Vol. 9 No. 4
August 2025

Opportunities Identification & Problem Solutions

INTERNATIONAL JOURNAL OF SYSTEMATIC INNOVATION

ACCSCIENCE PUBLISHING

ISSN(Print): 2077-7973 ISSN (Online): 2077-8767
DOI: 10.6977/IJoSI.202508_9(4)

https://ijosi.org/index.php/IJOSI

ISSN (Print): 2077-7973
ISSN (Online): 2077-8767
DOI: 10.6977/IJoSI.202508_9(4)

The International Journal of Systematic Innovation

ISSN (Print): 2077-7973
ISSN (Online): 2077-8767
DOI: 10.6977/IJoSI.202508_9(4)

INTERNATIONAL JOURNAL OF SYSTEMATIC INNOVATION

CONTENTS

AUGUST 2025 VOLUME 9 ISSUE 4

FULL PAPERS

DOI: 10.6977/IJoSI.202508_9(4).0001
J. Deng/Int. J. Systematic Innovation, 9(4), 1-12(2025)

TRIZ reverse-based new application identification for low-density, high-strength thin cement sheets

Jyhjeng Deng*

Department of Accounting and Information Management, Da-Yeh University, Taiwan

*Corresponding author E-mail: jdeng@mail.dyu.edu.tw

(Received 07 January 2025; Final version received 07 March 2025; Accepted 23 March 2025)

Abstract

The theory of inventive problem solving (TRIZ) reverse is a process in which known solutions are used as a basis for seeking new problems or applications and involves steps such as back-tracing product strengths to inventive principles, selecting catchwords, conducting database searches, analyzing patent lists, and identifying opportunities for patent exploitation. This paper explores the application of TRIZ reverse methodology to the identification of new markets for patented products, with a focus on low-density, high-strength thin cement sheets. Although research in the literature has used these methods, the details are often kept confidential. This study bridges that gap by offering a detailed examination of the TRIZ reverse process and its application to cement sheets, with specific demonstrations of patent search commands and a discussion of potential exploitation avenues. The insights provided here can facilitate a broader understanding and implementation of TRIZ reverse, thus empowering researchers to identify untapped market opportunities for existing technologies. The search results revealed various potential applications of thin cement sheets beyond solar panels. These include construction materials such as outer skins for wall panels, siding, partition walls, roofing, and ceilings; food industry applications such as food containers and boxes; wet-area lining boards; and smoke containment curtains for public buildings, including hotels, restaurants, prisons, hospitals, airports, and aircraft.

Keywords: Catchwords, High-Strength Thin Cement Sheets, Low-Density, New Market Identification, Patent Exploitation Opportunities, TRIZ Reverse.

1. Introduction

The concept of open innovation, proposed by Chesbrough (2003), suggests that in-house technologies or patents can be extended beyond their original domain by licensing them to other companies. However, there is no clear methodology for identifying new markets outside the primary domain of a given technology. Common practices in business research include market segmentation, which involves demographic analysis, geographic segmentation, psychographic segmentation, and behavioral segmentation, which are typically followed by studies of customer needs and pain points through surveys, questionnaires, focus groups, and customer feedback (Kotler et al., 2022). This process often continues until a business model canvas is built (Osterwalder & Pigneur, 2010). However, the identification of new markets largely relies on the subjective judgment of market researchers, and no systematic approach has been developed.

The theory of inventive problem solving (TRIZ) was developed by Altshuller (1984; 1996) to address inventive problems that involve contradictions in properties or functions. Conventionally, such problems have been handled by finding compromises between conflicting properties, neither of which is fully optimized. However, TRIZ takes a different approach by introducing an additional property to resolve the conflict between the original properties. For instance, the properties of strength and the requirement for a lightweight product may conflict in the design of a table (Dewulf, 2005). TRIZ resolves this by adding a property such as porosity or fragmentation, leading to a solution such as a porous or segmented table that is both strong and lightweight.

Whereas TRIZ follows a “problem seeks solution” methodology, TRIZ reverse applies a “solution seeks problem” approach, which involves identifying new potential applications or problem areas based on a known technical solution. Several researchers have contributed to the development of procedures for implementing TRIZ reverse (Glaser & Miecznik, 2009; Bianchi et al., 2010; Günther et al., 2021; Günther & Popova, 2022; Dewulf et al., 2023). For example, Glaser and Miecznik (2009) proposed a six-step process for applying TRIZ reverse to a fully implantable distraction system, a device that gradually increases the distance between two bone segments using a telescopic implement (Wittenstein, 2006). There were numerous steps in their process, including constructing contradiction pairs, selecting patent database search terms, and choosing the most promising applications.

Bianchi et al. (2010) later introduced a five-step procedure for searching alternative technology applications (ATAs), another term for identifying new markets. Their process began with defining the requirements for the technology, followed by TRIZ-based analysis, abstraction, and patent database searches. However, the specifics of the patented technology, patent search commands, and twenty applications were omitted.

Subsequent research by Günther et al. (2021) and Günther and Popova (2022) refined these processes. These authors developed a seven-step method for systematic knowledge and technology transfer and included detailed procedures for creating search commands for the German patent database (DPMA), although these were limited to specific patents, and the researchers did not address the broader issue of industrial applications.

Dewulf et al. (2023) simplified their process to give a five-step procedure, using software tools such as Patent Inspiration (AULIVE, 2024) to automate parts of the analysis. However, challenges in terms of implementation, particularly regarding the identification of suitable industries for technology transfer, remained unaddressed.

Based on previous research on TRIZ reverse, the TRIZ reverse procedure can be depicted as in Fig. 1.

Given the limitations of previous research, this paper aims to bridge the current gap in the area of TRIZ reverse implementation using low-density high-strength thin cement sheets as an example. There are two objectives in this study: first, we aim to provide a detailed procedure for TRIZ reverse that can be easily applied using the European Patent Database, and secondly, we aim to discuss the potential pitfalls of current methodologies in identifying promising applications and proposing remedies. The remainder of the paper is organized as follows: section 2 presents a review of the literature, section 3 describes the application of our TRIZ reverse method, and Section 4 concludes the study.

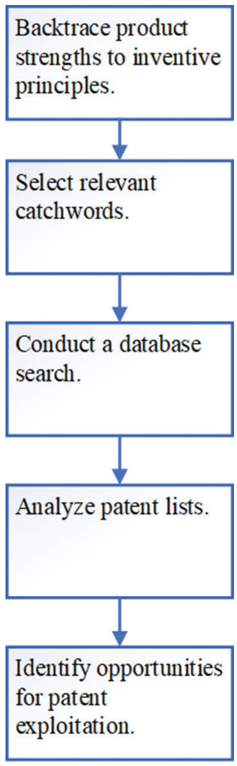


Fig. 1. TRIZ reverse procedure

2. Literature Review

Understanding new markets for innovative products is challenging for many large firms, due to the need for cohesive collaboration among various departments (Dougherty, 1990). Established theory suggests that successful market identification relies on the integration of insights from planners, technical teams, field operations, and manufacturing departments. Case studies have demonstrated that when these groups work together organically, a more accurate understanding of new markets can be formed through effective communication. In contrast, product failures often result from siloed relationships, where information is communicated in a linear, rigid manner, as this leads to frequent miscommunications.

Despite this finding, business professionals often rely on their judgment and intuition when exploring new markets, which lacks a systematic approach. Common tools such as questionnaires, interviews, and focus groups are used to gather direct insights from potential customers (Kotler et al., 2022), but this approach tends to target familiar types of customers, driven by preconceived notions, and often overlooks potential customers in other less obvious fields.

To overcome this psychological inertia, the TRIZ reverse methodology can be applied. By abstracting the product, the TRIZ reverse approach helps to identify potential customers in other industries. As illustrated in the context of ATAs by Bianchi et al. (2010), the TRIZ approach can be reversed: rather than progressing from a specific problem to an abstract problem, then progress from there to an abstract solution, and finally, to a specific solution, TRIZ reverse starts with a specific solution and works backward to a specific problem through abstract solutions and problems. Since all TRIZ tools are derived from patent databases, these databases can be used in accordance with TRIZ tools such as function analysis, evolution potential analysis, the 40 inventive principles, and frequently used terms in patents or research papers, to identify possible applications in other industries.

Günther et al. (2021) attempted to refine this process by providing a specific search command for the DPMA regarding a patented technology (DE10,2017,123,891 B4) of a collagen-based layer material (Harre et al., 2019). Their search command was constructed to identify patents discussing a layered arrangement or process involving collagen materials, with descriptions of properties such as pressure, temperature, and density, and where segmentation or bonding were ensured. Although the search command was effective in retrieving the patent itself, the study did not specify the industry in which the technology would be licensed, thus limiting its broader application.

To facilitate patent retrieval in DPMA, Günther et al. (2021) developed a search command tailored to the patented technology (DE10,2017,123,891), which pertains to a method for obtaining a collagen-based layer material. The command was created as follows: (BI=(Schichtanordnung UND Schichten ODER Sichtmaterial ODER Verfahren UND Kollagenmaterial)) UND (BI=(Eigenschaft? UND Druck? ODER Temperatur? UND Dichte?)) UND (BI= (segmentieren ODER zerlegen ODER teilen)) UND (BI= (Verbund? ODER Verbindung?)).

In this context, "BI" refers to "bibliographic information," indicating that the search command is applied to all fields in the patent document. The command "Schichtanordnung UND Schichten ODER Sichtmaterial ODER Verfahren UND Kollagenmaterial" translates to "Schichtanordnung" (layer arrangement) AND "Schichten" (layers) OR "Sichtmaterial" (visible material) OR "Verfahren" (process) AND "Kollagenmaterial" (collagen material). The command "Eigenschaft? UND Druck? ODER Temperatur? UND Dichte?" translates to "Eigenschaft?" (property?) AND "Druck?" (pressure?) OR "Temperatur?" (temperature?) AND "Dichte?" (density?). The question mark "?" is used as a wildcard that allows for variations of the word (e.g., "Eigenschaften" could be matched). The command "segmentieren ODER zerlegen ODER teilen" searches for documents containing any of these terms related to division or segmentation: "segmentieren" (segment), "zerlegen" (disassemble), or "teilen" (divide). The command "(Verbund? ODER Verbindung?)" searches for words related to composite or connection, such as "Verbund?" (composite?) or "Verbindung?" (connection?). In summary, this search command was designed to find patents that include a discussion of a layered arrangement or process involving collagen materials while also describing certain properties such as pressure, temperature, and density. The search also includes terms related to segmenting or dividing materials and ensuring that those segments form a composite or connection. The use of "?" as a wildcard enabled the search to capture variations of the base words, allowing for broader results.

Note that the five German words in the first BI are derived from the five most frequently used terms in the patent text, which can be retrieved using Voyant Tools (2024). The terms in the other BIs are based on the catchwords for the top three inventive principles: (35) property changes, (1) segmentation, and (40) composite materials (Mann, 2006).

