., 2020; Ahonen et al., 2018; Magsi et al., 2021). It is a well-known fact that an academic success or failure of a student is influenced by their parents' educational background, personal characteristics, psychological makeup, and other contextual circumstances, such as living environment. EDM helps identify the connection between these factors and the results of students in their examinations. Helal et al. (2018) used academic and demographic data to identify students who are weak in their studies and advised them to put extra effort into their studies so that they can do well in their academic pursuits. Their study is based on four classification algorithms, such as 'Naïve Bayes (NB)', 'Sequential Minimal Optimization (SMO)', 'J48', and 'Repeated Incremental Pruning to Produce Error Reduction (JRIP)'. They developed sub-models for the purpose with better efficiency. Hussain et al. (2018) found that internal evaluation marks are the most significant factors influencing the results of the students. Fernandes et al (2019) used a gradient boosting machine algorithm to predict the success of the students. They concluded that "grades" of previous exams and "absences" in the classroom are key features in forecasting the performance of the students in examinations. They further concluded that the status of the school, age of the students, and neighborhood of home location are prospective features in forecasting the success of the students.

Keeping in view the importance of the hidden knowledge in databases of students, this research article analyzed the data set of students with five machine learning algorithms in combination with six hybrid FS algorithms for selecting the best features to foresee the performance of the students with existing data. The hybridization involves either the FS algorithm or the machine learning algorithm, but in the present research paper, it is in the FS techniques. The research paper is structured

into five sections, starting with an introduction section. It is followed by a review of the literature on classification and hybrid FS algorithms in the context of EDM in Section 2. Six hybrid FS techniques and five classification methods, along with the methodology of the present study, are presented in Section 3. The analysis and results of the analysis, along with a brief interpretation, are embodied in Section 4. Section 5 wraps up the work in the form of a conclusion.

LITERATURE REVIEW

ML algorithms along with FS algorithms are used in many domains of study for classifying variables of interest (Nagahisarchoghaei et el, 2020; Shah and Du, 2021a; Shah and Du, 2021b; Shah and Du, 2021c). In educational data analysis, ML and Deep Learning (DL) models have been applied in the past to forecast the achievement of students. A review of select articles EDM used for predicting academic achievements of the students is presented in this section. These methods are applied to forecast the academic success of students that can be explained with many factors, i.e., socio-economic characteristics, lifestyle, family structure, peers of students, etc. Additionally, past studies have considered gender of student's friends, the time spent on social media for networking, family characteristics (size and income of the family, parents' education, parent profession), academic examination test score data, lab experiment results, being present in classes, GPA, prior classes grades, and demographic information such as age, gender, race, etc.). In addition, academic achievements before and after enrollment in a given degree program or class are used as a proxy for academic qualities in many studies.

Aulck et al. (2016), Daud et al. (2017), and Garg (2018) used demographic characteristics of the students to predict their academic achievement. To be specific, Aulck et al. (2016) made use of age data, while Kemper et al. (2020) concentrated on the country of origin of students for predicting the success or failure of students. Aluko, et al. (2018) and Luhaybi, et al. (2018) considered the grade point average (GPA) of the previous class for the purpose of predicting performance. GPA is frequently used as a pre-enrolment feature since the academic performance of the students, as reflected in their prior endeavors, always impacts future successes. Post-enrollment features such as marks in tests, midterm examinations, quizzes, and assignments have been used to assess student achievements by Asif et al (2017). Li et al. (2019) reported that participation in extracurricular activities has a big impact on academic achievement, and the success rate is also influenced

by attendance and participation in classroom discussion. Francis and Babu (2019) mentioned that the family characteristics like size and income can have a big impact in predicting performance of the students in many cases.

In a nutshell, EDM is an effective method for forecasting the educational accomplishment of the students. The two machine learning algorithms most frequently utilized in analyzing education data are random forest (RF) and decision tree (DT). Hussain et al (2018) applied J48, "Decision Tree", "JRIP", and "Gradient Boosting" Classifiers. Heuer and Breiter (2018) applied logistic regression, 'Decision Tree', 'Support Vector Machine' (SVM), and random forest (RF) classifiers for classifying the achievement of students as success or failure. To predict dropouts, Haiyang et al (2018) applied a Time Series Forest (TSF) classification model. Rizvi et al (2019) applied a decision tree-based approach for studying the role of demographics in online teaching in order to foresee triumphs of the students. Hlosta et al. (2017) used a framework of Ouroboros and classification models such as 'support vector machine', 'Random Forest', 'Naïve Bayes', 'Tree Boosting XG Boost', and logistic regression algorithms for the identification of students who are not performing well.

According to Wasif et al. (2019) student demographic and enrollment characteristics are helpful to ascertain academic success. They concluded that the Bayesian network performs better than the Decision Tree algorithm in predicting the achievement of students. To envisage academic accomplishment of the students, Alberto et al. (2021) employed methods such as multilayer perceptron (MLP), DT, RF, and extreme gradient boosting. The best performance was of MLP, and concluded that MLP accuracy was highest, i.e., 78.2%. Azizah et al. (2018) applied 'Naive Bayes' (NB) and 'Decision Tree' to forecast who will pass and who will fail. He concluded that NB had the best accuracy (63.8%).

Hybrid Models are applied for guessing students' future performance in different combinations. The following Table 1 provides a clear view of classification hybrid models along with references. However, the present study has not used these hybrid models. Hybridization is done for FS algorithms in this study.

Table 1:

List of Hybrid Classification Algorithms

Hybrid Model	Research Reference
Decision Tree + Logistic Regression	Kumar & Soni (2019). "Predicting Student Performance using Hybrid Classification Techniques." *IJCSIT*.
Random Forest + Support Vector Machine	Amado, et al. (2015). "A Hybrid Model for Predicting Student Performance in Educational Data Mining." *JEDM*.
Neural Network + Decision Tree	Silva, et al. (2017). "A Hybrid System for Predicting Students' Academic Performance." *IJCSE*.
K-Nearest Neighbours + Naive Bayes	Shabir & Arshad (2017). "Hybrid Machine Learning Model for Predicting Student Performance." *ICCCET*.
Gradient Boosting + Random Forest	Deshpande, & Agarwal. (2021). "Performance Prediction of Students using a Hybrid Ensemble Learning Model." *IJACSA*.
Support Vector Machine + Naive Bayes	Yaseen, Z., et al. (2020). "Predicting Student Performance Using Hybrid SVM and Naive Bayes." *IJCAI*.
KNN + Neural Network + Random Forest	Zhang, et al. (2019). "A Hybrid Ensemble Model for Student Performance Prediction." *AIED*.
XGBoost + Logistic Regression	Ali, & Qureshi (2020). "Predicting Academic Performance of Students using Hybrid Machine Learning Algorithms." *LJSS*.
Bagging + Boosting (Random Forest + AdaBoost)	Ebrahim, et al. (2018). "A Hybrid Ensemble Model for Academic Performance Prediction of Students." *JES*.

RESEARCH METHODOLOGY

The process of employing data mining tools to forecast and examine academic records of the students involves a methodical approach of gathering, preprocessing and examining a range of data sources pertaining to the demographics and academic credentials of pupils. Data cleansing, FS, classification model selection, training, and evaluation are important phases. This research article uses hybrid

FS algorithms to identify features with high predictability and classification algorithms to classify the students as good or bad performer. The research objectives, description of data sets, etc., are presented in the subsequent sections.

OBJECTIVE / QUESTIONS

The main research objective is to examine the impact of hybrid FS algorithms on the effectiveness of classification algorithms applied to the best set of predictors. Two research questions will be addressed by this study:

RQ1. Which features of data sets are crucial for predicting students' academic success?
RQ2. Which assemblage of hybrid FS and classification models works best together toward forecasting the success or failure of students?

Description of the Dataset:

The dataset comprises 395 students' records with 30 features for each record. This dataset has been used in many studies and is available publicly on many data repositories such as Kaggle and the UCI Data Repository. It was previously used to check the students' academic success and passing rates. There are three categories of attributes in this dataset (i) demographic features (sex, age, address, family size, Parent status, health), (ii) academic background features (type of school, failed attempts, time od studying in a day or week, support extended by school, paid activities, nursery, higher studies, absences from clases) and (iii) social-economic features (Academic qualification of Mother & Father, Job of Mother & Father, family support, reason for not performing, care of guardian, travelling time from home to school, internet used, romantic, family relation, free time, gout for outing, Weekday alcohol consumption, Weekend alcohol consumption). These characteristics are described in Table 2.

Table 2:

Description of Dataset

SN	Attribute	Type (Feature)	Measure	Description
1	School	Categorical	Nominal	'GP' – "Gabriel Pereira" or 'MS' – "Mousinho da Silveira"
2	Gender	Categorical	Nominal	Female or Male
3	Age (Years)	Categorical	Nominal	≥ 16' OR '< 16'
4	Address	Categorical	Nominal	'U' – "Urban" or 'R' – "Rural"
5	"Famsize"	Categorical	Nominal	' ≤ 3 or – > 3
6	"Pstatus"	Categorical	Nominal	"Living with parents": Yes or No
7	"M_edu"	Categorical	Nominal	"Education of Mother": '10^{th}, 12^{th} , 'Graduate' or 'Post Graduate' or 'Research'.
8	F_edu	Categorical	Nominal	"Education of Father" : 10^{th}, 12^{th}, 'Graduate' or 'Post Graduate' or 'Researcher".
9	"M_job"	Categorical	Nominal	"Job type of Mother" : 'Teaching', "health care professional", "Civil 'servant'" (e.g. "admin or police"), 'Home Maker' or "Other"
10	F_job	Categorical	Nominal	Father Job: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other'
11	Reason	Categorical	Nominal	Reason to choose this school: Nearness to 'home', 'reputation of the school', 'course' preference or 'other'
12	Guardian	Categorical	Nominal	"1-mother", "2-father" or "3-Other""
13	Travel_time	Categorical	Binary	<= 2 or > 2 Divided in to two categories based on hours
14	Study_time	Categorical	Binary	<= 2 or > 2 . Divided in to two categories based on hours
15	Failures	Categorical	Binary	Number of past class failures: <=3 or > 2
16	Schoolsup	Categorical	Binary	Additional School Support: '1-yes' or '2-no'
17	Famsup	Categorical	Binary	Family Educational Support: '1-yes' or '2-no'
18	Paid	Categorical	Binary	Extra paid classes within the course subject (Math or Portuguese) '1-yes' or '2-no'

19	Activities	Categorical	Binary	Extracurricular Accomplishments: '1-yes' or '2-no'
20	Nursery	Categorical	Binary	Attended Nursery School: 'yes' or 'no'
21	Higher	Categorical	Binary	Want to take Higher Education: 'yes' or 'no'
22	Internet	Categorical	Binary	Internet Access at Home: 'yes' or 'no'
23	Romantic	Categorical	Binary	With A romantic Relationship: 'yes' or 'no'
24	Famrel	Categorical	Binary	Quality of Family Relationships: (1 ("very bad") – 5 "Excellent") - converted to binary: < = 3 or : > 3
25	Freetime	Categorical	Binary	Free Time After School ("1- Very Bad" – "5- Excellent") converted to binary: <= 3 or > 3
26	Goout	Categorical	Binary	Going out with friends ("1- very bad" – "-5 Excellent") converted to binary: :<= 3 or : > 3
27	Dalc	Categorical	Binary	Work Day Alcohol Consumption (1 very bad – 5 Excellent) but converted to binary: :<= 3 or : > 3
28	Walc	Categorical	Binary	Weekend Alcohol Consumption ("1- very bad" – "5- Excellent") converted to binary: <= 3 or : > 3
29	Health	Categorical	Binary	Current health Status: ("1- very bad" – "5- Excellent") Changed to binary: :<= 3 or > 3
30	Absences	Categorical	Binary	Number of Absences: Altered in to Binary : ≤ 30 or : > 30
31	Passed (Target)	Categorical	Binary	"Yes" or "No"

"Source UCI Data Repository :https://archive.ics.uci.edu/dataset/320/student+performance". Note Zip Codes are not considered.