We entered the command in expert mode on DPMA (Fig. 2), resulting in 622 hits, as shown in Fig. 3. Note that in Fig. 2, the top three inventive principles are listed beside the command lines, indicating that command lines 2–4 correspond to these three principles. Note that DE10,2017,123,891 appears as item 376 in the search list (Fig. 4), demonstrating that the command successfully retrieves its patent. However, this paper does not specify the industry in which the patented technology will be licensed. Only rough IPC codes, including section, class, and subclass, are provided.

Günther and Popova (2022) used a similar approach for technology transfer with a patented self-synchronizable network (US 10,241,539 B2) (Wetzel et al., 2019), and crafted a search command to find patents that mentioned specific technical terms. Despite some success, their approach faced challenges in terms of identifying broader industry applications.

The command they used is repeated here for reference: (BI=(Phase ODER Netzwerk ODER Signal ODER Verzögerung ODER Knoten)) UND (BI=(Vibrieren UND Welle ODER Hz UND Gigahertz)) UND (BI=(Elektrisch ODER Atom ODER Drahtlos UND Feld)) UND (BI=(Muster UND UV ODER Synch ODER Frequenz)).

The command "(BI=(Phase ODER Netzwerk ODER Signal ODER Verzögerung ODER Knoten))" searches for patents where the following terms are present: "Phase," "Netzwerk" (network), "Signal," "Verzögerung" (delay), or "Knoten" (node). The command "(BI=(Vibrieren UND Welle ODER Hz UND Gigahertz))" looks for patents containing the

DOI: 10.6977/IJoSI.202508_9(4).0001
J. Deng/Int. J. Systematic Innovation, 9(4), 1-12(2025)

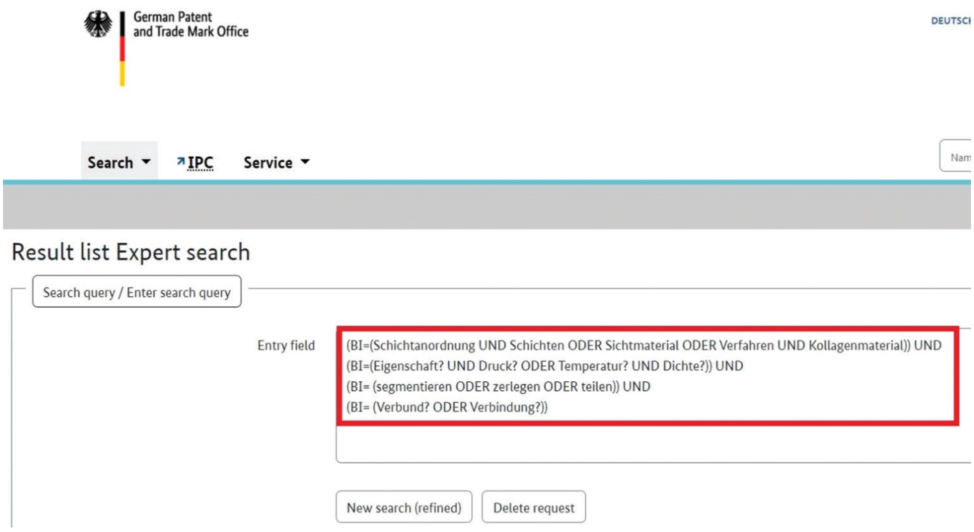


Fig. 2. Search of the German patent database in expert mode for patent DE10, 2017, 123, 891

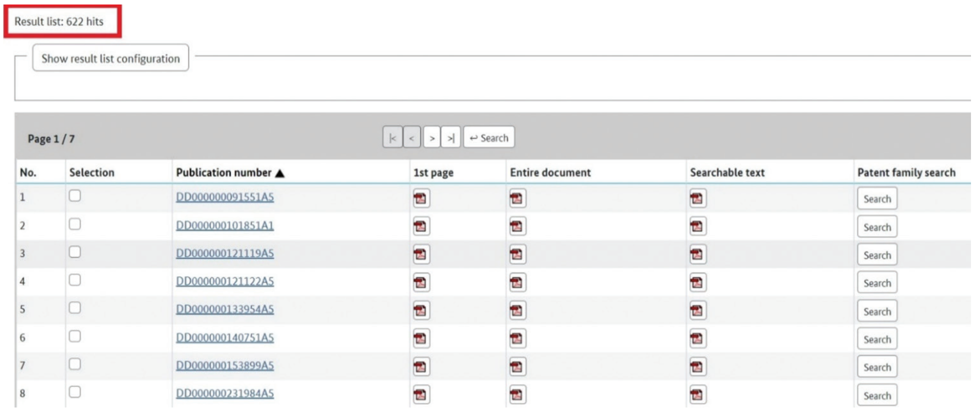


Fig. 3. Results of the search shown in Fig. 2

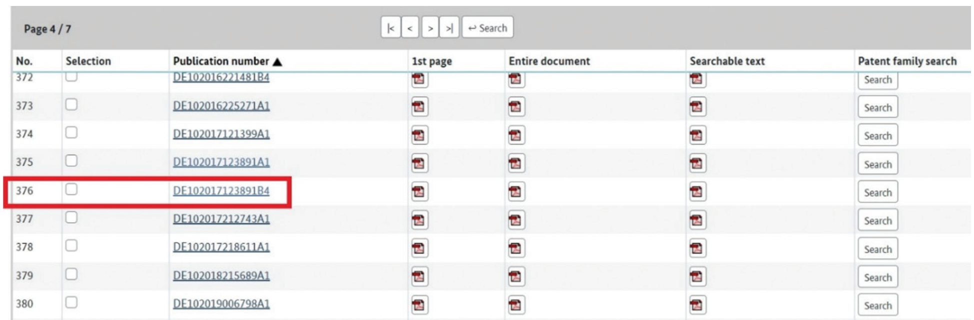


Fig. 4. DE10, 2017, 123, 891 appears in the list of results

words "Vibrieren" (vibrate) and "Welle" (wave) or the terms "Hz" (hertz) and "gigahertz." The command "(BI=(Elektrisch ODER Atom ODER drahtlos UND Feld))" searches for documents containing either "Elektrisch" (electric), "Atom" (atom), or "Drahtlos" (wireless) and "Feld" (field). The command "(BI=(Muster UND UV ODER Synch ODER Frequenz))" specifies that the patents must include the terms "Muster" (pattern) and "UV" (ultraviolet) or the terms "Synch" (sync) or "Frequenz" (frequency). In summary, these commands are used to find patents that mention specific combinations of technical terms. The search will return documents that (i) contain any of the terms related to phases, networks, signals, delays, or nodes; (ii) include either combinations of vibrations with waves or hertz with gigahertz; (iii) mention electric, atomic, or wireless concepts in conjunction with fields; and (iv) discuss patterns with ultraviolet aspects, or include terms related to synchronization or frequency.

Note that the five German words in the first BI are derived from the five most frequently used terms in the patent text, which can be retrieved using Voyant Tools (2024). The terms in the other BIs originate from the catchwords for the top three inventive principles: (18) mechanical vibration, (28) mechanics substitution, and (32) color change (Mann, 2006). In Fig. 5, the top three inventive principles are listed beside the command lines, indicating that command lines 2–4 correspond to these three principles. The search yielded 471 hits, but due to space constraints, the results are not displayed here.

The patent family associated with US10241539B2 includes CN106462177 (A), CN106462177 (B), EP2957982 (A1), EP2957982 (B1), KR102029320 (B1), KR20170021303 (A), TW201601566 (A), TWI721948 (B), US2017139438 (A1), and WO2015193512 (A1), as retrieved from the European Patent Office (EPO) (Fig. 6). This patent has been applied in China, Taiwan, Korea, and the USA and has also been applied by the EPO – which provides a centralized procedure for granting patents in Europe – and the World Intellectual Property Organization, which is responsible for the Patent Cooperation Treaty. DPMA also includes EP patents, but the search results do not include EP2957982 (A1) and EP2957982 (B1), as shown in Fig. 7. In summary, this search command cannot locate its patent in the DPMA, indicating that the recall ability of this command is weak. In the domain of classical information retrieval, and particularly in patent searches, recall is defined as the ratio of the number of relevant patents retrieved to the total number of relevant patents available in the database, as shown in Eq. (1) (Bonino et al., 2010). It measures the completeness of the retrieval process, i.e., how well the system finds all the relevant patents.

$$\text{Recall} = \frac{\text{Number of relevant patents retrieved}}{\text{Total number of relevant patents available}} \tag{1}$$

In this case, the number of relevant patents retrieved was low because the correct patents, EP2957982 (A1) and EP2957982 (B1), could not be retrieved using this search command. Consequently, the recall rate for this search command is low.

Our proposed new methodology is described below. Low-density, high-strength thin cement sheets are designed to satisfy the loading requirements for solar panels (Deng et al., 2023). The purpose of these large, thin cement sheets is to replace the

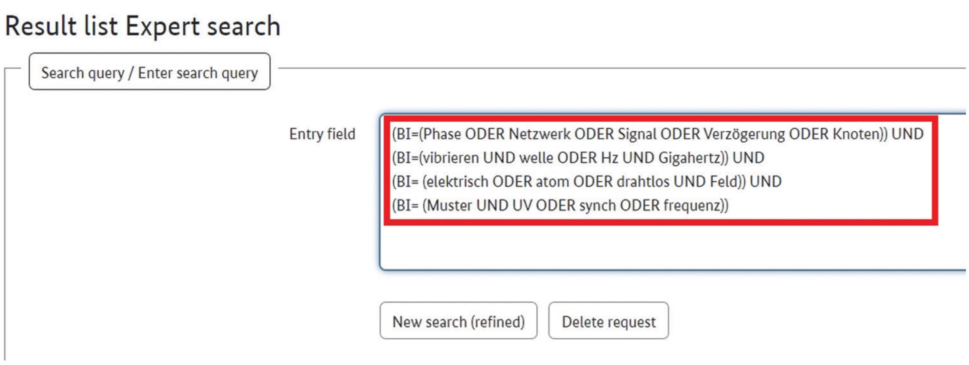


Fig. 5. Search of the German patent database in expert mode for US10241539

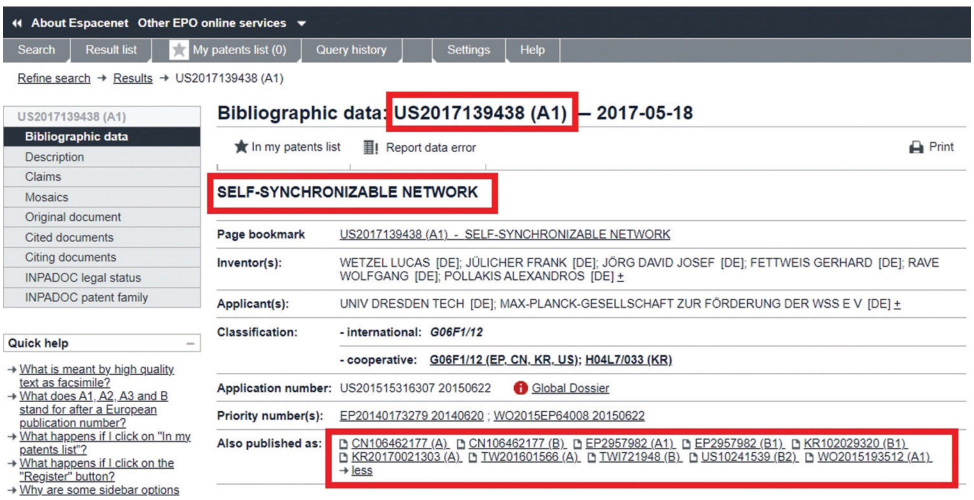
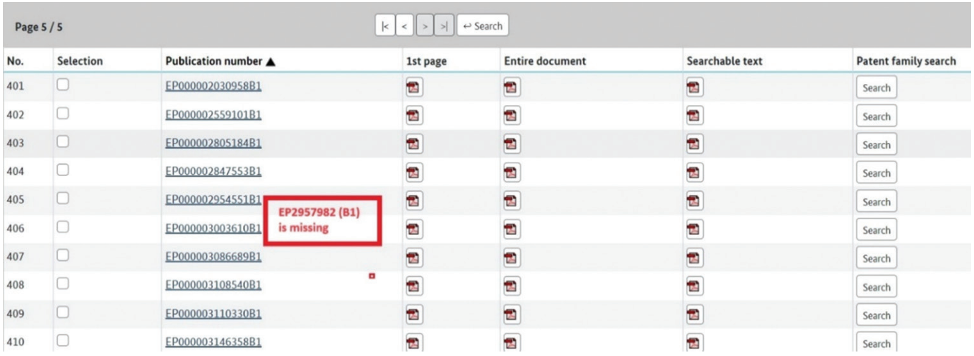


Fig. 6. Patent family for US10241539B2

DOI: 10.6977/IJoSI.202508_9(4).0001
J. Deng/Int. J. Systematic Innovation, 9(4), 1-12(2025)


Page 5 / 5

No.	Selection	Publication number ▲	1st page	Entire document	Searchable text	Patent family search
401		EP000002030958B1				Search
402		EP000002559101B1				Search
403		EP000002805184B1				Search
404		EP000002847553B1				Search
405		EP000002954551B1				Search
406		EP000003003610B1				Search
407		EP000003086689B1				Search
408		EP000003108540B1				Search
409		EP000003110330B1				Search
410		EP000003146358B1				Search

EP2957982 (B1) is missing

Fig. 7. EP2957982 (B1) is missing from the search list

glass in conventional solar panels, thereby creating a lightweight solar panel weighing less than 10 kg. This weight reduction would enable a one-person installation of solar panels rather than a two-person task, thereby significantly increasing the efficiency of the solar panel installation process. These sheets are used as backsheets in glassless solar panels, as illustrated in Fig. 8 (from a Taiwanese patent by Hsiao et al., 2022).

As a pilot run, Deng et al. (2023) created an A4-sized thin sheet with a thickness of 7 mm that could withstand a load of 1.5 kg. Seaweed powder was mixed with Portland cement, a foaming agent, calcium sulfoaluminate, and water. The resulting sheet was sandwiched between layers of ethylene vinyl acetate and a backsheet, resulting in a composite cement sheet. There are two advantages of a sandwiched cement sheet of this type: first, it can support loads of up to 13 kg in a static mechanical loading test without bending, for over 8 h, and second, it can be quickly recovered at the end of its life cycle.

In contrast to conventional low-density, thin cement sheets that support less than 300 grams, the seaweed-powder-based thin cement sheet can withstand a load of 1.5 kg, demonstrating a load-bearing performance five times greater than that of traditional sheets. Moreover, its composite structure significantly enhances its strength, enabling it to support loads of up to 13 kg. This invention effectively resolves the contradictions between thinness and strength, as well as between weight and strength. Although this invention has not been patented, it serves as a perfect example of a technology to which TRIZ reverse can be applied. An exploration of how this technology can be applied to industries beyond solar panels is presented here to help clarify the process of TRIZ reverse.

In summary, given the limitations of existing approaches, this paper aims to bridge the current gap in the implementation of TRIZ reverse using low-density high-strength thin cement sheets as a case study. The two objectives in this study are (i) to provide a detailed procedure for TRIZ Reverse that can be easily applied using the European Patent Database and (ii) to discuss

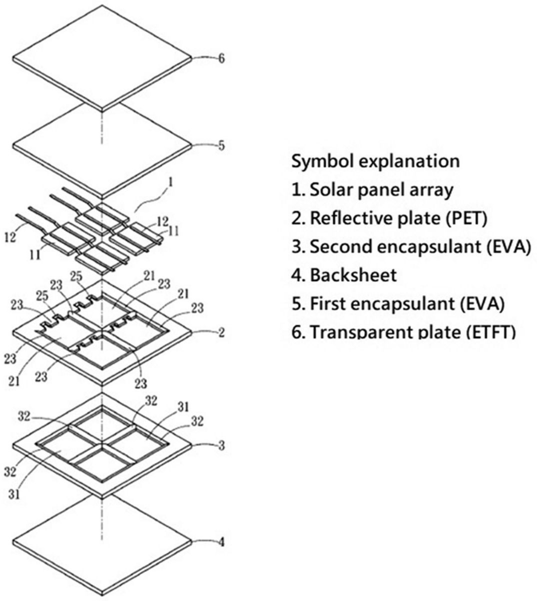


Fig. 8. Detailed view of a glassless panel (Hsiao et al., 2022)
Abbreviations: ETFT: Ethylene tetrafluoroethylene; EVA: Ethylene vinyl acetate; PET: Polyethylene terephthalate
Note: For the meaning of other numbered labels, please refer to the original patent (https://tiponet.tipo.gov.tw/twpat1/twpatusr/00020/GA-I769951.pdf?324551295).

the potential pitfalls of current methodologies in terms of identifying new applications and to propose remedies.

3. Application of TRIZ Reverse

To address the contradiction between the attributes of thinness and strength of the cement sheets, we first identify the relevant engineering characteristics from the 39 engineering parameters recognized in TRIZ: length of a stationary object (Parameter 4) and strength (Parameter 14). Although the goal is to improve strength, the length of the stationary object

DOI: 10.6977/IJoSI.202508_9(4).0001
J. Deng/Int. J. Systematic Innovation, 9(4), 1-12(2025)

poses a challenge. By consulting the contradiction matrix, we can identify the following inventive principles: dynamics (Principle 15), spheroidality–curvature (Principle 14), mechanics substitution (Principle 28), and copying (Principle 26). These principles can guide the development of solutions that can effectively resolve the identified contradiction.

Similarly, to tackle the contradiction between weight and strength, the corresponding engineering characteristics were identified as the weight of a stationary object (Parameter 2) and strength (Parameter 14). The contradiction matrix suggests the following inventive principles: composite materials (Principle 40), copying (Principle 26), cheap short-lived objects (Principle 27), and segmentation (Principle 1).

Table 1. Catchwords for inventive principles.

IPs	Frequency	Catchwords
26: Copying	2	Optical, virtual, reflect (ion), UV, IR
1: Segmentation	1	Split, segment, multi-, divide, micro
14: Spheroidality–curvature	1	Curve, spiral, rotary, circular, sphere
15: Dynamics	1	Dynamic, variable, flexible, free, adapt
27: Cheap, short-lived objects	1	Disposable, cheap, replace, inexpensive, simplification
28: Mechanics substitution	1	Electrical, magnetic, optical, acoustic, wave
40: Composite materials	1	Composite, fiber, (inter-) layer, grid, filler

Abbreviations: IPs: Inventive Principles; IR: Infrared; UV: Ultraviolet.

Table 1 below lists the inventive principles along with their corresponding catchwords, which were used to guide the search for new applications. The column marked "Frequency" represents the number of times the inventive principles appear in these two contradictions.

Using Voyant Tools (2024), we also extracted the five most frequent words from the work by Deng et al. (2023) as cement (172), sheet (151), solar (71), thin (59), loading (56), and density (56), meaning that "cement" occurs 172 times, "sheet" 151 times, and so on. To search, we entered the command: ("cement sheet" AND load* AND thin AND density AND solar) into the Espacenet database, which returned the patent AU2009279384 (A1). The patent search results are shown in Figs. 9 and 10. We note that in Fig. 9, only the English patent database is selected rather than the entire Espacenet patent database; this is because full-text search is available only in the English patent database, whereas we can only search for keywords in the title or abstract in the Espacenet database. Patent AU2009279384 (A1) describes the use of cement sheets as outer skins for wall panels. From the website for these wall panels, it was found that these cement sheets can also be used for cement sidings, partition walls, roofing, and ceilings (see https://lonwow.en.made-in-china.com/product/fKyxcNHJABVe/China-Fiber-Cement-Exterior-Wall-Panel-with-Australia-Standard-4-5-18mm-Thickness.html). When the command was adjusted to exclude the word "solar" ("cement sheet" AND load* AND thin AND density), it returned 86 patents, which broadened the scope to various construction-related applications. The cooperative patent classification (CPC) codes in the section, class, and subclass for those 86 patents are illustrated in Fig. 11. The meanings of the specified CPC codes are as follows:

1. C04B: Lime; magnesia; slag; cements; compositions thereof; ceramics; refractories;

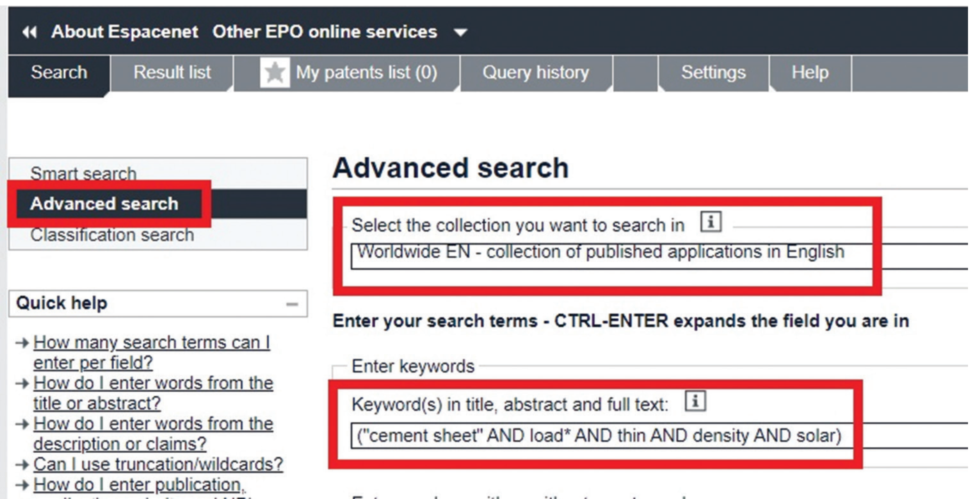


Fig. 9. Advanced search for all requirements using Voyant Tools

DOI: 10.6977/IJoSI.202508_9(4).0001
J. Deng/Int. J. Systematic Innovation, 9(4), 1-12(2025)