Experimental Setup:

Environment

To evaluate the rank/importance of attributes in the given dataset with select FS/ hybrid FS methods and to achieve the research objective, a set of six hybrid FS and five classification algorithms is identified based on a review of the literature.

For the purpose of analysis and implementation of a combination of two sets of algorithms, Python libraries are used as listed in Table 3. The steps involved in the computation are listed below.

Step 1: Selection of the Dataset

The dataset, as described in Table 2, is selected for applying hybrid FS algorithms and classifiers. The dataset is identified and downloaded from the existing repositories of datasets available at many sources on the Internet.

Step 2: Data Preprocessing

This step involves functions like loading the dataset, handling missing values, encoding categorical variables, and dividing the dataset into a training and a test dataset. The authors have used 80% data for training the models and 20% for testing the goodness of the model. Balancing of the training data set from the perspective of the target variable is also ascertained at this stage before taking the next step.

Step 3: Selection of Algorithms for FS and Prediction of the Target Variable of the Dataset

This step consists of the selection of six hybrid FS and five classification algorithms based on a review of the literature in the research domain.

Step 4: Generate Base Model Prediction Results

Results are calculated using five classification algorithms by considering all features of the given data set.

Step 5: Evaluation of the Performance of Hybrid FS and Classification Algorithms Using Quality Metrics/ Parameters

Computations on the selected dataset are done with five classification algorithms to forecast the performance of students. This is done for a different set of attributes, having their importance derived using hybrid FS algorithms. Metrics are applied and compared to identify which combination of algorithms achieves better values of quality metrics for a given set of attributes.

Libraries Used

This section presents a list of libraries that can be used for the hybrid FS method, classification algorithms, visualization, and evaluation matrices. The details are given in Table 3.

Table 3:

Libraries used for Computation and Visualization

Libraries of FS Methods		
Hybrid FS Method	**Libraries**	**Comments**
Hybrid PSO	Pyswarm	It combines with filter and wrapper methods
Filter-Embdded	Scikitlearn, MLxtend, TPOT, Borutpy	It combines filter and embedded methods
Hybrid GA	Pygad, inspired, Deap	GA searched the best subset of features. The hybrid part combines it with filter, wrapper, and embedded methods.
Wrapper-Embedded	Scikitlearn, MLxtend, TPOT, Borutapy	It combines the wrapper method with embedded methods.
Filter-Wrapper	Scikitlearn, MLxtend, TPOT, Borutpy	It combine filter with wrapper method.
Ensemble Based Mathods	Scikitlearn, mlxtend,Borutapy	It calculates importance features using multiple FS methods and ensemble best on voting
Libraries of Classification Algorithms		
Algorithm	Libraries	Comments
Random Forest	Sklearn.Emsemble	Library imports Random Forest Classifier
SVM	Sklearn.svm	Library imports SVM classifier
Decision Tree	Sklearn.decision-treeclassifier	Library imports decision tree classifier
Logistic Regression	Sklearn.linear_model	Library imports logistic regression classifier
KNN	Sklearn.neighbors	Library imports KNN Classifier
Libraries for Evalution Metrices and Visualization		
Evaluation Metrices	Scikit_learn.metrices	Calculates “accuracy (ACC)”, “precision (PRE)”, “recall (REC)”, “F1 Score (F1_S), ROC Curve (ROCC), and AUC curve (AUCC)
Visualization	Matplotlib, seaborn	For plotting graphs

Evaluation Metrics

Five key measures are chosen to evaluate the prediction ability of classification models. These measures are ACC, PRE, F1_S, PRE, REC, AUCC, and AUCC. All measures inferred together offer a more complete view of the model's performance. To find out how successfully the models applied in the research article detect real positives and negatives in more detail, with these measures.

Interpretation of Results

The present section presents ranks (a measure of the importance of the attributes) using six hybrid FS algorithms. In the next stage of computation, five classification algorithms are used to forecast the results of the students. This is achieved in four parts, i.e., considering all attributes, followed by considering the 20 most important attributes, the 15 most important attributes, and the 10 most important attributes. This is being done to conclude a tradeoff between loss of information (resulting in the selection of fewer attributes) and its influence on the outcome of classifiers in predicting the results of the students in the examination. Generally, it is concluded that by removing highly non-significant predictors, the performance of the classifier improves and reduces the requirement of computing power. But by eliminating more variables, it negatively impacts the performance of the classifier. The results are presented in the same sequence in subsequent sub-sections.

Ranks of Features as Calculated Using Six Hybrid FS Algorithms

This section presents ranks of attributes with six hybrid FS algorithms that are showcased in Table 4, along with a reference.

Table 4:

Ranks of all attributes with Hybrid Methods

Feature	Hybrid PSO Method (Xue, et al, 2016)	Hybrid GA Method (Xue, et al, 2016)	Filter Wrapper Method (Dash & Liu, 1997)	Wrapper-Embedded Method (Saeys et al, 2007)	Ensemble-Based Method (Saeys, et al, 2008)	Filter-Embedded Method (Guyon & Elisseeff, 2003)
	Rank	Rank	Rank	Rank	Rank	Rank
school	6	22	6	27	25	30
Sex	20	19	26	12	20	25
Age	24	5	3	13	14	4
Address	10	4	5	15	10	22
Famsize	16	26	27	20	22	20
Pstatus	17	29	23	28	28	29
Medu	9	25	9	5	6	2
Fedu	3	13	17	4	2	7
Mjob	8	21	30	2	3	5
Fjob	13	30	12	1	5	11
Reason	30	18	29	3	7	10
Guardian	27	27	21	6	9	15
Traveltime	12	17	28	29	29	28
Studytime	2	16	4	17	11	3
Failures	14	1	1	7	1	1
Schoolsup	28	24	25	21	8	8
Famsup	5	11	24	22	26	21
Paid	22	9	14	16	15	23
Activities	11	8	22	11	17	9
Nursery	15	6	8	23	23	24
Higher	26	2	10	26	16	17
Internet	19	10	20	25	24	26
Romantic	1	28	13	19	21	12
Famrel	7	3	19	14	12	14
Freetime	23	20	18	18	18	19
Goout	21	7	7	10	4	13
Dalc	18	12	16	30	30	27
Walc	29	14	15	24	27	6
Health	25	15	2	8	13	16
Absences	4	23	11	9	19	18

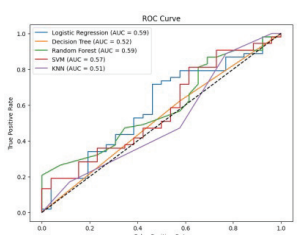
From Table 4, it can be seen that each model generates different ranks of the attributes. The reasons could be the assumption of each FS algorithm in the context of the nature of the features/ attributes of the data. Secondly, the subroutines used in Python do not consider pre-processing of data for the assumption (Sikri et al., 2023). This results in a wrong ranking in many cases

Results of Five Classification Algorithms & Quality Indexes

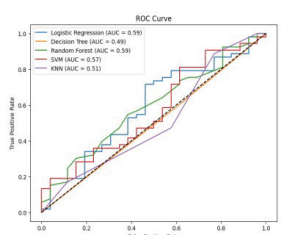
This section presents the performance of five classification algorithms in predicting the success or failure of the students. These algorithms are logistic Regression (LR), Decision Tree (DT), Random Forests (RF), Support Vector Machines (SVM), and K-nearest neighbor (KNN). Six quality parameters are used for making a comparison of the performance of five classification algorithms in conjunction with six hybrid FS methods. These quality parameters are (i) ACC, (ii) PRE, (iii) REC, (iv) F1_S, (v) AUCC, and (vi) ROCC. The values of these parameters for five classification algorithms without any FS are presented in Table 5. The results presented in Table 3 are treated as base model results and will be used for making comparisons when non-significant attributes are eliminated. The ROCCe of five classification models for six FS algorithms are shown in Fig. 1(a) to Fig. 1(f). It is worth mentioning that there is no role for the FS algorithm since all features are used for the purpose of classifying students into two categories, i.e., pass or fail. Evaluation matrices of Table 5 are considered only to bring symmetry in the presentation in relation to other results, wherein fewer numbers of features are selected to run classification algorithms, i.e., 20,15, and 10 attributes.

Fig 1:

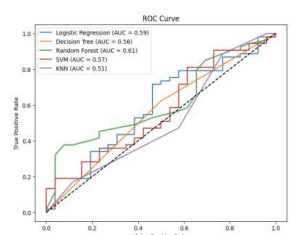
ROC Curves of classification models for all Features



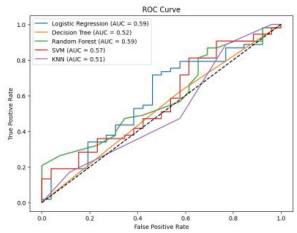
ROC Curve of Hybrid PSO Model
(c)



ROC Curve of Hybrid GA Model
(b)

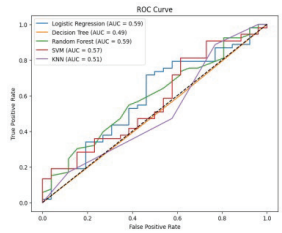


ROC Curve of Filter-Wrap Model
(a)



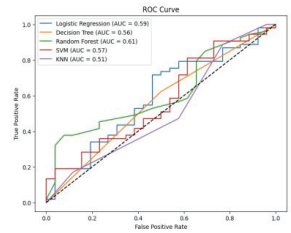
ROC Curve of Wrapper-Embedded Model

(d)



ROC Curve of Ensemble Model

(e)



ROC Curve of Filter-Embedded Model

(f)

Table 5:

Evaluation Metrics for all features using different Classification Algorithms

Hybrid PSO Model					
Algorithm Name	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.64557	0.686567	0.867925	0.766667	0.592888
DT	0.607595	0.729167	0.660377	0.693069	0.580189
RFt	0.632911	0.671429	0.886792	0.764228	0.59688
SVM	0.670886	0.675325	0.981132	0.800000	0.569666
KNN	0.670886	0.701493	0.886792	0.783333	0.508345
Hybrid GA Model					
Algorithm Name	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.64557	0.686567	0.867925	0.766667	0.592888
DT	0.582278	0.700000	0.660377	0.679612	0.541727
RF	0.683544	0.712121	0.886792	0.789916	0.595428
SVM	0.670886	0.675325	0.901132	0.800000	0.570029
KNN	0.670886	0.701493	0.886792	0.783333	0.508345
Filter Wrap Method					
Algorithm Name	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.64557	0.686567	0.867925	0.766666748	0.592888
DT	0.594937	0.714286	0.660377	0.686274448	0.560958
RF	0.658228	0.691176	0.886792	0.776859033	0.66582
SVM	0.670886	0.675325	0.911132	0.775703619	0.569666
KNN	0.670886	0.701493	0.876792	0.779407332	0.508345

Wrapper-Embedded Method					
Algorithm Name	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.64557	0.686567	0.867925	0.766666748	0.592888
DT	0.582278	0.708333	0.641509	0.673266937	0.551524
RF	0.683544	0.705882	0.90566	0.793388062	0.582003
SVM	0.670886	0.675325	0.901132	0.772056539	0.569666
KNN	0.670886	0.701493	0.826792	0.759006076	0.508345
Ensemble-Based Method					
Algorithm Name	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.64557	0.686567	0.867925	0.766667	0.592888
DT	0.531646	0.666667	0.603774	0.633664	0.494194
RF	0.658228	0.691176	0.886792	0.776859	0.643324
SVM	0.670886	0.675325	0.981132	0.800000	0.569666
KNN	0.670886	0.701493	0.886792	0.783333	0.508345
Filter-Embedded Method					
Algorithm Name	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.64557	0.677419	0.792453	0.730435	0.592888
DT	0.556962	0.686567	0.867925	0.766667	0.522859
RF	0.620253	0.687500	0.622642	0.653466	0.604862
SVM	0.670886	0.682540	0.811321	0.741380	0.569666
KNN	0.670886	0.675325	0.931132	0.782862	0.508345

Ranks of Top 20 attributes as identified by six hybrid FS algorithms

This section presents a list of the 20 most important attributes as identified by six hybrid FS algorithms. It can be seen that different features are identified by six different hybrid FS algorithms as given in Table 6. Justification for the variation of ranks is given in section 4.2.

Table 6:

Ranks of Top 20 attributes with Hybrid Methods

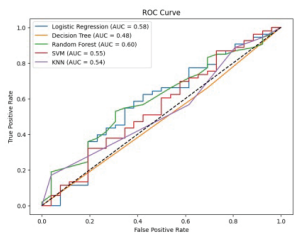
By Using Hybrid PSO Method		**Hybrid GA Method**		**Filter-Wrap Method**		**Wrapper-Embedded Method**		**Ensemble Method**		**Filter-Embedded Method**	
Feature	Rank	Feature	Rank	Feature	Rank	Feature	Rank	Feature	Rank	Feature	Rank
Age	15	School	16	School	19	Sex	16	Age	16	School	18
Address	19	Sex	17	Age	2	Age	11	Address	20	Sex	16
Famsize	4	Age	1	Address	14	Address	18	Famsize	18	Age	12
Pstatus	12	Address	2	Pstatus	12	Medu	4	Medu	3	Famsize	13
Medu	8	Famsize	18	Medu	16	Fedu	2	Fedu	4	Pstatus	19
Mjob	6	Pstatus	3	Fedu	9	Mjob	1	Mjob	2	Medu	2
Fjob	9	Medu	19	Mjob	20	Fjob	5	Fjob	7	Mjob	1
Reason	7	Fedu	20	Fjob	18	Reason	3	Reason	5	Guardian	3
Guardian	10	Mjob	4	Reason	17	Guardian	7	Guardian	6	Studytime	15
Traveltime	20	Traveltime	5	Traveltime	8	Studytime	9	Studytime	10	Failures	4
Failures	1	Studytime	6	Studytime	4	Failures	6	Failures	1	Schoolsup	14
Paid	3	Failures	7	Failures	3	Famsup	13	Famsup	11	Famsup	10
Activities	13	Schoolsup	8	Schoolsup	5	Paid	8	Paid	9	Paid	9
Nursery	18	Famsup	9	Activities	13	Activities	14	Activities	17	Activities	6
Internet	16	Activities	10	Nursery	10	Romantic	15	Romantic	13	Higher	17
Romantic	2	Internet	11	Higher	15	Famrel	19	Famrel	19	Romantic	5
Freetime	14	Freetime	12	Famrel	11	Freetime	20	Freetime	12	Famrel	8
Goout	5	Goout	13	Dalc	6	Goout	12	Goout	15	Freetime	11
Health	17	Dalc	14	Walc	7	Health	17	Health	14	Goout	7
Absences	11	Health	15	Health	1	Absences	10	Absences	8	Dalc	20

ROC Curves and Evaluation Matrix of Classification Models in Conjunction With Hybrid FS Models For Top 20 Features/ Attributes Of Data.

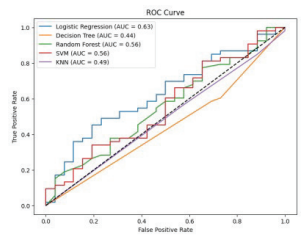
The results of five classifiers for the top 20 attributes as identified by different hybrid FS models are given in Table 7, and a graphical presentation of the ROC curve is shown in Figure 2 (a-f). Based on the values of accuracy, precision, recall, F1 Score, and AUC, and the ROC curve, the following can be inferred.

Fig 2:

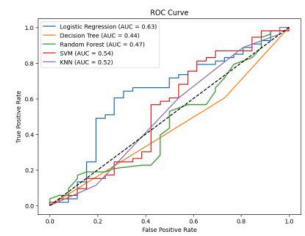
ROC Curves of five classification models and six hybrid FS with top 20 Features



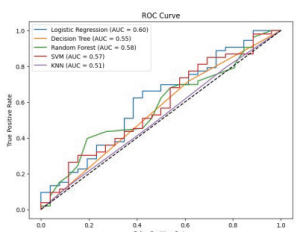
ROCC of Hybrid PSO Model
(a)



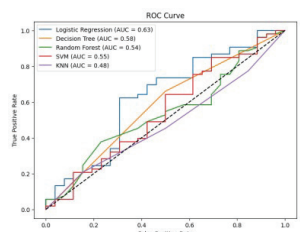
ROCC of Hybrid GA Model
(b)



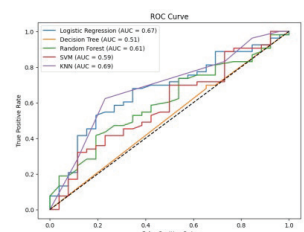
ROCC of Filter-Wrap Model
(c)



ROCC of Wrapper-Embedded Model
(d)



ROCC of Ensemble Model
(e)



ROCC of Filter-Embedded Model
(f)

It can be seen from the quality parameters given in Tables 5 and 7 that the values of precision, recall, and F1-score for the 20 top attributes are slightly higher in comparison to when all attributes are considered by the classification models. Therefore, it can be concluded that removing non-significant attributes not only improves the performance of classification models but also reduces the need for computing power. In addition, it reduces the need for capturing more data in the future (Dahiya et al., 2017; Singh & Singh, 2019). Secondly, the ROC curve had

improved for two FS methods, i.e., wrapper-embedded model and filter-embedded algorithms. It is evident from making a comparison between fig 1(d) and 2(d), and fig1(f) and 2(f).

Table 7:

Evaluation Metrics for Top 20 features using different Classification Algorithms

Hybrid PSO Model					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.658228	0.703125	0.879057	0.781310	0.582729
DT	0.544304	0.670377	0.660665	0.665486	0.484035
RF	0.658228	0.703125	0.889057	0.785235	0.600508
SVM	0.680126	0.680000	0.972264	0.800283	0.552975
KNN	0.698228	0.721176	0.896762	0.799441	0.535559
Hybrid GA Model					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.632911	0.691429	0.886792	0.777019	0.628447
DT	0.493671	0.692653	0.684906	0.688758	0.441219
RF	0.632911	0.728752	0.890189	0.801422	0.556967
SVM	0.698228	0.680556	0.924528	0.784001	0.558055
KNN	0.687595	0.711765	0.899057	0.794523	0.486575
Filter Wrap Method					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.645577	0.696056	0.90566	0.787144	0.626996
DT	0.493671	0.717451	0.673774	0.694928	0.436502
RF	0.607595	0.737419	0.832453	0.782060	0.469158
SVM	0.670886	0.675325	0.961132	0.793271	0.535559
KNN	0.658228	0.711176	0.886792	0.789334	0.516328
Wrapper-Embedded Method					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.670886	0.686567	0.867925	0.766666748	0.603774
DT	0.607595	0.708333	0.641509	0.673266937	0.550798
RF	0.645570	0.705882	0.905660	0.793388062	0.575472
SVM	0.670886	0.675325	0.901132	0.772056539	0.574746
KNN	0.607595	0.701493	0.826792	0.759006076	0.513425

Ensemble-Based Method					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.683544	0.705882	0.905660	0.793388	0.62627
DT	0.607595	0.729167	0.660377	0.693069	0.580189
RF	0.645570	0.686567	0.867925	0.766667	0.538099
SVM	0.683544	0.684211	0.981132	0.806202	0.545718
KNN	0.569620	0.690794	0.873585	0.771510	0.478955
Filter-Embedded Method					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.658228	0.685714	0.905660	0.780487	0.671263
DT	0.569620	0.679245	0.679245	0.679245	0.510160
RF	0.632911	0.714286	0.754717	0.733945	0.607402
SVM	0.645570	0.676056	0.905660	0.774193	0.590711
KNN	0.645570	0.698413	0.830189	0.758621	0.686865

ROCC and Evaluation Metrics of Classification Models in Conjunction With Top 15 Features Identified by Hybrid FS Models

This section presents the top 15 attributes of the data set which are a high association with the target variable. The attributes are shown in Table 8. It can be seen that there are differences in the ranks of the attributes due to inherent weaknesses of the black box approach of computing using the built-in Python subroutine. Secondly, all methods cannot be applied to all kinds of datasets, which is an inherent weakness of machine learning algorithms.

Table 8:

Name and Ranks of Top 15 attributes as identified with six Hybrid Methods

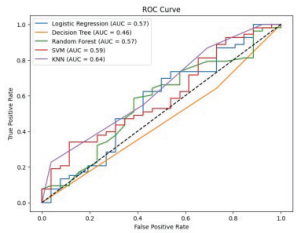
By Using Hybrid PSO Method		**By Using Hybrid GA Method**		**By Using Filter-Wrap Method**		**By Using Wrapper-Embedded Method**		**By Using Ensemble Method**		**By Using Filter-Embedded Method**	
Feature	Rank	Feature	Rank	Feature	Rank	Feature	Rank	Feature	Rank	Feature	Rank
Sex	12	Address	1	Age	2	Age	11	Medu	3	School	13
Famsize	3	Famsize	2	Address	14	Medu	4	Fedu	4	Sex	9
Pstatus	11	Pstatus	3	Pstatus	12	Fedu	2	Mjob	2	Age	11
Medu	8	Medu	4	Fedu	9	Mjob	1	Fjob	7	Famsize	6
Mjob	14	Fjob	5	Travel-time	8	Fjob	5	Reason	5	Guard-ian	1
Fjob	9	Guard-ian	6	Study-time	4	Reason	3	Guard-ian	6	Study-time	7
Reason	6	Study-time	7	Failures	3	Guard-ian	7	Study-time	10	Failures	8
Guard-ian	7	Failures	8	School-sup	5	Study-time	9	Failures	1	School-sup	12
Study-time	15	School-sup	9	Activi-ties	13	Failures	6	Famsup	11	Famsup	3
Failures	1	Paid	10	Nursery	10	Famsup	13	Paid	9	Paid	4
Paid	2	Activi-ties	11	Higher	15	Paid	8	Roman-tic	13	Higher	14
Activi-ties	10	Higher	12	Famrel	11	Activi-ties	14	Freetime	12	Roman-tic	5
Roman-tic	4	Roman-tic	13	Dalc	6	Roman-tic	15	Goout	15	Famrel	10
Goout	5	Famrel	14	Walc	7	Goout	12	Health	14	Goout	2
Absenc-es	13	Free-time	15	Health	1	Absenc-es	10	Absences	8	Dalc	15

The ROCC for five classification algorithms, i.e., LR, RF, DT, SVM, and KNN, along with different Hybrid FS algorithms wherein the top 15 features of the dataset are selected for analysis, are shown in Fig. 3(a-f).