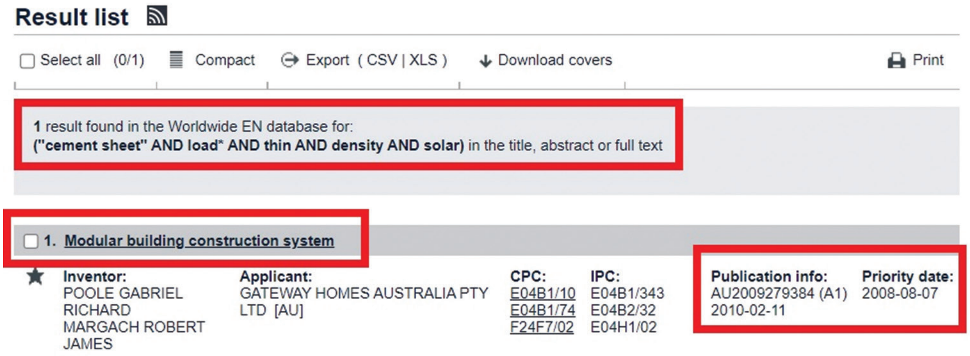


Fig. 10. Results of the search shown in Fig. 9

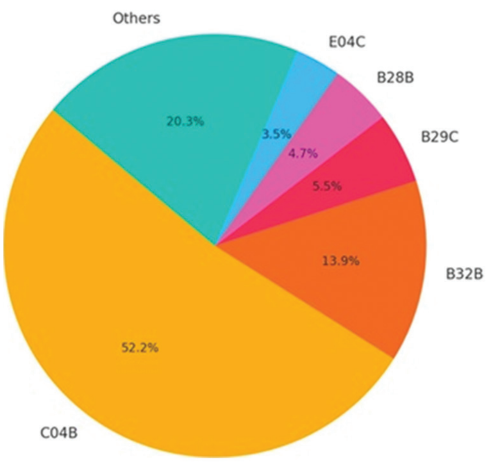


Fig. 11. The distribution of CPC codes among the 86 patents
Abbreviation: CPC: Cooperative patent classification

treatment of natural stone. This code relates to the production and processing of cement, ceramics, and other similar materials.

2. B32B: Layered products, i.e., products built up of strata of flat or non-flat, e.g., cellular or honeycomb, form. This involves the creation and processing of layered or laminated products, including materials such as composites.
3. B29C: Shaping or joining of plastics; shaping of substances in a plastic state, in general. This pertains to the techniques and methods used in molding, shaping, and joining plastics or other materials in a plastic state.
4. B28B: Shaping clay or other ceramic compositions; shaping materials in a plastic state, not otherwise provided for. This code covers the processes involved in shaping ceramic materials and other substances in a plastic state.
5. E04C: Structure elements; building elements; built-up structures. This involves the construction of structural elements such as building frameworks, modular constructions, and other components used in construction.

Each patent has several CPC codes. For example, one of the 86 patents is US2022363602 (A1), which has the following CPC codes: B28B1/16 (EP, US), B28B1/522 (EP, US), B28B1/525 (EP, US), B28B23/0087 (EP), C04B14/06 (US); C04B14/185 (US), C04B18/24 (US), C04B28/02 (US), C04B28/04 (EP), C04B40/0064 (US), E04C5/073 (US), C04B2111/00612 (EP, US), E04C5/073 (EP), and Y02W30/91 (EP). Each CPC code is reduced to its section, class, and subclass, so that B28B1/16 is reduced to B28B. Following this scheme, the patent US2022363602 (A1) has the following classifications: B28B (4), C04B (7), E04C (2), and Y02W (1), meaning that B28B occurs four times, C04B seven times, etc. This process was repeated for all 86 patents, resulting in the data presented in Fig. 11.

Specifically, Fig. 11 presents a pie chart showing the percentage distribution, with CPC codes representing less than 3% grouped under "Other." The specific CPC codes displayed include C04B (764), B32B (204), B29C (80), B28B (69), and E04C (51). The analysis was done using ChatGPT (OpenAI, 2024). The CPC codes indicate that most of the patents are related to construction materials, except for one instance of C04B, which is related to the process of making cement.

When the details of the 86 patents were examined, one particular example – WO199515849 – stood out, highlighting a new direction for the use of cement sheets (Andersen and Hodson, 1995). This patent suggested that cement sheets could be used in food containers and boxes.

We further refined our search by inputting the command ("cement sheet" AND load*) AND (curve OR sphere), which retrieved 35 patents. Of these, WO2020/234622 (Priyadarshana et al., 2020) proposed a new application for wet-area lining boards (where "wet area" refers to spaces such as shower rooms and restrooms). In addition, patent US20150086773 (Grundy et al., 2015) indicated that lightweight cement sheets could be utilized for:

Table 2. Applications of cement sheets

Command in search	Patents retrieved	Patent number	Applications
("cement sheet" AND load* AND thin AND density AND solar)	1	AU2009279384 (A1)	Outer skins for wall panels, sidings, partitions, roofing, and ceilings of buildings
("cement sheet" AND load* AND thin AND density)	86	WO199515849	Food containers and boxes
("cement sheet" AND load*) AND (curve OR sphere)	35	WO2020/234622	Wet-area lining boards
		US20150086773	High-rise buildings or structures built on less stable ground, areas prone to extreme weather conditions
("cement sheet" OR "cement plate" OR "cement board") AND (segment OR divide) AND (dynamic OR flexible)	137	CA2808343A1	Green roofs and living walls
		US5607758	Smoke containment curtains for public buildings such as hotels, restaurants, prisons, hospitals, airports, and aircraft

1. Lightweight building solutions: Due to their reduced density and weight, these materials are suitable for construction projects where minimizing structural load is crucial, such as in high-rise buildings or structures built on less stable ground.
2. Building exteriors: These aerated panels are particularly well-suited for use on building exteriors, especially in areas prone to extreme weather conditions.

Lastly, we input the command ("cement sheet" OR "cement plate" OR "cement board") AND (Segment OR divide) AND (Dynamic OR flexible), which returned 137 patents. Of these, patents CA2808343A1 (Hellwig, 2011) and US5607758A (Schwartz, 1997) revealed new applications. Patent CA2808343A1 indicated that cement sheets, plates, or boards could be used in the following applications:

1. Green roofs: The system is particularly suitable for creating environmentally friendly, energy-efficient roofs that contribute to urban sustainability. It is applicable in both new constructions and retrofits to existing roofs, due to its lightweight nature.
2. Living walls: This technology can be applied to the construction of living walls on buildings, providing esthetic, environmental, and insulation benefits. This would be especially useful for vertical gardening in urban spaces.

The patent described in the document US5607758 pertains to a "smoke containment curtain." This is designed to contain smoke within a confined area, such as a ceiling segment or corridor, and is particularly useful in environments where fire safety is critical. Cement sheets, plates, or boards can be incorporated into this design. The specific applications of this patent include:

1. Public buildings: The smoke containment curtain can be used in public buildings such as hotels, restaurants, prisons, and hospitals, where fire safety regulations require systems to prevent the spread of smoke and toxic fumes during a fire.
2. Airports and aircraft: The curtain can be applied in airports and aircraft to contain smoke, especially in confined spaces, where toxic fumes can be deadly.

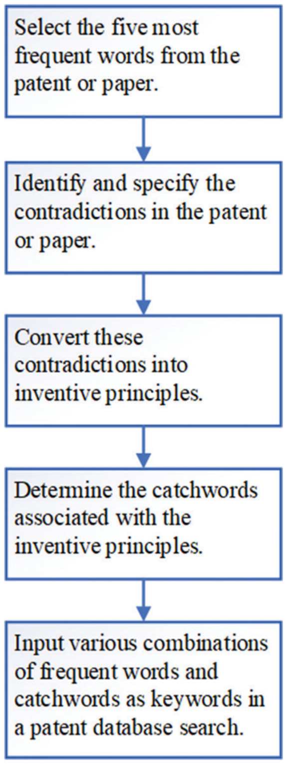


Fig. 12. Patent search procedure

We created a summary of the applications based on the patents retrieved using these commands in searches of the English Espacenet patent database, as shown in Table 2. All patent summaries were generated using ChatGPT (OpenAI, 2024).

The search results revealed that low-density, high-strength thin cement sheets have potential applications beyond solar panels, extending into various domains including construction, food packaging, and safety solutions. We note that full-text searches of the English patent database were limited to only 10 terms, whereas the German patent database allowed for more than 10 terms in full-text searches, making the English patent database less powerful for comprehensive full-text patent searches.

To address the previous issue of low recall in retrieving the correct patent in Fig. 7, it is suggested that various keyword combinations from Table 2 can help mitigate this problem. Finally, to ensure the completeness of the patent search in TRIZ reverse, a flowchart is presented in Fig. 12.

4. Conclusion

Identifying new markets for innovative products or technologies is essential for firms and typically involves collaboration among planners, technical teams, field operations, and manufacturing departments (Dougherty, 1990). Conventionally, this process is driven by marketing surveys and focus groups, meaning that researchers often have preconceived notions about potential markets. Such psychological inertia can limit the scope of market exploration.

The TRIZ reverse methodology overcomes this inertia by abstracting the contradictions inherent in new products or technologies and retrieving the corresponding inventive principles. Using catchwords derived from these principles (Mann, 2006), specific search commands can be formulated to uncover relevant patents, thereby revealing potential new applications. Although previous studies have laid the groundwork for using TRIZ reverse, they have primarily focused on the German patent database (Günther et al., 2021; Günther & Popova, 2022) or specialized software such as Patent Inspiration (Dewulf et al., 2023), and have often fallen short of identifying specific applications for new products, with only vague patent classification codes.

This paper addresses these gaps in two significant ways: first, we have demonstrated the use of the publicly accessible Espacenet patent database, thus enabling broader access to TRIZ reverse technology for English-speaking users, and second, we have delved into the details of the patents retrieved, thereby identifying specific applications for the product under investigation.

The product considered in this work consisted of low-density, high-strength thin cement sheets, which were designed to meet the loading requirements for solar panels (Deng et al., 2023). These large, thin cement sheets were developed to replace conventional solar panel glass and consist of lightweight panels weighing less than 10 kg, hence allowing for a one-person installation process. This invention resolves the contradictions between thinness and strength as well as between weight and strength. Although not yet patented, this product serves as an ideal example to illustrate the use of TRIZ reverse technology.

Four different search commands were used to explore potential applications:

1. (“cement sheet” AND load* AND thin AND density AND solar), which directly addressed the invention but retrieved only one patent: AU2009279384 (A1) (Poole and Margach, 2010).
2. (“cement sheet” AND load* AND thin AND density), which broadened the search by excluding the solar constraint, resulting in 86 patents.
3. (“cement sheet” AND load*) AND (curve OR sphere), which focused on loading capacity and inventive principle 14: Spheroidality – curvature, retrieving 35 patents.
4. (“cement sheet” OR “cement plate” OR “cement board”) AND (Segment OR divide) AND (Dynamic OR flexible), which expanded the scope to include thicker cement boards or plates, considering inventive principles 1: Segmentation and 15: Dynamics, and retrieved 137 patents.

The search results revealed several potential applications beyond solar panels, including:

1. Construction industry: outer skins for wall panels, siding, partition walls, roofing, and ceilings.
2. Food industry: food containers and boxes.
3. Wet-area lining boards.
4. High-rise buildings or structures on less stable ground and in areas prone to extreme weather conditions.
5. Green roofs and living walls.
6. Smoke containment curtains for public buildings such as hotels, restaurants, prisons, hospitals, airports, and aircraft.