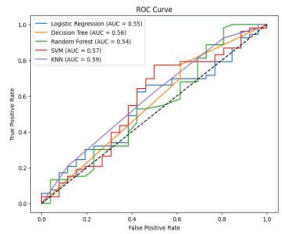
wing can be inferred.

Fig 3:

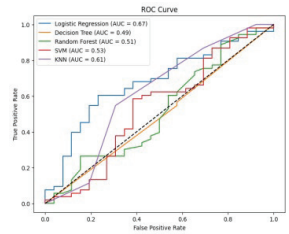
ROCC of five classification models and six hybrid FS with the top 15 Features



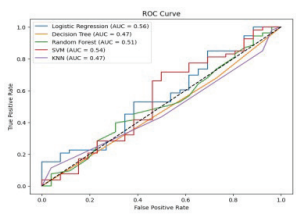
ROCC of Hybrid PSO Model
(a)



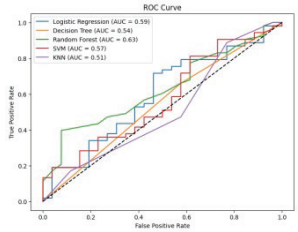
ROCC of Hybrid GA Model
(b)



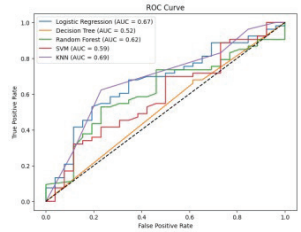
ROCC of Filter-Wrap Model
(c)



ROCC of Wrapper-Embedded Model
(d)



ROCC of Ensemble Model
(e)



ROCC of Filter-Embedded Model
(f)

The evaluation matrices with the top 15 ranked attributes are given in Table 9. It can be seen from the values of quality parameters that in the case of the hybrid PSO model of the FS algorithm, there is a slight improvement in the performance of the SVM classification algorithm. For the Hybrid GA model of FS, the performance of the decision tree has improved, but AUCC did not indicate better performance. Similar changes are seen for other FS models. ROCC had improved only for the filter-ensemble algorithm, as evident from Fig. 1(f), Fig. 2 (f), and Fig. 3(f).

Table 9:

Evaluation Metrics for Top 15 features of five Classification and six hybrid FS Algorithms

Hybrid PSO Model					
Algorithm Name	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.658228	0.696970	0.867925	0.773109	0.570392
DT	0.518987	0.641509	0.641509	0.641509	0.455372
RF	0.607595	0.677419	0.792453	0.730435	0.568215
SVM	0.683444	0.689189	0.962264	0.803155	0.592888
KNN	0.683544	0.718755	0.867925	0.786325	0.636067
Hybrid GA Model					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.64557	0.676056	0.90566	0.774194	0.549347
DT	0.594937	0.698113	0.698113	0.698113	0.555878
RF	0.670886	0.707692	0.867925	0.779661	0.541364
SVM	0.658228	0.680556	0.924528	0.784444	0.573295
KNN	0.683544	0.700000	0.924528	0.796748	0.594702
Filter Wrap Method					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.658228	0.680556	0.924528	0.784444	0.674891
DT	0.518987	0.666667	0.566038	0.612245	0.489115
RF	0.632911	0.687555	0.830189	0.752137	0.513062
SVM	0.658228	0.680556	0.924528	0.784444	0.533382
KNN	0.683544	0.718755	0.867925	0.786325	0.61357
Wrapper-Embedded Method					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.64557	0.692308	0.849057	0.762712	0.558055
DT	0.544304	0.654545	0.679245	0.666667	0.474238
RF	0.620253	0.676923	0.830189	0.745763	0.508345
SVM	0.683544	0.684211	0.981132	0.806202	0.544993
KNN	0.594937	0.652174	0.849057	0.737705	0.473149

Ensemble-Based Method					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.64557	0.686567	0.867925	0.766667	0.592888
DT	0.582278	0.723431	0.660377	0.690467	0.541727
RF	0.64557	0.686567	0.867925	0.766667	0.62881
SVM	0.670886	0.675325	0.981132	0.800000	0.570029
KNN	0.670886	0.701493	0.886792	0.783333	0.508345
Filter-Embedded Method					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.658228	0.685714	0.935660	0.791422	0.671263
DT	0.569621	0.686275	0.660377	0.673077	0.519956
RF	0.607595	0.689655	0.754717	0.720721	0.617562
SVM	0.645571	0.676056	0.905667	0.774196	0.590711
KNN	0.645572	0.698413	0.830189	0.758621	0.686865

ROC Curves and Evaluation Metrics of Classification Models in Conjunction With Top 10 Features Identified by Hybrid FS Models.

The ranks of the top 10 attributes with hybrid six FS algorithms are given in Table 10. It was expected that the top 10 out of 31 attributes would have more similarities in the ranks, but in this case, the different attributes are also identified as top 10 attributes by these six algorithms. The ROC curve for five classification algorithms along with six hybrid FS methods is given in Fig. 4 (a-f).

Table 8:

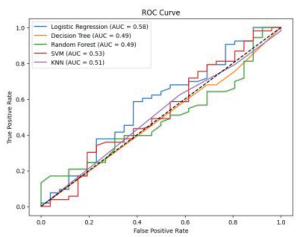
Showing Ranks of Top 10 attributes with Hybrid Methods

Hybrid PSO Method		**Hybrid GA Method**		**Filter-Wrap Method**		**Wrapper-Embedded Method**		**Ensemble Method**		**Filter-Embedded Method**	
Feature	Rank	Feature	Rank	Feature	Rank	Feature	Rank	Feature	Rank	Feature	Rank
Sex	6	Sex	1	Famsize	3	Medu	4	Medu	3	Sex	6
Famsize	3	Address	2	Fedu	2	Fedu	2	Fedu	4	Address	7
Mjob	7	Pstatus	3	Fjob	10	Mjob	1	Mjob	2	Famsize	5
Reason	10	Fedu	4	Reason	7	Fjob	5	Fjob	7	Mjob	1
Failures	1	Mjob	5	Travel-time	8	Reason	3	Reason	5	Failures	2

Paid	2	Fjob	6	Study-time	4	Guard-ian	7	Guard-ian	6	Activi-ties	3
Activi-ties	8	Reason	7	Failures	5	Study-time	9	Study-time	10	Famrel	4
Roman-tic	4	Guard-ian	8	Famsup	9	Failures	6	Failures	1	Free-time	8
Goout	5	Study-time	9	Paid	6	Paid	8	Paid	9	Goout	9
Health	9	Failures	10	Roman-tic	1	Absenc-es	10	Absences	8	Dalc	10

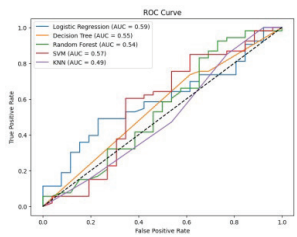
Fig 4:

ROCC of five classification models and six hybrid FS with top 10 Features



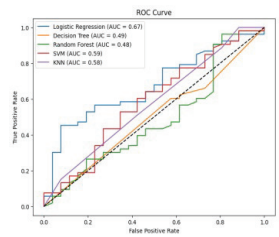
ROCC of Hybrid PSO Model

(a)



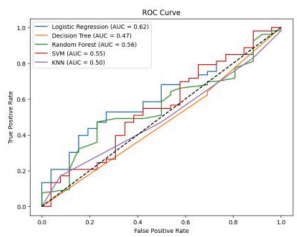
ROCC of Hybrid GA Model

(b)



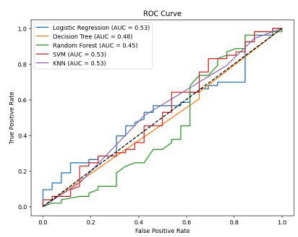
ROCC of Filter-Wrap Model

(c)



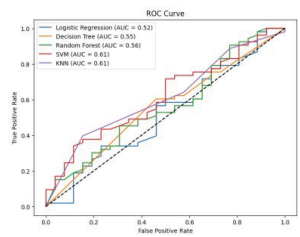
ROCC of Wrapper-Embedded Model

(d)



ROCC of Ensemble Model

(e)



ROCC of Filter-Embedded Model

(f)

It is evident from Figure 4(a-f) that the area under the ROCC curve has reduced when 15 or 20 attributes are considered for classification purposes. In addition, in many cases, the classification algorithm curves are below the straight line. This indicates that the performance of classification algorithms had not improved with FS for a small set of 10 features in place of 15, 20, or all features. Similar trends are exhibited by other parameters calculated and given in Table 11 for making a comparison in the performance of the five classification models.

Table 11:

Evaluation Metrics for Top 10 features using five Classification Algorithms and six hybrid FS methods

Hybrid PSO Model					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.670886	0.701493	0.886792	0.783333	0.575472
DT	0.544304	0.654545	0.679245	0.666667	0.485123
RF	0.556962	0.645161	0.754717	0.695652	0.489115
SVM	0.670886	0.684932	0.943396	0.793651	0.527576
KNN	0.594937	0.666667	0.792453	0.724138	0.512337
Hybrid GA Model					
Algorithm	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.645578	0.701493	0.886792	0.783333	0.589623
DT	0.620253	0.654545	0.679245	0.666666	0.554064
RF	0.620253	0.645161	0.754717	0.695652	0.542091
SVM	0.658228	0.684932	0.943396	0.793651	0.568578
KNN	0.648576	0.666667	0.792453	0.724138	0.493832
Filter Wrap Method					
Algorithm Name	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.658228	0.680556	0.924528	0.784001	0.674528
DT	0.544304	0.673469	0.622642	0.647059	0.49492
RF	0.582278	0.666667	0.754717	0.707965	0.484761
SVM	0.658228	0.680556	0.894528	0.773008	0.585269
KNN	0.658228	0.691176	0.886792	0.776859	0.57656

Wrapper-Embedded Method					
Algorithm Name	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.64557	0.676056	0.885661	0.766792	0.61611
DT	0.518987	0.647059	0.622642	0.634616	0.46807
RF	0.544304	0.644068	0.716981	0.678571	0.556967
SVM	0.683544	0.649189	0.912264	0.758565	0.552975
KNN	0.594937	0.656716	0.830189	0.733333	0.497097
Ensemble-Based Method					
Algorithm Name	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.658228	0.660556	0.904528	0.763526	0.530842
DT	0.518987	0.653061	0.603774	0.627451	0.481495
RF	0.607595	0.696429	0.735849	0.715597	0.448839
SVM	0.683544	0.689189	0.912264	0.785191	0.525036
KNN	0.607595	0.677419	0.792453	0.730435	0.534107
Filter-Embedded Method					
Algorithm Name	**ACC**	**PRE**	**REC**	**F1_S**	**AUCC**
LR	0.658228	0.620556	0.874528	0.725971	0.520682
DT	0.556962	0.687555	0.622642	0.65349	0.553701
RF	0.645571	0.698413	0.810189	0.75016	0.559507
SVM	0.658228	0.610556	0.874528	0.719082	0.614296
KNN	0.670886	0.701493	0.846792	0.767325	0.612482

It can be seen from the results presented in the earlier sections that the performance of five classification algorithms in conjunction with six hybrid FS algorithms has improved when the top 20 attributes are used. This is based on the values of quality parameters, i.e., precision, recall, and F1-score. Similar trends were seen for many combinations of classification and FS algorithms when the top 15 attributes are used in computation. Further, a reduction in the number of features did not result in an improvement in the goodness of fit of the classification algorithms. The results of ROCC did not indicate improvement in forecasting success of classifiers with a reduction in the number of features, except for when 20 features are selected for the purpose of classification. Accuracy is not always a good measure of the goodness of fit of models; therefore, it is not emphasized in making conclusions. However, there is a need to rerun the algorithm by changing the process of discretization.

Conclusion

It is always desirable to use FS algorithms before implementing machine learning/supervised classifiers, which enhances the values of goodness-of-fit parameters of models. However, it is right to pre-process the features of data as per the assumption/requirement of a given algorithm. Sikri et al. (2023) had shown a significant difference for the chi-square FS method in the rank of features when computation is done with prior right pre-processing of data with respect to the chi-square method of FS. It is, therefore, necessary to optimize even a simple operation, such as discretization or any other assumptions of the algorithm used for either FS or classification. A discrete predictor with four categories, when used with two categories (if needed as per the assumption of the algorithm), makes not only a difference in the ranking of the attributes but also impacts the performance parameters of classification models. Similarly, an artificial neural network as a predictor used with weights and biases between [-1 to +1] may give different results when the limit of inputs is changed to [0 to +1] or in any other range.

The summary of the best combination of the hybrid FS method and supervised classifiers is presented in Table 12. The results are the best of the ROC curve. It can, therefore, be concluded that Logistic Regression and the Filter wrapper method of FS are the best.

Table 12:

Combination of Feature Selection model and classifier with the Highest ROC

No. of Predictors/ Features	Classifiers	Hybrid FS Algorithm	Value of ROC Curve	Comments
30 (All)	Random Forest	Ensemble	0.64	Many Cases curve is below straight line
20	Logistic	Filter-Embedded	0.67	-do-
15	Logistic	Filter-Wrapper	0.67	-do-
10	Logistic	Filter Wrapper	0.67	-do-

REFERENCES

Ahonen, L., Cowley, B. U., Hellas, A., & Puolamaki, K. (2018). Biosignals reflect pair-dynamics in collaborative work: EDA and ECG study of pair-programming in a classroom environment. *Scientific Reports*, *8*, 3138. https://doi.org/10.1038/s41598-018-21248-1

Alberto, R., Alfonso, G. B., Guillermo, H., Javier, P., & Pablo, C. (2021). Artificial neural network analysis of the academic performance of students in virtual learning environments. *Neurocomputing*, *423*, 713–720. https://doi.org/10.1016/j.neucom.2020.08.105.

Alisha Sikri, N. P. Singh, & Surjeet Dalal (2023)**.** Chi-square method of feature selection: Impact of pre-processing of data**,** International Journal of Intelligent Systems and Applications in Engineering**,** 11(3s): 241-248.

Aluko, R. O., Daniel, E. I., Oshodi, O. S., Aigbavboa, C. O., & Abisuga, A. O. (2018). Towards reliable prediction of academic performance of architecture students using data mining techniques. *Journal of Engineering Design and Technology*, *16*(3), 385–397. https://doi.org/10.1108/JEDT-12-2017-0117.

Antoniou, P. E., Arfaras, G., Pandria, N., Athanasiou, A., Ntakakis, G., Babatsikos, E., & Bamidis, P. (2020). Biosensor real-time affective analytics in virtual and mixed reality medical education serious games: Cohort study. *JMIR Serious Games*, *8*(1), e17823. https://doi.org/10.2196/17823.

Asif, R., Merceron, A., Ali, S. A., & Haider, N. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, *113*, 177–194. https://doi.org/10.1016/j.compedu.2017.05.005.

Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting student dropout in higher education. In *Proceedings of the ICML Workshop on #Data4Good: Machine Learning in Social Good Applications*, New York, NY, USA, 24 June 2016. https://arxiv.org/pdf/1606.06364.pdf

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. In *Proceedings of the International Conference on Computing Networking and Informatics (ICCNI)*, Lagos, Nigeria, 29–31 October 2017.

Azizah, E. N., Pujianto, U., & Nugraha, E. (2018). Comparative performance between C4.5 and Naive Bayes classifiers in predicting student academic performance in a virtual learning environment. In *Proceedings of the 4th International Conference on Education and Technology (ICET)*, Malang, Indonesia, 26–28 October 2018 (pp. 18–22).

Baker, N. K., D'Mello, S., Ocumpaugh, J., Baker, R., & Shute, V. (2016). Using video to automatically detect learner affect in computer-enabled classrooms. *ACM Transactions on Interactive Intelligent Systems*, *6*(2), 1–26. https://doi.org/10.1145/2908190

Bohong, Y.; Zeping, Y.; Hong, L.; Yaqian, Z.; Jinkai, X. In-classroom learning analytics based on student behavior, topic and teaching characteristic mining. Pattern Recognit. Lett. 2020, 129, 224–231.

Byung-Hak, K., Ethan, V., & Ganapathi, V. (2018). GritNeikikik8t: Student performance prediction with deep learning. *arXiv*. https://arxiv.org/abs/1804.07405.

Cetintas, S., Si, L., Xin, P., & Hord, C. (2010). Automatic detection of off-task behaviors in intelligent tutoring systems with machine learning techniques. *IEEE Transactions on Learning Technologies*, *3*(3), 228–236. https://doi.org/10.1109/TLT.2010.29.

Chadrashekar, G., & Sahni, F. (2014). A survey on feature selection methods. Computers and Electrical Engineering, 40(1), 16-28. DOI: 10.1016/j.compeleng.2013.11.04.

Dahiya, Shashi; S. S. Handa, and N P Singh (2017). A feature selection enabled hybrid bagging algorithm for credit risk evaluation, Expert Systems, 34 (6): 1-11

Dash,M. & Liu, H. (1997). Feature selection for classification, Intelligent Data Analysis, 1(1), 131-156. DOI: 10.1016/s1088-467X(97)00008-5.

Daud, A., Aljohani, N. R., Abbasi, R. A., Lytras, M. D., Abbas, F., & Alowibdi, J. S. (2017). Predicting student performance using advanced learning analytics. In *Proceedings of the 26th International Conference on World Wide Web Companion*, Perth, Australia, 3–7 April 2017. https://doi.org/10.1145/3038912.3052588.

Farzaneh, A., & Fadlalla, A. (2017). Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems*, *24*, 32–58. https://doi.org/10.1016/j.accinf.2017.02.002

Fernandes, E. Holanda, M., Victorino, M., Borges, V., Carvalho, R., Van Erven, G. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil, Journal of Business Research, 94, 335-343.

Francis, B. K., & Babu, S. S. (2019). Predicting academic performance of students using a hybrid data mining approach. *Journal of Medical Systems*, *43*(6), 162. https://doi.org/10.1007/s10916-019-1350-7.

Garg, R. (2018). Predicting student performance of different regions of Punjab using classification techniques. *International Journal of Advanced Research in Computer Science*, *9*(5), 236–241. https://doi.org/10.26483/ijarcs.v9i5.6031.

Ghosh, I., & Chaudhuri, D.T. (2021). FEB-stacking and FEB-DNN models for stock trend prediction: A performance analysis for pre and post COVID-19 periods. *Decision Making: Applications in Management and Engineering*, *4*, 51–84. https://doi.org/10.5267/j.dmae.2021.2.004

Goldberg, P., Sümer, O., & Stürmer, K. (2021). Attentive or not? Toward a machine learning approach to assessing students' visible engagement in classroom instruction. *Educational Psychology*, *33*(1), 27–49. https://doi.org/10.1080/01443410.2021.1897027

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection, Journal of Machine Learning Research, 3.1157-1182.

Haiyang, L., Wang, Z., Benachour, P., & Tubman, P. (2018). A time series classification method for behaviour-based dropout prediction. In *Proceedings of the IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, Mumbai, India, 9–13 July 2018. https://doi.org/10.1109/ICALT.2018.00046.

Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U., & Sattar, M. U. (2020). Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Applied Sciences*, *10*(11), 3894. https://doi.org/10.3390/app10113894

Helal, S., Jiuyong, L., Lin, L., Esmaeil, E., Shane, D., Duncan, M., & Qi, L. (2018). Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, *161*, 134–146. https://doi.org/10.1016/j.knosys.2018.07.028.

Heuer, H., & Breiter, A. (2018). Student success prediction and the trade-off between big data and data minimization. In *DeLFI 2018—Die 16. E-Learning Fachtagung Informatik* (pp. 219–230). Gesellschaft für Informatik, Bonn.

Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017). Ouroboros: Early identification of at-risk students without models based on legacy data. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 6–15), New York, NY, USA, 13–17 March 2017. Association for Computing Machinery. https://doi.org/10.1145/3027385.3027408.

Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Computational Intelligence and Neuroscience*, *2018*, 6347186. https://doi.org/10.1155/2018/6347186.

Hussain, S., Dahan, N. A., Ba-Alwib, F., & Najoua, R. (2018). Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, *9*(2), 447–459. https://doi.org/10.11591/ijeecs.v9.i2.447-459.

Jackson, L. (2013). Get the 411: Laptops and tablets in the classroom, January 04. Education World. Retrieved on February 02, 2025 from https://www.educationworld.com/a_tech/tech/tech194.shtml.

Kemper, L., Vorhoff, G., & Wigger, B. U. (2020). Predicting student dropout: A machine learning approach. *European Journal of Higher Education*, *10*(1), 28–47. https://doi.org/10.1080/21568235.2019.1613427.

Koh, H., & Tan, G. (2005). Data mining applications in healthcare. Journal of Healthcare Information Management, 19(2), 64–72. https://pubmed.ncbi.nlm.nih.gov/

Lan, K., Wang, D., & Fong, S. (2018). A survey of data mining and deep learning in bioinformatics. *Journal of Medical Systems*, *42*(8), 139. https://doi.org/10.1007/s10916-018-0987-4

Li, F., Zhang, Y., Chen, M., & Gao, K. (2019). Which factors have the greatest impact on student's performance. *Journal of Physics: Conference Series*, *1288*, 012077. https://doi.org/10.1088/1742-6596/1288/1/012077.

Luhaybi, M. A., Tucker, A., & Yousefi, L. (2018). The prediction of student failure using classification methods: A case study. *Computer Science and Information Technology*, *2018*, 79–90. https://doi.org/10.5121/csit.2018.80906.