There were certain limitations on this research, particularly in the patent search process, as only the abstracts were reviewed due to time constraints. This approach may have missed some potential applications.

Acknowledgment

The author acknowledges the use of ChatGPT (OpenAI, 2024), which proves to be extremely helpful in summarizing patent CPC codes and content.

References

Altshuller, G. (1984). *Creativity as an Exact Science: The Theory of the Solution of Inventive Problems*. Gordon and Breach, New York.

Altshuller, G. (1996). *And Suddenly the Inventor Appeared - TRIZ, the Theory of Inventive Problem Solving*. 2nd ed. Technical Innovation Center, Worcester, MA.

Andersen, P.J., Hodson, S.K. (1995). *Hinges for Hydraulically Settable Materials*. MISUMI India, United States.

AULIVE Homepage. Available from: https://www.patentinspiration.com [Last accessed on 2024 Jul 01].

Bianchi, M., Campodall'Orto, S., Frattini, F., & Vercesi, P. (2010). Enabling open innovation in small-and medium-sized enterprises: How to find alternative applications for your technologies. *RandD Management*, 40(4), 414–431.

Bonino, D., Ciaramella, A., & Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, 32(1), 30–38. https://doi.org/10.1016/j.wpi.2009.05.008

Chesbrough, H.W. (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business Press, Massachusetts.

Deng, J.J., Lin, T.H., Wang, J.S., Hsiao, Y.C., Tu, G.Y., & Huang, Q.H. (2023). A method of producing low-density, high-strength thin cement sheets: Pilot run for a glass-free solar panel. *Materials(Basel)*, 16(23), 7500. https://doi.org/10.3390/ma16237500

Dewulf, S. (2005). *Directed variation solving conflicts in TRIZ Part 1. TRIZ Journal*. Available from: https://www.metodolog.ru/triz-journal/archives/2005/09/07-01.html

Dewulf, S., Günther, S., Childs, P.R., & Mann, D. (2023). Opening up new fields of application with TRIZ reverse-conceptual framework, software application, and implementation challenges. In: *International TRIZ Future Conference*. Springer Nature, Switzerland, Cham, p55–69.

Dougherty, D. (1990). Understanding new markets for new products. *Strategic Management Journal*, 11, 59-78.

DPMA. *German Patent Search in Expert Mode in English*. Available from: https://depatisnet.dpma.de/depatisnet/depatisnet?action=experte&switchtolang=en [Last accessed on 2024 Jul 31].

Glaser, M., & Miecznik, B. (2009). TRIZ for reverse inventing in market research: A case study from WITTENSTEIN AG, identifying new areas of application of a core technology. *Creativity and Innovation Management*, 18(2), 90–100. https://doi.org/10.1111/j.1467-8691.2009.00516.x

Grundy, S., Lau, K., Liu, X., Mueller, T.P., Peng, J.Z., & Terzian, S. (2015). *Aerated Fiber Cement Building Products and Methods of Making the Same*. James Hardie Technology Limited, Dublin, LE.

Günther, S., Popova, S.L, & Garzon, S. (2021). TRIZ reverse-specification and application of a 7 step-by-step approach for systematic knowledge and technology transfer. In: *Proceedings of the 16th International Conference Trizfest*. p184–195.

Günther, S., & Popova, S.L. (2022). *TRIZ Reverse-Case Study in the Field of Basic Research of Physics*. TRIZfest, Malaysia, p198–207.

Harre, K., Pietsch, T., Firzlaff, D., & Tobiasch, E. (2019). *Verfahren Zum Bereitstellen eines Kollagenbasierten Schichtmaterials und Daraus Hergestelltes Biokompatibles Formteil*. Aesculap AG, German.

Hellwig, R.T. (2011). *Living Roof and Wall Systems Using Cultivated Mineral Wool Mats to Support BLAVES, Methods of Cultivation and Innoculants Therefor*. Archiphyte LLC, United States.

Hsiao, Y.C., Chiang, Y.W., Lin, C.H., & Huang, K.H. (2022). *Solar Photovoltaic Panel Package Structure.* Taiwan Patent TW, Taiwan.

Kotler, P., Keller, K.L, & Chernev, A. (2022). *Marketing Management (Global Edition)*. 16th ed. Pearson, United Kingdom.

Mann, D.L. (2006). Principle-guided patent searches. Systematic Innovation E-zine. Available at: https://www.systematic-innovation.com/ezine

OpenAI. (2024). *ChatGPT (GPT-4) [Large Language Model]*. Available from: https://www.openai.com [Last accessed on 2024 Aug 20].

Osterwalder, A., & Pigneur, Y. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. John Wiley and Sons, United States.

Poole, G., & Margach, R.J. (2010). *Modular building construction system, WO Patent No. 2010/015042 A3*. Available from: https://ipsearch.ipaustralia.gov.au/patents/2009279384

Priyadarshana, G., Karunaratne, V., Amaratunga, G., Kottegoda, N., & Siriwardhana, A. (2020). Asbestos free nano-hybrid for fiber-cement composite applications. PCT Patent Application, United States.

Schwartz, W.C. (1997). *Smoke Containment Curtain*. BGF Industries, Inc, United States.

Shulyak, L., & Rodman, S. (1997). *40 Principles: TRIZ Keys to Innovation*. Technical Innovation Center, Worcester, MA.

Wetzel, M., Julicher, F., Jorg, D.J., Fettweis, G., Rave, W., & Pollakis, A. (2019). *Self-Synchronizable Network*. Wetzel, United States.

Wittenstein, L. (2006). *Distraction Device*. Payne, United States.

Voyant Tools Homepage. Available from: https://voyant-tools.org [Last accessed on 2024 Jul 31].

AUTHOR BIOGRAPHIES

Dr. Jyhjeng Deng has been a faculty member at DaYeh University in Taiwan since 1994. He received his Ph.D. in Industrial Engineering from Iowa State University. He is currently a Professor in the Department of Accounting and Information Management at DaYeh University and serves as a board member of the Society of Systematic Innovation. In 2024, he was elected as a member of Sigma Xi. His research interests include systematic innovation – particularly TRIZ, patent analysis, and circumvention – as well as data analysis in deep learning and artificial intelligence.

Strengthening the absorptive capacity of the national innovation system through university-industry research collaboration: A Theory of Inventive Problem Solving approach

Abeda Muhammad Iqbal[1]*, Narayanan Kulathuramaiyer[2], Adnan Shahid Khan[2], Johari Abdullah[2]

[1]Department of Business Administration, Faculty of Economics, Business and Accounting (FEBA), iCATS University College, Sarawak, Malaysia

[2]Department of Network Computing, Faculty of Computer Science and Information Technology (FCSIT), Universiti Malaysia Sarawak (UNIMAS), Sarawak, Malaysia

*Corresponding author E-mail: miabeda@icats.edu.my

(Received 17 July 2024; Final version received 12 March 2025; Accepted 3 June 2025)

Abstract

There has been a universal recognition that university-industry research collaboration (UIRC) is vital to strengthen the national innovation system (NIS) and economic growth. Despite a series of research studies on the significance of UIRC, present baseline models to enhance the capabilities of NIS through URIC are still scarce, specifically in developing countries. This research has highlighted that absorptive capacity has a vital influence on the NIS as well as on the research and innovative activities of an individual. Moreover, this research highlights how education and training (E&T) can enrich the absorptive capacity of NIS and UIRC using the Theory of Inventive Problem Solving (TRIZ) approach. The methodology involves applying TRIZ tools, such as function modeling, contradiction analysis, and inventive principles to identify effective strategies for improving the absorptive capacities of universities and industries and consequently of NIS. Thus, proposed solutions include enhancing the education, training systems and programs, promoting collaboration between universities and industries, and decreasing aids, such as foreign-educated and skilled workforce, which can strengthen the absorptive capacity of NIS. Analysis of this research suggests that a strong E&T system and upgrading the standard of education are crucial factors for improving the absorptive capacity of NIS. Recommendations include developing policies that foster a culture of knowledge, promoting interdisciplinary research, and incentivizing innovation. Future research directions include exploring comparative analysis with other developed and developing countries' strategies in customizing the education and research system to enhance the outcomes of research and innovations.

Keywords: Absorptive Capacity, Education and Training, National Innovation System, Theory of Inventive Problem Solving, University-Industry Research Collaboration

1. Introduction

Innovations and new technologies heavily rely on our capacity to assimilate and apply novel concepts and ideas, which is termed as absorptive capacity (Kale et al., 2019). This role is analogous to the robust foundation of a house that sustains its structure. Nations worldwide are increasingly recognizing the factors to enhance their capability to learn and apply emerging technologies, viewing it not just as a beneficial pursuit but as an absolute necessity (Albort-Morant et al., 2016). This realization stems from confronting challenging economic issues that can be effectively addressed and securing a promising future through the development of new and advanced technologies. Both industry and academia play pivotal roles in the domain of inventions and technology. The central component of a vast machine facilitates the exchange of ideas and advances technology (Iqbal et al., 2015; Khan, 2022; Ye & Wang, 2019; Yu et al., 2022), and its absorptive capacity plays a crucial role in the realm of innovation. Despite our awareness of universities and industries and their collaboration significance, producing new technology is always challenging, particularly in developing nations (Iqbal et al., 2021). In industrialized nations, absorptive capacity is widely acknowledged as a fundamental factor in promoting technological innovation and transformation within university-industry research collaboration (UIRC) and contributing significantly to long-term economic progress (Terstriep & Lüthje, 2018). Given the present state of the international economy and the imperative of ensuring future prosperity, countries worldwide recognize the need to strengthen their national innovation systems (NISs) by enhancing their absorptive capacity (Ahn, 2016). Numerous research studies have consistently emphasized the critical significance of both UIRC and absorptive capacity, recognizing their profound influence on the NIS. Despite the recognized importance of absorptive capacity, the rate of technical innovation resulting from UIRC remains relatively low, particularly in developing nations (Iqbal et al., 2022). Various research endeavors have sought to identify the constraints on producing a number of research and innovations in UIRC and explore ways to address them. Unfortunately, many of these initiatives have focused on superficial solutions, such as hosting workshops and seminars or employing knowledgeable staff who provide only short-term remedies for underlying issues (Iqbal, 2018; Iqbal et al., 2022). Recognizing the pivotal roles of universities and industries as integral components of the NIS, actively contributing to technological innovation, it becomes imperative to examine the impact of absorptive capacities of universities and industries and their research collaborations on the broader innovation landscape (Chryssou, 2020).