Magsi, H., Sodhro, A. H., Al-Rakhami, M. S., Zahid, N., Pirbhulal, S., & Wang, L. (2021). A novel adaptive battery-aware algorithm for data transmission in IoT-based healthcare applications. *Electronics*, *10*(3), 367. https://doi.org/10.3390/electronics10030367

Marte, J. (2014). Here's How Much Your High School Grades Predict Your Future Salary, May 20. Retrieved on January 21, 2025 from https://www.washingtonpost.com/ news/wonk/wp/2014/05/20/heres-how-much-your-high-school-grades-predict-how-much-you-make-today/.

Nagahi, M., Jaradat, R., Nagahisarchoghaei, M., Ghanbari, G., Poudyal, S., & Goerger, S. (2020). Effect of individual differences in predicting engineering students' performance: A case of education for sustainable development. In *Proceedings of the International Conference on Decision Aid Sciences and Applications (DASA)*, Online, 8–9 November 2020.

Nagahisarchoghaei, M., Dodd, J., Nagahi, M., Ghanbari, G., & Poudyal, S. (2020). Analysis of a warranty-based quality management system in the construction industry. In *Proceedings of the International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, Online, 26–27 October 2020.

Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017). A neural network approach for students' performance prediction. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)*, Vancouver, BC, Canada, 13–17 March 2017 (pp. 274–283). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027401

Rizvi, S., Rienties, B., & Khoja, S. A. (2019). The role of demographics in online learning: A decision tree-based approach. *Computers & Education*, *137*, 32–47. https://doi.org/10.1016/j.compedu.2019.04.010.

Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. Bioinformatics, 23(9), 2507-2517. DOI: 10.1093/bioinformatics/btm344.

Saeys, Y., Abeel, T., & Vande Peer, Y. (2008). Robust feature selection using ensemble feature selection techniques, Machine Learning and Knowledge Discovery in Databases (ECML.PKDD.2008), Lecture Notes in Computer Science, 5212, 313-325, Springer. DOI:10.1007/978-3-590-87481-2_21.

Shah, C., & Du, Q. (2021a). Spatial-aware collaboration-competition preserving graph embedding for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *19*(1), 1–5. https://doi.org/10.1109/LGRS.2021.3081070.

Shah, C., & Du, Q. (2021b). Modified structure-aware collaborative representation for hyperspectral image classification. In *Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Brussels, Belgium, 11–16 July 2021. https://doi.org/10.1109/IGARSS47720.2021.9554932.

Shah, C., & Du, Q. (2021c). Collaborative and low-rank graph for discriminant analysis of hyperspectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 5248–5259. https://doi.org/10.1109/JSTARS.2021.3082013.

Shah, C., Du, Q., and Xu, Y. (2022). Enhanced TabNet: Attentive interpretable tabular learning for hyperspectral image classification. *Remote Sensing*, *14*(3), 716. https://doi.org/10.3390/rs14030716

Singh, N.P. & Singh, D. (2019). Impact of feature selection methods on the performance of Credit Risk Classification Algorithm. Proceeding of IEEE 13th International Conference Application of Information and Communication Technologies 23-25, October 2019 | Baku, Azerbaijan, pp 101-106.

Sodhro, A. H., & Zahid, N. (2021). AI-enabled framework for fog computing driven e-healthcare applications. *Sensors*, *21*(21), 8039. https://doi.org/10.3390/s21248039

Vazan, P., Janikova, D., Tanuska, P., Kebisek, M., & Cervenanska, Z. (2017). Using data mining methods for manufacturing process control. *IFAC-PapersOnLine*, *50*(1), 6178–6183. https://doi.org/10.1016/j.ifacol.2017.08.111

Velazquez, R. (2023). Virtual reality in education: Benefits, uses, and examples, March 22. Retrieved on April 06, 2025 from https://soeonline.american.edu/blog/benefits-of-virtual-reality-in-education

Wang, W., Yu, H., & Miao, C. (2017). Deep model for dropout prediction in MOOCs. In *Proceedings of the 2nd International Conference on Crowd Science and Engineering (ICCSE'17)*, Beijing, China, 6–9 July 2017.

Wasif, M., Waheed, H., Aljohani, N. R., & Hassan, S. U. (2019). Understanding student learning behavior and predicting their performance. In *Cognitive Computing in Technology-Enhanced Learning* (pp. 1–28). IGI Global. https://doi.org/10.4018/978-1-5225-7191-7.ch001.

Whitehill, J., Mohan, K., Seaton, D., Rosen, Y., & Tingley, D. (2017). Delving deeper into MOOC student dropout prediction. *arXiv*. https://arxiv.org/abs/1702.06404.

Winstead, S. (2025). Using tablets in School: How to implement 1:1 technology using tablets in the classroom, April 06. My eLearning World. Retrieved on April 25, 2025, from https://myelearningworld.com/10-benefits-of-tablets-in-the-classroom/.

Xing, W., & Dongping, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, *57*(3), 547–570. https://doi.org/10.1177/0735633119831374

Xue, B. M., Browne, W.N., & Yao, X. (2016). A survey on evolutionary computation approaches to feature selection; IEEE Transactions on Evolutionary Computation, 20 (4), 606-626.

Zaletelj, J., & Košir, A. (2017). Predicting students' attention in the classroom from Kinect facial and body features. *Journal of Image and Video Processing*, *2017*, 80. https://doi.org/10.1186/s13640-017-0252-9.

Zauzmer, J. (2020). More students are learning on laptops and tablets in class. Some parents want to hit the off switch. *The Washington Post*. https://www.washingtonpost.com/local/education/more-students-are-learning-on-laptops-and-tablets-in-class-some-parents-want-to-hit-the-off-switch/2020/02/01/d53134d0-db1e-11e9-a688-303693fb4b0b_story.html

INTERNATIONAL JOURNAL OF TECHNICAL AND NATURAL SCIENCES

DOI: https://doi.org/10.69648/NQDI9754

International Journal of Natural and Technical Sciences (IJTNS), 2025; 5(1): 77-95

ijtns.ibupress.com

Online ISSN: 2671-3519

Application: 01.06.2025

Revision: 16.06.2025

Acceptance: 27.06.2025

Publication: 30.06.2025

Demiri, V. (2025). Cloud computing: Implementation, management, and security. International Journal of Natural and Technical Sciences, 5(1), 77-95. https://doi.org/10.69648/NQDI9754

Valdeta Demiri

International Balkan University, Skopje, North Macedonia

Correspondence concerning this article should be addressed to Valdeta Demiri

Email: valdetademiri2@gmail.com

BALKAN JOURNALS

Cloud Computing: Implementation, Management, and Security

Valdeta Demiri

Abstract

Cloud computing, a cutting-edge advancement in information and communication technology, offers significant advantages. Includes the flexibility of computing capabilities, unlimited processing power, and economic benefits for individuals and enterprises, as services are paid according to their use. Cloud services, while offering numerous benefits from information technology, must primarily be secure to be reliable.

Encryption methods can protect data; this approach highlights the limitations of cloud services, rendering them unreliable and inefficient in their offerings. All this makes data owners, especially enterprises, have doubts when their data storage is not under their control, which often occurs in the cloud.

This research paper aims to address challenges and dilemmas of cloud computing, providing information on its organisation, models, and standard services. It particularly focuses on the security aspects of storing enterprise data and the approach to service provider reliability.

The topic brings in key points for implementation, management, and security that enterprises should consider when selecting a cloud technology provider. Through field studies, this research paper demonstrates how the provider of these services fulfils these key points.

Keywords: Cloud Computing, Scalability, Multi-Tenancy, Privacy, SaaS

INTRODUCTION

Technologies such as clusters, networks, and, more recently, cloud computing have now become a reality. The purpose of their creation is to exploit these powerful and virtualized resources, apparently. As a whole, for various purposes of individuals, companies, and beyond. These information services offered are a good business model that utilises information power in public services, as gas, electricity, and telephone. Therefore, the consumer pays for what they consume.

Large computer companies use cloud computing. It describes the cloud-based services these companies provide. The primary purpose is to provide data storage, providing programs, for businesses from any place where they are located.

Cloud computing is the illusion of infinite resources, or the ability to pay "as much as you need" no more, no less; it is defined as: pay as you use, use whenever you need these computing resources (networks, servers, data, applications or services), get as much as you need, use when you want these resources.

- Pay per usage
- Management of the system as desired
- Virtualised services

Cloud computing also includes APIs, which are development programs that allow programmers to use or further develop the applications they create.

CLOUD COMPUTING

Cloud Computing is an essential term in the world of information technology. It is rapidly developing and attracting an increasing number of service providers as well as clients. The rapid growth has been made possible by the emergence of new computing technologies that allow the use of computer infrastructure and data storage at an affordable cost.

Cloud Computing is a computing model that incorporates the concept of "outsourcing" computing resources, which are scalable and provisioned on demand without requiring significant initial investments in IT infrastructure.

Several definitions attempt to define the term cloud computing. According to the American National Institute of Standards and Technology, it is a model that supports convenient services, such as networks, applications, and other services, that can be rapidly assigned or retrieved (Mather et al., 2009).

Although Cloud Computing has become familiar, experts believe that earlier forms of the Cloud have existed for decades.

- **Cloud 1.0** – The Cloud 1.0 network represents an abstract model of TCP/IP, where network communication tiers communicate with each other without knowing exactly who or where the other side is.
- **Cloud 2.0 –** represents data on the "World Wide Web", where various documents can be published or downloaded without knowing their exact path.
- **Cloud 3.0** – model is an abstraction of the complex infrastructure of servers etc. Simply put, Cloud 3.0 is considered an extension and generalisation of the Cloud 1.0 and 2.0 models

Concept cloud computing is becoming an integral part of today's companies, regardless of their industry or size. In recent years, a growing wave of companies has adopted the cloud to reduce costs, better manage their products and services, and scale their business more easily. The challenge that any company faces when considering incorporating cloud into its system is determining the skills needed to hire the person who will be responsible for it.

Since the concept of "cloud" is relatively new in the computing world, it is challenging to provide a definitive list of skills that a cloud analyst should possess. In recent years, cloud positions have included engineers, analysts, architects, as well as developers and administrators.

Many companies that utilised the web did not understand the role of developers. Additionally, they were unsure of the appropriate compensation for employees working in this sector. We are currently in a period of rapid evolution.

Many things still need to find their place. Since the cloud market is in transition, the list of necessary skills related to it cannot be clearly defined. Analysts predict that the position of cloud architect will be the first to be evaluated, due to the need to create a long-term architectural plan.

Cloud Computing means using many computers together to do tasks. Although cloud computing has existed for a while, IT teams now see it differently, because its flexibility lets them execute more services and applications in the back office. As shown in *Figure 1*, cloud computing offers flexible resources through virtualization.

Figure 1.

Cloud Computing

A Brief History of Cloud Computing

The term cloud computing is not a new one; it is used for many things nowadays. It has changed how we see the possibilities of using it today. The power and scale of the cloud have evolved significantly since its inception.

Over time, as technology and business environments have advanced, the status quo of cloud computing has changed. What was previously known as cloud computing was the same in principle, but the use of information today has undergone a significant scale of change.

Main Goals

Companies strive in order to reduce their operating expenses. Utilising technology can help address this by eliminating the need for significant capital investments. In the face of technological changes, it is beneficial to allow a cluster of servers to handle the processing needs of small companies.

A large number of companies have service-level agreements in place. For those unfamiliar with service level agreements, they are commonly referred to as SLAs (Service Level Agreements).

Service-Level Agreement

IaaS offers a certain quality of service. An SLA typically includes availability guarantees, ensuring uninterrupted service. Most IaaS providers focus their SLA on guaranteeing uptime and defining the guaranteed uptime during which the service will be available for a given period.

Most servers are in safe places, but it is important to be careful still. Considering that something with the cloud can go wrong. The possibility is always out of the question.

SLAs aim for a 99.99% uptime level but cannot guarantee that the data will consistently meet this level. This is something that many administrators need to keep in mind. SLAs play an important role in cloud computing and set the standard for uptime. While the cloud may always be online, the data it stores may not be accessible if an issue occurs. Although the chances of this happening are low, it can happen.

Therefore, it can be a key objective of cloud computing to provide SLAs with a certain level of uptime. The technical team should advise the CIO that this problem can create problems for supervisors. IT can be chosen easily, including data integrity of SLA, a thing which will help to prevent confusion or problems that arise from losing non-productive data.

Another advantage that cloud technology offers its users is shared processing power. Rather than constantly worrying about updating their hardware, the users can pay one provider to do that job for them. In most cases, the average user is not interested in or does not worry about the technical details of their computers. Most users work on finishing their job without having technical problems. Using this technology makes the job easier, especially for people who often deal with small problems, but technical problems slow down the work.

WHAT IS CLOUD COMPUTING?

In simple terms, it can be broken down into browser-based applications hosted on a remote server. For a normal user, that is all they need to know about cloud computing.

It is a way for small organisations to compete with large organisations, which means saving a significant amount of money, and a way to leverage energy efficiency in operations. Cloud computing is tied to the Internet technology that surrounds us.

Providers connect to the network as an extension of their machine. As long as a user is connected to the internet, the power of cloud computing comes into play, offering numerous benefits. One example would be processing power.

Notable feature, as the growing popularity of free storage space on servers will greatly benefit users in the future. A significant benefit of this is the prevention of data loss. With data management on a large number of computers on a network, the likelihood of data loss is significantly reduced.

BENEFITS OF CLOUD COMPUTING

Because the clients do not own the infrastructure in cloud computing, they can remove the previous costs and instead pay for what they use. So many services pay only for what they use. So many use the option "pay after use" similar to utility billing.

By sharing computing power between the users, they are used in an effective way, and servers pass less time sitting idle. It helps to lower the costs and speed up application development. Also, it gives access to the customers to more computing power when they need it, without having to build or maintain their systems for maximum requirements.

Increased availability of higher bandwidth speeds has also facilitated the model. Greater capacity, however, there are also issues to consider, due to the greater availability of increased bandwidth speeds.

There are significant benefits to using this technology. With new technology always there are new possible risks. Still, companies that need to offer cloud computing resources will gather several features, enabling organisations using cloud technology to be more efficient. As shown in *Figure 2,* it offers several key benefits.

Figure 2.

Benefits of Cloud Computing

FLEXIBILITY

With the model "server for rent", businesses have more flexibility in how they use the technology. Instead of buying and maintaining their servers, they adjust their resources, choosing the space or power they need at any time. By updating applications, the process can be significantly accelerated and made more efficient.

SCALABILITY

With cloud computing, one person can quickly transition from a small to a large scale. Organisations can be a good example if they can process large amounts of data at a given time and then return to normal without requiring significant resources from those servers. Often, it is more cost-effective for many organisations to rent mighty computing power than to purchase the equipment.

Cyclical and seasonal businesses can be a good fit for a "server rental" facility that serves cloud computing. A cyclical business, such as tax preparation, can utilise its resources during the first six months of the year when demand is high and then scale back its use when demand is low.

CAPITAL INVESTMENTS

Companies expecting a significant increase in the use of the coming years are expected to invest a large amount of money in the infrastructure for cloud computing. Most of them do not even know specifically how it will be used, but they are confident that the demand will exist in the coming years.

IT expenses consume a significant portion of the budget from money that companies can waste in different fields, like marketing, development, and human resources. Most basic IT purchases, such as hardware, will not be a problem as long as the tasks, assets, and functions are carried out.

PORTABILITY

In the global economy, organisations need to have people in the field, away from headquarters, to manage things. With cloud computing technology, they can utilise their computing power wherever their people are, as long as the users have access to it. Access to it is available wherever companies do business today, so this should not be an issue.

DISADVANTAGES OF CLOUD COMPUTING

Just like with any new and changing technology, cloud computing has potential downsides. Such problems should be taken seriously by organisations that rely on outside help for complex tasks. IT experts should always keep in mind potential downsides that could develop in the future.

RELABILITY

Companies that offer cloud services should inform the users that they are trustworthy, just like electricity. Power is always trustworthy and available; cloud providers should think about doing the same thing. It is important for businesses to rely on the cloud to finish the main parts of their work. To improve reliability, one could plan around hosting the most crucial functions in-house, while delegating most non-critical processing to a vendor's cloud.

There are some worries about data. It would be helpful to know where to save the data, as many providers of cloud prefer not to share this information. You can

make this location information an important part of your deal with a provider. If you are ready to work with different terms, most vendors would agree to share this. If they do not share where the data is being stored, this is a sign that they do not know themselves, but it is not a good sign.

SECURITY

Keeping essential data safe is always an important topic, but with the increase in technology, the information is being spread faster. Using rare clients could potentially be compromised. SLAs should state how cloud providers will protect the data.

If companies are not cautious, this can be turned into a dangerous risk. With frequent reports of lost data or stolen data, and mishandled data, they could seriously undermine efforts to adopt cloud technology (Velte et al., 2010).

The concept of cloud is a term coined to make people more comfortable. However, while providers can quickly understand what that means, technologies can still pose concerns about the ability to secure private data that shouldn't be exposed. *Figure 3* shows how cloud security protects online data.

Figure 3.

Security in the Cloud

FEW OR NO REFERENCES

Due to privacy concerns, cloud vendors are largely incapable of providing use cases of companies using their services. Only a small number of big companies can report their cloud computing practices.

As a result, many organizations are uneasy about using cloud computing resources, despite it being the most popular term in the tech world.

Drawbacks of this technology are made worse by the fact that fewer companies are known to be using the technology due to some internal problems with the cloud. This means that small start-up companies will likely need to leverage some of the big ones before they begin adopting the cloud.

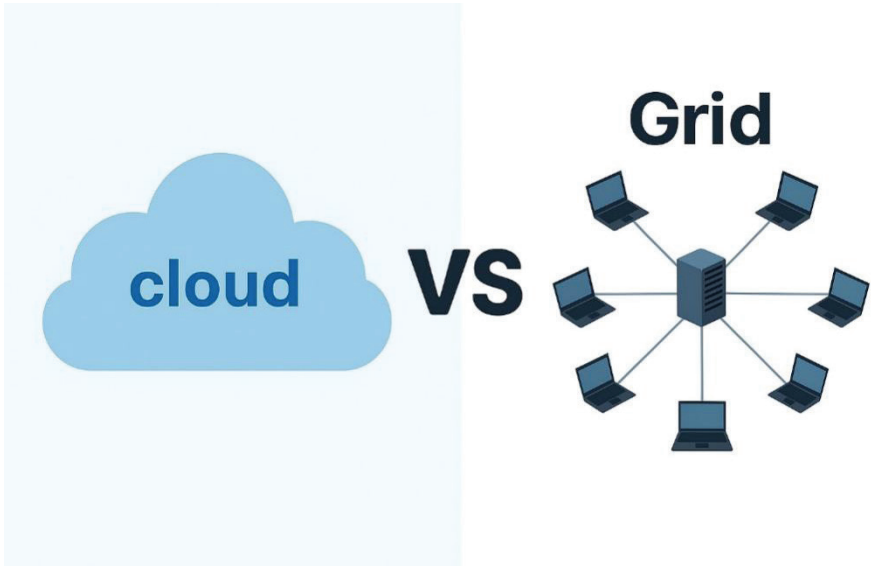
GRID COMPUTING VS CLOUD COMPUTING

Grid Computing is often misunderstood with cloud computing, but they are not the same thing. Grid computing is the same as cloud computing; they work together like one big virtual supercomputer to handle large tasks. Today, so many cloud platforms use cloud computing network behind the scenes to offer "pay per use. However, it is essential to recognise that cloud computing is not merely a continuation of grid computing, but a next-generation evolution of the network services model, leading to significant advancements in the field (Hugos & Hulitzky, 2011).

Commercial service platforms are another type of Sass. They give users an important online role to have access to different services. These platforms are often used in financial systems or for reservations. For example, the users can reserve flights or other services, and the system decides automatically the price and arranges everything based on what the user needs. In Figure 4, the differences between grid and cloud computing are highlighted.

Figure 4.

Grid computing or Cloud computing



CLOUD WEB SERVICES

How Cloud Computıng Works and What It Offers

Key concepts in cloud computing are "virtual" or "virtualisation". These infrastructures do not provide physical access to the devices that are being used. Instead, the entire infrastructure is built on a software system controller for the hardware of the computer as a whole (such as those from manufacturers VMware, Microsoft Hyper-V, or Citrix).

These platforms create several virtual units from a physical unit, offering them as services, platforms, or even virtualised infrastructures. These virtual units are utilised by businesses offering services, with a payment typically calculated based on the time used, whether in the form of virtual devices or services. For this reason, cloud computing is often defined as a service that "pays as you use" (pay-as-you-go).

Other types of series are web services, which originated from the development of technology based on cloud infrastructure, but whose development and ownership costs have always been and remain high.

Communication as a Service (CaaS)

Technology that allows one or more users to use it. Creating this model follows progression, considering that in this model, we also have software for communication provided as a service. However, a vital part of this model is also the hardware for communication and the guarantee of service quality. This model has initiated the process of evolution within the telecommunications industry, unlike the SaaS model, which originated from the software-as-a-service sector.

The providers of these types of solutions take care of managing the hardware; therefore, this model is based on a specific focus on developing the conditions for efficient communication. Communication in this model can be created based on IP, sending messages in the moment (instant messaging), applications for interaction, and video-audio conferencing, and others.

CaaS lets small businesses stay flexible without the high costs of doing it alone. Service providers are typically prepared to optimise the load points for their customers by offering services that enable increased capacity, device support, modes, or area coverage, as required by their customers.

Group function can change dynamically, ensuring the feature keeps peace with customers that resource providers will not have the service stop working.

CaaS almost manages itself and is managed by the customer. This eliminates the need for the business customer to make any capital investment in infrastructure. With CaaS, customers can upgrade their communications services without having to build facilities tailored to their specific needs. This allows these customers to allocate their budget and personal resources where their business can utilise them most effectively.

Infrastructure as a Service (IaaS)

IaaS is a model based on the cloud that offers users access to computing resources online, offering a cost-effective solution for businesses. This model incorporates the physical technology essential for building all models, including this one. Even companies responsible for providing service models have this type of service as the basis for developing their business, making it a financially wise choice for many.