This study aims to enhance the absorptive capacities of universities and enterprises to address the aforementioned challenge. A notable limitation identified in existing research is the absence of a visual thinking strategy, leading to a constrained level of predictability in outcomes (Czinki & Henntschel, 2015; Kasravi, 2010). Given the intricate interdependencies between the two key constituents of the NIS, universities and industry, a comprehensive evaluation necessitates the adoption of a visual thinking strategy (Ersin, 2009; Terninko et al., 1998). To address this, the study advocates the implementation of the Theory of Inventive Problem Solving (TRIZ) paradigm and associated resources. Utilizing the concepts of TRIZ in a non-technical domain is not new; much research has already contributed to such a domain (Hammer & Kiesel, 2019; Lin & Chen, 2021; Navas et al., 2015; Walter, 2015; Wang et al., 2015). TRIZ, traditionally associated with technical problem-solving, has shown significant potential in social, management, and humanities domains, particularly in business and management. Mann and Spain (2000) emphasize that TRIZ can effectively address social, management, and humanities issues by providing a structured approach to decision-making and strategic planning in organizations. Furthermore, Ruchti et al. (2001) highlight that TRIZ can enhance business decision-making processes, demonstrating its versatility beyond engineering applications. The methodology's 12 principles serve as valuable tools for navigating complex business challenges, fostering innovative solutions, and improving management practices (Ruchti et al., 2001). Despite its technical origins, the integration of TRIZ into non-technical domains remains underexplored, indicating a need for further research to fully leverage its capabilities in business contexts. Ishrat et al. (2023) conducted an extensive and exhaustive discussion on TRIZ in social, management, and humanities domains, particularly in business and management.

In contrast to conventional strategies involving concessions or trade-offs, TRIZ excels at resolving contradictions, rendering it a potent tool for generating innovative solutions to challenges (Hammer & Kiesel, 2019). Utilizing a two-stage process of "cause and effect analysis and inventive principles," TRIZ furnishes a visual thinking tool that enables the systematic resolution of problems (Kasravi, 2010). Typically, the initial phase of the TRIZ methodology involves pinpointing the core problem by eliminating peripheral concerns and assumptions. This necessitates breaking down the issue into its fundamental components, comprehending each facet, expressing the elements in the most fundamental or elementary manner, and ultimately liberating oneself from the constraints of the language used to articulate the problem (Ersin, 2009). During this phase, the TRIZ approach relies on the problem solvers' capacity to contemplate the fundamentals of the issue and recognize its pivotal elements. To delve into the complexities of modeling relationships, we employed key TRIZ tools in this study, such as function modeling and cause-effect chain analysis (Ersin, 2009). Beyond providing comprehensive problem-solving techniques, TRIZ also illuminates the exploration of inconsistencies between parameters that are deteriorating and those that are improving. Furthermore, it furnishes a comprehensive roadmap that addresses all parameters (Su and Lin, 2008; Ekmekci, 2019). Consequently, we present an enriched perspective to enhance the absorptive capacities of the NIS through the application of TRIZ tools and solutions.

There are five primary contributions from this research. First, it enriches the discourse of absorptive capacity and its implications for research institutions, particularly universities and industries, by presenting a practical solution using the TRIZ theory to enhance their potential for technological innovation. In addition, it stands as the pioneering application of TRIZ theory to UIRC investigations, substantiating its effectiveness in this context. Third, it supplements the existing body of literature on the NIS by showcasing how TRIZ theory can be harnessed to illuminate the critical elements of UIRC. Fourth, it identifies supportive factors that can enhance UIRC's capacity for innovation and, consequently, fortify the NIS. In conclusion, this study underscores the significance of education and training (E&T), standards of education (SE), on absorptive capacity. By unlocking the latent potential of businesses and universities to contribute invaluable research and innovations to their respective countries, the study provides policymakers with practical insights and relevance.

The rest of this article is organized as follows: in Section 2, conceptual design and hypothesis development are presented, while research design and methodology are presented in Section 3. Results and analysis are covered in Section 4, followed by a conclusion and future direction in Section 5.

2. Conceptual Design and Hypothesis Development

Absorptive capacity holds a prominent role in NISs and is integral to the theory of economic growth, facilitating faster development in innovation-driven economies (Spreafico, 2022). It denotes the ability to assimilate, process, adapt, and apply new information for business purposes, aiding both innovation and industrial processes (Miguélez & Moreno, 2015). Various definitions exist; one perspective, as proposed by Müller et al. (2020), defines absorptive capacity as the capability to absorb, process, and utilize new information effectively. Another definition, by Naqshbandi & Tabche (2018), views it as the power to transform knowledge into a productive form, while Ishrat et al. (2023) characterize it as the capability to identify the importance of new external knowledge, integrate and employ it to attain economic objectives. In essence, absorptive capacity is the technological competence rooted in a firm's internal technical capabilities, playing a pivotal role in fostering innovation (Murovec & Prodan, 2009). One of the primary challenges hindering collaboration between university and industry personnel during the establishment of UIRC is the deficiency in technological competency or absorptive capacity (Carrasco-Carvajal, 2023). In lower-income nations, the relatively subdued educational systems have resulted in a less conspicuous presence of absorptive capacity, which is a pivotal element in research and innovations (Chryssou, 2020). The quality of E&T is emphasized as a determining factor for research breakthroughs (Zahra & George, 2002). Particularly in developing nations, industries often rely on external technical support, such as collaborations with universities or the recruitment of highly qualified foreign expertise (Iqbal, 2018).

Universities, recognized as centers of learning and education, play a pivotal role in the advancement of research-oriented systems (Iqbal et al., 2012; 2015; Kim & Park, 2023). Similarly, the presence of a well-structured education system, institutions for training management, and a robust technology infrastructure contributing to the production of highly educated and skilled professionals has a profound impact on the success of research and development endeavors (Iqbal et al., 2015). The quality of E&T, as highlighted by Müller et al. (2020), significantly influences research and innovation capabilities. Similarly, several studies (Heiden et al., 2015; Iqbal et al., 2010; 2013) elaborate that high-tech nations consistently achieve breakthroughs and enhance their market position by elevating competitiveness among their citizens through high-quality training and education. Moreover, numerous studies have identified a direct correlation between the volume of innovations and E&T, emphasizing that highly educated and trained individuals play a crucial role in driving progress in research and innovation (Kobarg et al., 2018). In line with this observation, the hypothesis of the study is: E&T have a positive influence on the absorptive capacity of universities and industries and on NIS (H1a).

However, it is observed that universities, especially in lower-income nations, persist in training and educating their staff using outdated methods (Akram, 2020). Additionally, there is a lack of flexibility in the curriculum to meet industry requirements due to various operational constraints, such as adherence to traditional teaching syllabi, theory-centric curricula, mindset-oriented curricula, and adherence to self-produced curricula (Topkaya, 2015). Consequently, graduates from such programs may face challenges in transforming knowledge into practical applications (Gu, 2021). Li et al. (2023) suggest that institutions of higher learning and training should modernize their curricula, advocating for a research-focused curriculum as it would yield greater benefits in terms of research and innovation compared to a theory-based one. Theoretically-based curricula lead to graduates who lack practical experience and are ill-prepared for engaging in research and innovation or enhancing their absorptive capacities. Overall, the E&T systems in developing nations remain of low quality and are in urgent need of the introduction of research-based teaching and training methodologies (Alexander & Yuriy, 2018; Chen et al., 2019; Iqbal et al., 2010; Mallana et al., 2013). The standard of E&T has been shown to be a significant determinant of information absorption and is closely linked to absorptive capacity, as indicated by Iqbal et al. (2013). Thus, this study posited the following hypothesis: standard of education as a reinforcing factor of E&T has a positive influence on the absorptive capacity of universities and industries and on NIS (H1b).

Furthermore, several studies have demonstrated that developing nations rely on foreign-educated and skilled workers (FESW) to enhance their research and innovations due to their lower educational standards (Dooley & Gubbins, 2019). FESW constitutes a critical resource that businesses can leverage to improve production (Patston et al, 2018). Similarly, a substantial body of research has found that FESW has a significant impact on national research and innovation. This impact includes promoting technological spillovers, facilitating the integration of international trade, fostering a more competitive business climate, and stimulating enterprise development.

Despite these contributions, FESW may not significantly contribute to the growth of a nation's intellectual capital (Knudsen et al., 2001). Davies et al. (2014) define intellectual capital as the collective knowledge, experience, learning, and teamwork skills possessed by a workforce, enabling them to solve problems and add value to a company. Similarly, Crown et al. (2020) characterize intellectual capital as the application of individual intelligence to generate novel ideas and information. Intellectual capital, considered a nation's internal competitive advantage, plays a fundamental role in fostering innovation, essential for the nation's development (Truong & Nguyen, 2023). A dynamic environment and high intellectual standards in education are imperative for sustaining and enhancing research and innovation. Countries should prioritize developing intellectual capital and fostering innovative capabilities rather than relying on the employment of FESW (Farzaneh et al., 2022). Industrialized nations have become significant players in global competition, research, and innovation by consistently generating intellectual capital and enhancing their innovation systems, gradually reducing dependence on FESW (Rehman et al., 2011). According to Arshad et al. (2023), intellectual capital has gained prominence in recent years, contributing to creative ideas and technological advancements in research and innovation firms. As a result, the economy has advanced to a higher stage, intensifying competition through these technological developments. Nations strive to become market leaders or maintain their positions in the global economy by enhancing the number of inventors from the supply side of the economy. Many countries have enhanced their E&T systems to produce highly educated, skilled workers and inventors. In summary, it is extensively documented that an organization's innovation capacity is directly associated with its intellectual capital rather than relying on FESW (Ahmed et al., 2020). Intellectual capital is rapidly becoming a critical factor at the top of the value chain for institutions and organizations. Thus, this study posited the following hypothesis: FESW does not have a significant influence on the absorptive capacity of universities and industries and on NIS (H2).

3. Research Design and Methodology

This study employed an open-ended questionnaire to collect data through a survey approach based on the positivist paradigm. In this paradigm, knowledge is initially derived from facts, evidence, and reasoned thinking. Subsequently, assertions are formulated, and hypotheses are tested using statistical methods. This approach proves valuable in uncovering important discoveries and insights while describing norms and correlations between variables (Ozgun et al., 2022). When the target population is geographically dispersed, the survey method becomes an appropriate strategy for data collection. Surveys serve as an effective means of gathering information from participants, aiding in accurately documenting norms, defining variable correlations, and understanding generalized interactions across constructs, thereby yielding crucial insights and discoveries (Alrowwad et al., 2020). Survey research possesses the capability to characterize the relationships between variables (Groeneveld et al., 2015). The primary focus of the survey approach in this study is quantitative, involving the collection of a substantial amount of participant data through an open-ended questionnaire. To authenticate the instruments, a panel of knowledgeable adjudicators was employed, as recommended by Dong et al. (2023). A pilot test of the questionnaire was conducted to ensure that its wording and structure were suitable for the survey, allowing the researcher to verify the appropriateness of the data needed from the target group. Following the confirmation of the instruments' validity, an online questionnaire was distributed to the respondents, with an anticipated total of about 500 participants.

To achieve a 95% confidence level within a population of 500, Krejcie and Morgan's table suggests a sample size of 210 respondents; however, for increased precision, this study gathered data from 214 respondents. The study instrument included absorptive capacity as the dependent variable and E&T, FESW, and SE as independent and reinforcing variables, accordingly. Relevant variables were selected and measured based on the parameters within the scope of the present study. Table 1 provides a comprehensive overview of the main variables and constructs within the study, detailing the number of items associated with each.

Table 1. Variables, constructs, and items of research instruments

Variables	Constructs	N	Items
Dependent variable	Absorptive capacity	1	Competencies of understanding and learning capabilities.
		2	Ability to transform new knowledge into commercial ends.
Independent variable	Education and training	1	Number of educational organizations.
		2	Number of training management institutions.
		3	Number of educated people for research and innovation.
	Foreign-educated and skilled workers	1	Talent and intelligence in the national innovation system.
		2	knowledge and expertise in the national innovation system.
		3	Number of research and innovations in the national innovation system.
Reinforcing factor	Standards of education	1	Research based on teaching and training curriculum.
		2	Introducing a new syllabus according to new challenges.

The study centered around the dependent variable of absorptive capacity, encompassing competencies related to comprehension and learning (Item 1) and the application of newly acquired knowledge for commercial purposes (Item 2). E&T served as the independent variable, with components such as the quantity of educational organizations (Item 1), training management institutions (Item 2), and educated individuals for research and innovation (Item 3). Another independent variable was FESW, covering the number of research and innovations in the NIS (Item 3), knowledge and expertise in the NIS (Item 2), and skills and intelligence in the NIS (Item 1). The reinforcing factor was the SE, incorporating a curriculum for teaching and training based on research (Item 1) and the introduction of a new syllabus in response to emerging challenges (Item 2). These constructs and items were thoughtfully selected and measured to explore their interconnections and contributions to the absorptive capacity of UIRC.

The methodology followed a structured TRIZ problem-solving process to systematically address the challenges within UIRC. To ensure logical consistency, the application of TRIZ encompasses four key stages: problem identification, cause analysis, contradiction analysis, and application of inventive principles. To assess the practical implications of the solutions developed through TRIZ, this study additionally incorporated partial least squares (PLS) analysis. PLS serves as a complementary tool that validates the effectiveness of the strategies identified through the TRIZ framework, thus ensuring that the proposed solutions for enhancing absorptive capacity are grounded in empirical data. The process began with function analysis to identify and outline the functions of the system, followed by cause-and-effect chain analysis (CECA), which was used to trace the root causes of weak absorptive capacity within UIRC. Following this, a contradiction analysis was conducted to identify conflicting factors, such as the dependence on foreign expertise versus the need to develop local talent. Once contradictions were identified, appropriate TRIZ inventive principles, such as prior action, dynamicity, and phase transition, were applied to propose structured solutions. Finally, the effectiveness of these proposed solutions was validated through PLS analysis, which served as a quantitative measure to determine the impact of TRIZ-based strategies on absorptive capacity. By following this structured paradigm, the study ensured that TRIZ was applied in a methodological and systematic manner, avoiding any logical confusion. The detailed methodology can be visualized in Fig. 1.

Moreover, the TRIZ theory, along with the validated statistical programs Statistical Package for the Social Sciences and Smart PLS, was employed to develop and validate our hypothesis. TRIZ models and tools were used to determine the most effective method for improving the absorptive capacity of UIRC. The TRIZ method encourages critical thinking for creative problem-solving. In this context, a function model (Fig. 1) was created to illustrate the seamless process of strengthening the NIS. On the other hand, the CECA (Fig. 2), which is the second feature of TRIZ, revealed

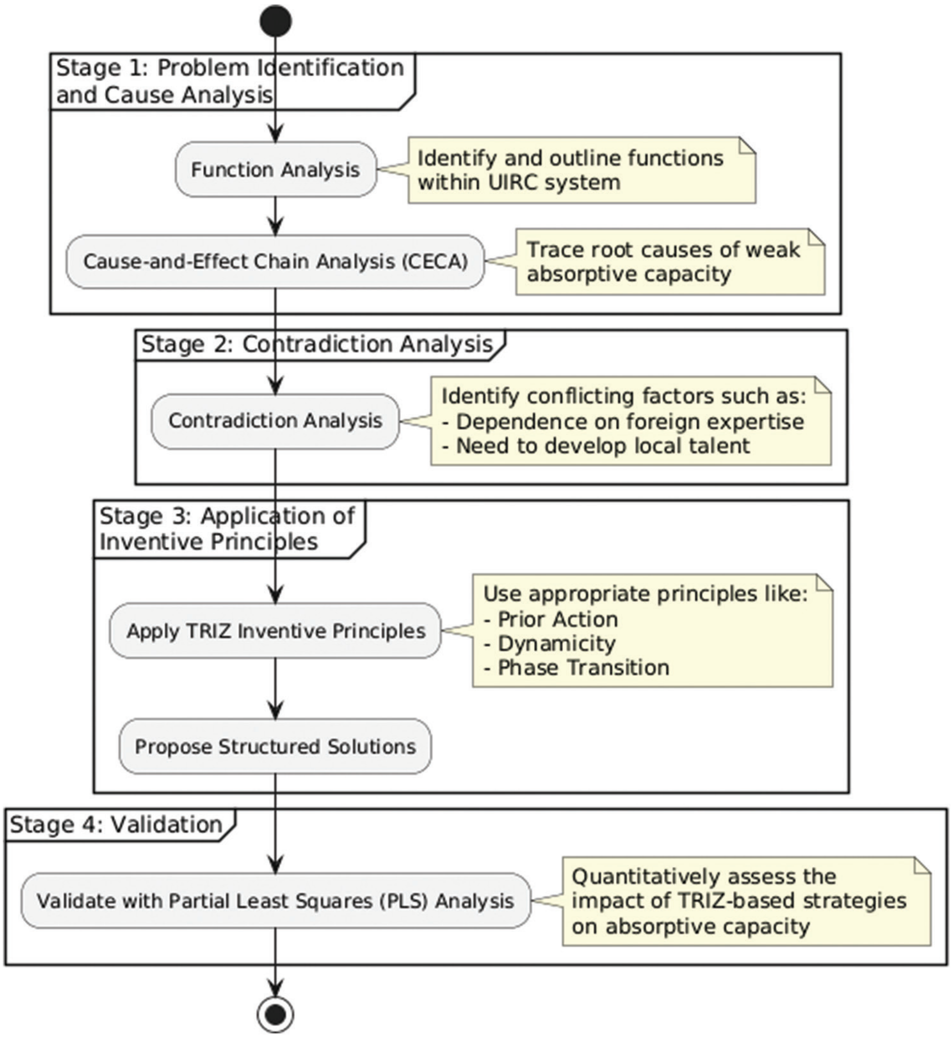


Fig. 1. The Theory of Inventive Problem Solving-based methodology with partial least squares validation for university-industry research collaboration

potential root causes with the function model that can be systematically addressed.

To achieve this, a TRIZ matrix was utilized to identify parameters that were worsening and improving, followed by the identification of proposed creative principles. Table 2 presents one possible approach to resolving the issue. Concurrently, the TRIZ-based framework (Fig. 3) illustrates a tangible, concise, accurate, and systematic solution to the problem.

3.1. Function Model

A function model, also referred to as an activity model or process model, is a systematic representation of the functions, actions, activities, procedures, and operations within a specific field of study. It serves as an organized depiction of the various components that contribute to the functioning of a system, aiding in the identification of ongoing situations, opportunities, information needs, and the description of functions and processes (Valjak & Bojcetic 2022). A function model represents the key processes in UIRC and their impact on strengthening the NIS. Initially, the function "providing real problems" was included, but further analysis revealed that its effectiveness depends on how well industries can absorb and apply academic insights. Simply presenting real-world challenges does not automatically lead to innovation unless structured mechanisms are in place to facilitate knowledge application. The revised function model, therefore, shifts the focus from merely identifying problems to enhancing problem-solving capabilities within collaborations. Similarly, the function "provide knowledge" was reconsidered because knowledge alone does not guarantee innovation. Industries often face difficulties integrating theoretical research into practical solutions. If knowledge transfer occurs without proper application pathways, research outputs may remain underutilized. To address this, the

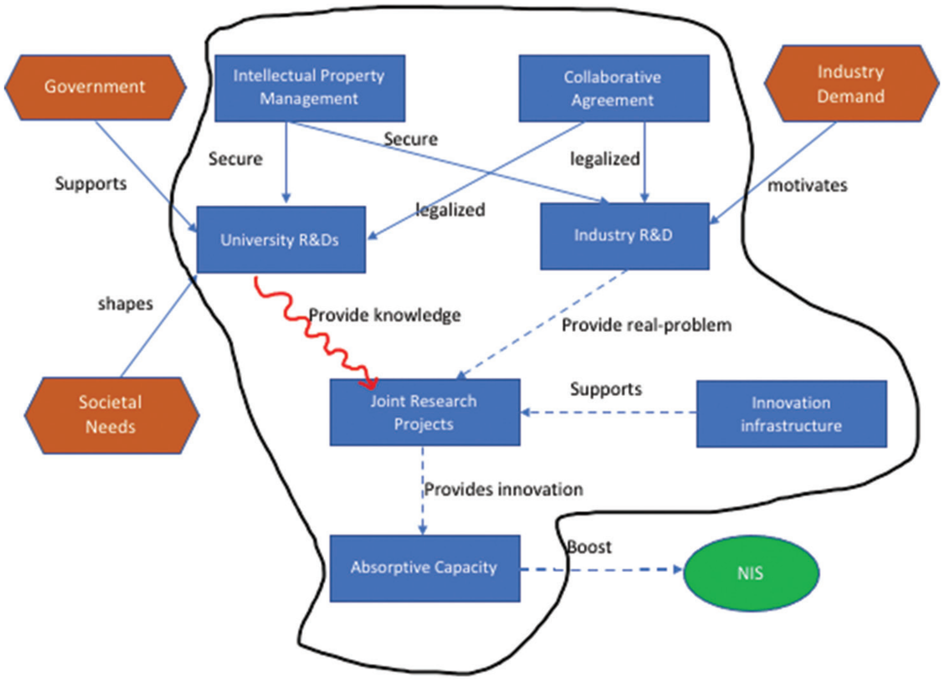


Fig. 2. University-industry research collaboration function model

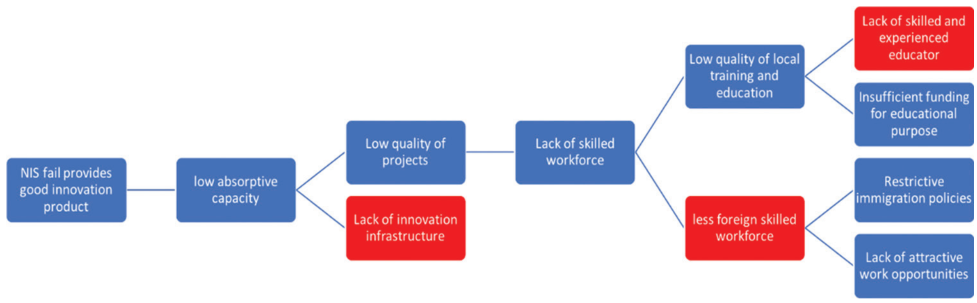


Fig. 3. Cause and effect analysis for the national innovation solution's failure to provide a good innovation product

Table 2. Inventive principles

Inventive principles	
# 10 Prior action # 28 Replacement of mechanical system	# 15 Dynamicity # 36 Phase transition

revised function model now emphasizes co-creation of knowledge, where universities and industries collaborate actively to ensure that research is directly applicable to industrial needs.