According to Wikipedia, IaaS is a cloud model where computing resources like platforms and virtual machines are offered to users as a service. IaaS allows organi-

zations to go into the infrastructure of IT on demand, supported by large investments in technology, services, and data centre systems.

Compared to traditional outsourcing, which demands comprehensive due diligence and tools during the contract, IaaS centres on a service delivery model that provides unspecified, standardised infrastructure specifically for customer applications.

Modern On-Demand Computing

On-demand computing (ODC) is a system where companies take resources only when they need them. These resources can include saving, the processing power, and the software. ODC is useful mostly for projects that are short and for everyday tasks. It was developed to help organizations handle changing computing demands efficiently, since resource requirements often rise and fall over time.

MAIN CHARACTERISTICS OF PAAS

The main features of PaaS include services to develop, test, distribute, run and control applications, supporting their entire development lifecycle.

Support for shared user architecture helps solve developer concerns about the application's use by multiple users. Paas providers often include tools for managing and securing applications, as well as connecting with web services and databases.

The capacity to build deploy, undefined, or distributed teams greatly increases the productivity of PaaS offerings. PaaS helps to understand how applications function and how users behave using these control panels, which tell metrics how the performance and activity of the user is. Some PaaS offerings utilise a pay-per-use billing model.

SOFTWARE AS A SERVICE (SAAS)

Application sits in the top layer of the cloud. Users can reach these services from portals on the internet provided by this layer. Instead of using installed programs on their computers, users now have access to the same functions on the internet. For example, traditional applications such as Word, Excel, and PowerPoint are available as web services.

This model is called "Software as a Service"; it lets users avoid dealing with installation updates and testing. Salesforce.com represents a SaaS model, providing the possibility of using business applications that are deployed, allowing users to adapt the application at any time and from any location.

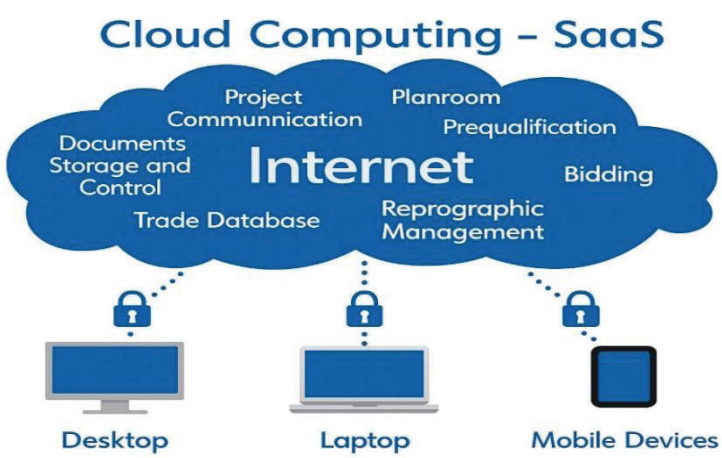
Different from traditional software, SaaS is hosted online and accessed via the internet. SaaS is hosted online and has access from the internet. As new web technologies, SaaS continues to grow in popularity, and often it offers subscription plans. Cable service has become easier to access, allowing users to access it from numerous locations worldwide. The significant ISP to increase bandwidth, combined with the continued introduction of increasingly powerful microprocessors and inexpensive storage devices, is providing a robust platform for the design, distribution, and use of software in all areas of business and personal computing. SaaS is mostly used to offer software to business clients at a low cost. It gives them the same benefits as traditional software without installation, management, or a high upfront cost. Most of the customers do not care about the details of software deployment; they just need to use it in their work. Many types of software can be offered through SaaS, including accounting, email, human resources, security in IT, and web content management.

The primary difference between SaaS and previous Internet-delivered solutions is that SaaS solutions are designed to run specifically within a browser. SaaS applications function directly in a browser. Their architecture is designed to support many users at the same time.

This shows significant differences from traditional clients with server services or service provider solutions, which are typically based on catering to a captive audience. How SaaS delivers software through the cloud is shown in Figure 5.

Figure 5.

Software as a Service



IMPLEMENTATION ISSUES

A variety of software components and application frameworks can be utilised in the development of SaaS applications. Using components and application frameworks can lower the time and the cost of a traditional product into a SaaS solution. According to Microsoft, Saas can be grouped into four maturity levels. Main characteristics are ease of configuration and scalability. Each level is built on the previous one by adding one of these characteristics.

SaaS Architectural Maturity Level 1 - Ad-Hoc/Custom

The first maturity level does not have real maturity. Every client gets a version to personalize the application, which functions in its instance on the hosting servers. Moving traditional non-network applications to this level usually requires less development, and it can lower the operational costs by combining the management of hardware and software.

SaaS Architectural Maturity Level 2 – Configurability

The second level allows the clients to use separate instances in the same application with configurable settings. This enables vendors to cater to the diverse needs of each customer using detailed settings. It also allows vendors to ease the maintenance burden by updating the shared code base.

SaaS Architectural Maturity Level 3 - Multi-Tenancy Efficiency

At the third level, multi-tendency allows one application to be used by all clients in an effective way, with a noticeable impact on end users; however, it ultimately has a limited ability to scale.

SaaS Architectural Maturity Level 4 – Scalability

The fourth level of maturity, scalable, improves from a shared user architecture. It can run so many copies of the application in hundreds or thousands of servers. Servers can be added or removed based on the requirements, without requiring further changes to the application software architecture.

Privacy

So many people believe that privacy is just part of the information. Even though they are strongly connected, privacy has its own unique issues and challenges. In

this chapter, these aspects in contact with traditional informatics will be discussed (Rittinghouse & Ransome, 2010).

Especially in the industry where there are fewer regulations outside of healthcare and financial services, the responsibility of saving privacy is wrongly placed on the service provider instead of the business unit that owns the data. Often, privacy is treated as an optional requirement, one box to check among many other important obligations.

What is privacy?

In this chapter, these aspects will be examined more concretely in the context of cloud computing, and we will compare those models with the traditional computing models. Especially, in industries that are less strictly regulated outside healthcare and financial services, to save privacy, which is often wrongly placed by the cloud service provider instead of the business unit that owns the real data. Often, privacy is treated as an optional requirement, in a box among many other important regulations.

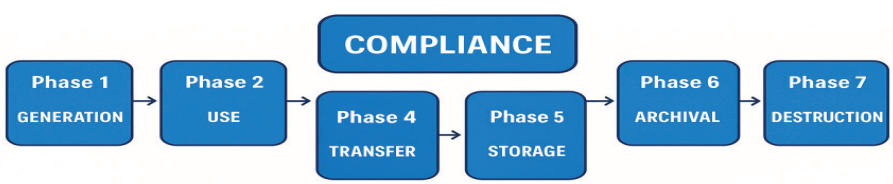
What is the data lifecycle?

Data lifecycle refers to the phases through which personal data and organizations pass, from the moment they are created till the moment they reach their form or it is no longer needed.

Personal data should be managed carefully this entire cycle. In cloud environments, it is important to consider how each stage of the lifecycle may be affected, so that the protection of personal information is maintained at every possible point, as illustrated in *Figure 6* below.

Figure 6.

Cycle of data KPMG



What are the main privacy concerns in the cloud?

Privacy advocates have pointed out so many concerns about cloud computing. These often include the security and the privacy. Except that, some other important issues to keep in mind are: People, personal information which are collected have the right to know exactly the information that an organization holds about them. This right is important, especially when it comes to marketing activities. In some places, marketing is being regulated differently from using data, and marketing efforts usually aim at individual end users in accordance with the organization's privacy policies.

When the data is saved or processed in the cloud, a key concern is whether the organization can fully meet the rights. Specifically, the organization must be able to fulfill these rights. Specifically, the organization should be able to give the subject access to the data in which all the information is saved when they request it. If a person decides to exercise this right, the organization must have processes in place to respond properly to the person's data handled according to the law and policies. Is it possible for a cloud provider to fully erase all the data when an organization requests it?

Compliance: Which rules are applied in the cloud? Which laws, standards, and contracts cover these data, and who is responsible for following them? How has the shift to the cloud impacted existing privacy compliance requirements? The cloud may traverse multiple jurisdictions; for example, data may be stored in various locations across multiple states within the United States.

Storage: Where is data saved in the cloud? Has it been moved to the centre of data or to another place? Are they saved together with the other organizations that use the cloud? Privacy laws change based on the location, and they can limit an organisation's ability to transfer specific types of personal information to other places. When data is saved in the cloud, transfers can sometimes happen without the organization realizing it, which can lead to violations of local privacy laws.

Is the Cloud Model Trustworthy?

Most of modern cloud computing is built on this trusted high-level platform that functions in servers using different types of virtualized technology. These services are provided through large data centres that operate under service level agreements to secure consistent performance. Cloud providers design their offers to meet the customer requirements of the clients, including specific levels of service

that users expect. From the user perspective, the cloud is like a single point of access for all computing tasks. These services can be used anywhere in the world as long as there is an internet connection. Cloud growth has also been supported by open standards and open-source software, which can make it easier for services to work together.

CONCLUSION

Cloud computing is currently the most widely used concept in the IT world, encompassing both the education and professional sectors. It is one of the most important transformations the IT field has ever undergone. As organizations embrace cloud adoption, they enter a new environment that relies on the trust of the provided services.

Cloud has already helped companies increase their competitiveness today and will continue to play a crucial role in sustaining it in the future. Those who continue to reject cloud solutions, viewing them as inflexible, are likely to fail due to their lack of skills and the associated cost burdens.

While cloud computing is truly very important and you might already be using it, whether for personal tools or in your business, here is an overview of the benefits and drawbacks we have learned from this research. Cloud computing is a truly affordable way to access all the resources they need in one place. It is a more efficient way to spread out your resources, making it easier to access things from a distance.

However, is cloud computing truly secure in every way? For example, even though cloud computing includes backups, if something goes wrong, will you have your backup? Is it as private as you think it is? Are your data truly secure, or could someone else gain access?

The primary concern with cloud computing stems from the feeling of losing control over protected information. Actual measures do not fully address the challenges of storing and processing data with cloud providers from third parties. Best practices recommend extending enterprise control measures to the cloud environment by using trusted methods. Implementing these techniques can help in reducing the concerns associated with cloud computing today, because they have the potential to improve data security and privacy in the cloud to ensure proven advantages in business innovation through cloud participation.

These approaches also address potential issues and abuses that may arise from increased reliance on cloud computing, as well as strategies for maintaining security against such attacks.

Specifically, new threats require new measures to preserve and enhance security. Among these are tools to monitor and understand privacy leaks, perform authentication, and ensure service availability against cloud-based attacks.

REFERENCES

Hugos, M., & Hulitzky, D. (2011). *Business in the cloud.* John Wiley & Sons, Inc.

Mather, T., Kumaraswamy, S., & Latif, S. (2009). *Cloud security and privacy.* O'Reilly Media.

Rittinghouse, J. W., & Ransome, J. F**.** (2010). *Cloud computing: Implementation, management, and security.* [s.n.].

Velte, A. T., Velte, T. J., & Elsenpeter, R. (2010). *Cloud computing: A practical approach.* The McGraw-Hill

IJTNS INTERNATIONAL JOURNAL OF TECHNICAL AND NATURAL SCIENCES

BALKAN JOURNALS

VOLUME 5 | ISSUE 1 | JUNE 2025 ijtns.ibupress.com