Fig. 1 in this context illustrates the key components playing a significant role in research and discoveries. For instance, programs for instruction and training are depicted, showcasing their role in supporting basic research, which in turn, forms the foundational knowledge and comprehension base for the NIS. E&T programs play a crucial role in equipping individuals with the necessary knowledge and skills for research and development initiatives. On the other hand, absorptive capacity is responsible for the recognition, absorption, and utilization of external information, enabling industries to identify and leverage insights from external sources, fostering innovation and growth. This function is intricately connected to UIRC, as the successful transfer and exchange of knowledge depend on effective collaboration between universities and industries. The shared role of collaboration is evident in the intertwined dynamics of absorptive capacity, UIRC, and E&T programs. It underscores the critical importance of collaborative efforts and absorptive capacity in achieving positive outcomes and advancing the national innovation system. Moreover, the functional model depicted in Fig. 2 offers a comprehensive overview of the interactions among various components that contribute to the goal of enhancing the absorptive capacity of NIS through knowledge transfer and effective UIRC. Within the framework of UIRC and NIS, it underscores the significance of E&T, collaborative efforts, and the utilization of knowledge in driving innovation and fostering growth.

3.2. Cause and Effect Chain Analysis

CECA highlights that while real-world problems exist in many industries, the challenge lies in their ability to absorb and implement academic research into practical solutions. The issue is not the absence of real problems but the lack of structured mechanisms to bridge academic research and industrial application. Similarly, knowledge provision alone is not sufficient—without clear implementation strategies, industries may struggle to translate academic insights into innovation. The revised function model thus emphasizes solution-driven collaboration and industry-academia co-creation, ensuring that research findings are effectively utilized.

Additionally, TRIZ analytical thinking employs CECA to identify conflicts related to goals. CECA involves investigating the underlying causes and their relationships to an observed effect, with the primary aim of graphically representing the outcome (Davies et al., 2014). This effect is usually a negative outcome, drawback, or issue that the project aims to address. CECA is utilized to delve into the fundamental reasons behind a specific occurrence (Crown et al., 2020). For instance, in the context of the NIS, the primary issue is its limited absorptive capacity. Incorporating FES can potentially enhance the volume of research and innovation projects. However, CECA studies and previous research indicate that FESW may have a negative impact on the unemployment rate.

Table 3 highlights the problems and their underlying causes in the hiring of FESW.

In the context of the research, the table delineates specific challenges along with their respective underlying causes. The first challenge pertains to insufficient absorptive capacity, and it is discerned that a workforce lacking in education and competence is the primary driver of this issue. Challenge 2, centering on the presence of a skilled and educated immigrant labor force, is traced back to the inadequacy of a competent and educated local workforce. Moreover, Challenge 3 underscores the problem of unemployment, attributing it to low educational standards and the recruitment of foreign workers as the main contributing factors. This table succinctly summarizes the central issues related to employment and absorptive capacity, providing clarity by establishing a direct link between each challenge and its underlying causes. Table 4 and Fig. 3 indicate that the country's lack of educated and trained workers is the primary reason for low absorptive capacity. Therefore, while FESW can increase the workforce in research companies, it may also contribute to a decline in the employability rate.

Table 3. Challenges and root causes

Challenges	Root causes
Low absorptive capacity	Lack of a local educated and skilled workforce
Foreign-educated and skilled workforce	Lack of an educated and skilled workforce
Unemployment	Number of hiring foreign workforce Low standard of education

Table 4. If then contradiction analysis

If-than-but	Statements
If	Hired foreign-educated and skilled workforce
Then	Enhanced number of research and innovation
But	Unemployment of the local workforce increased
Improving parameter	Productivity
Worsening parameter	Force (Intensity)
Recommended inventive principle	Prior action, dynamicity, replacement of mechanical system, and phase transition

Based on Fig. 3, we formulated an engineering contradiction model as shown in Table 4. Table 4 provides problem modeling using engineering contradiction. This model analyzed the basic challenge to enhance the research collaboration between universities and industries and NIS, specifically hiring the FESW to improve the number of research and innovations. However, hiring FESW will increase unemployment among the local educated workforce. The detailed solutions of this model are discussed in the following section.

From the contradiction analysis, it has been analyzed that the basic challenge to enhance the research collaboration between universities and industries and NIS is the lack of an educated and skilled workforce. To resolve the matter, it is mandatory to refurbish the absorptive capacities of the researchers by improving the SE. In this regard, Table 3 provides a comprehensive overview of the issues along with their underlying causes.

3.2.1. Applying inventive principle

The application of inventive principles is geared toward stimulating creative thinking to address technical challenges (Krejcie & Morgan, 1970). The objective of research on innovative principles is to refine principal definitions by providing a comprehensive list of analogous examples and customizing definitions for specific domains, such as manufacturing, business, informatics, and chemistry (Valjak & Bojectic, 2023). In this context, TRIZ recommends inventive principles (Table 2) as potential solutions for the challenges. Table 2 delineates specific innovative principles designed to address challenges, with each principle offering a distinct perspective on problem resolution. The first principle, "prior action," underscores the importance of establishing robust policies and strategies in advance to navigate through challenging situations effectively. The second concept, "dynamicity," highlights the significance of adaptability in plans and programs to address shifting conditions. The third principle, "replacement of mechanical systems," suggests that better outcomes may result from modernizing or replacing antiquated systems. The fourth and final premise, "phase transition," proposes a modification or overhaul of the educational system to achieve constructive outcomes. Each tenet presents a unique strategy for overcoming obstacles, all contributing to the shared objective of enhancing employment rates and absorptive capacity. The inventive principles were developed through a thorough examination of creative solutions from previous patents.

The TRIZ toolkit outlines four potential solutions to address the issues of unemployment and insufficient absorptive capacity. For instance, the innovative principles suggest implementing phase transition, dynamicity, prior action, and mechanical system replacement to enhance the nation's employability and increase its absorption capacity. To delve into specifics, the principle of previous activity advocates for robust policies and plans for each affair, while dynamicity encourages flexibility in these areas. Similarly, phase transitions and mechanical system replacements amplify the notion of action shifts. The innovative ideas presented in the TRIZ toolkit offer insights into potential strategies that can be applied to enhance absorptive capacity and increase the employment rate. Table 5 provides a detailed explanation of the recommended creative principles, offering a comprehensive guide for implementing strategies to address the challenges of absorptive capacity and unemployment.

Table 5 provides a comprehensive overview of specific creative principles, their associated imaginative triggers, and the potential solutions that can be derived from their application. The inventive trigger in focus is the imperative need for a change in the national curriculum, index, and standard of education. The first principle, "prior action," advocates taking the necessary steps proactively and in advance. The second principle, "dynamicity," underscores the importance of a dynamic environment for optimal performance, with potential solutions centered around enhancing flexibility and adaptability. These solutions

Table 5. Inventive principles descriptions

Invention principle	**Inventive trigger as a guideline**	**Potential solutions from the suggested inventive principles**
# 10 Prior action	Perform the required change in advance	The standard of education/index of education/and curriculum of education must be changed at the national level.
# 15 Dynamicity	The environment must be dynamic to provide optimal performance/ enhancing flexibility and adaptability	Increase the dynamicity of education and training programs to better align with industry needs/ SHuman capital must be transferred into intellectual capital, so the absorptive capacity can be enhanced.
# 28 Replacement of mechanical systems	Replace the system that is fixed with changes over time	The number of local employment opportunities must be increased by providing up-to-date education and training.
#36 Phase transition	Using the phenomenon of phase change	Increase the innovative capabilities of the nation by changing the old prospectus to a new one.

involve making E&T programs more dynamic to meet industry demands and transforming human capital into intellectual capital to enhance absorptive capacity. The third principle, "replacement of mechanical system," suggests replacing outdated E&T with more modern approaches to increase local employment, replacing fixed systems that are unable to adapt to changing times. The fourth principle, "phase transition," utilizes the phenomenon of phase shift to propose practical ways to enhance the country's innovation capacity by repurposing outdated prospectuses. This approach suggests not only replacing old frameworks with new ones but also adapting and improving existing structures to meet modern demands.

Table 6 offers a systematic overview of the challenges at hand, their root causes, and recommended solutions for effective problem resolution. The primary challenge involves the low absorptive capacity within universities and industries, primarily attributed to an

Table 6. Challenges, root causes, and solutions

Challenge	**Root cause**	**Solution**
Challenge 1: Low absorptive capacity in universities and industries	Lack of an educated and skilled workforce	Solution 1: Prioritize hiring foreign-educated and skilled workforce to improve absorptive capacity
		Solution 2: Increase dynamicity in education and training programs to better align with industry needs.
Challenge 2: High unemployment rate	Low standard of education	Solution 3: Replace outdated mechanical systems with modern ones to improve productivity and competitiveness
		Solution 4: Implement phase transition in education by increasing the index of education to better prepare graduates for industry demands

inadequately educated and skilled workforce. The table delineates two principal approaches to tackle this issue: prioritizing the hiring of foreign-educated and skilled workers to enhance absorptive capacity and enhancing the dynamism of E&T programs to align more effectively with the evolving needs of industry. The secondary challenge pertains to a high unemployment rate, rooted in the inadequacies of the education system. In response, the table proposes two strategic measures: first, integrating contemporary mechanical systems to enhance competitiveness and productivity, and second, implementing a phase transition in education by elevating the educational index to better prepare graduates for industry expectations. These collective strategies aim to address the underlying issues, fostering an environment that supports improved absorptive capacity and reduced unemployment rates through strategic interventions in education, workforce dynamics, and systemic changes.

Following a comprehensive explanation of the innovative principles, they were practically applied to address real-time problems. Table 6 outlines the identified challenges along with the root cause and solutions. Solution 1, prioritizing the employment of foreign-educated and talented labor, and Solution 2, enhancing dynamicity in E&T programs, offer effective measures to overcome Challenge 1, characterized by low absorptive capacity in universities and industries. On the other hand, Solution 3, replacing outdated mechanical systems, and Solution 4, implementing phase transition in education through improving the index of education, are proposed to tackle Challenge 2, which involves the high unemployment rate resulting from the hiring of skilled workers with foreign education. These solutions reflect a strategic approach to address the identified challenges and promote positive outcomes.

Fig. 4 serves as a visual aid to enhance understanding of the relationship between each solution and the identified challenges. These visuals facilitate the decision-making process by providing a clear representation of how well each solution addresses the specific contradictions and underlying issues outlined in the problem statement. They contribute to a comprehensive and organized approach to evaluating and ranking solutions based on their effectiveness in resolving the identified challenges related to absorptive capacity and unemployment. Furthermore, Fig. 4 clearly demonstrates that employing creative ideas can effectively tackle issues related to low absorptive capacity in academic institutions and businesses, as well as address national unemployment. For instance, leveraging past practices and introducing dynamism can serve as proactive measures for a workforce with foreign education and expertise. Similarly, addressing the second problem can involve implementing creative ideas, such as replacing outdated mechanical systems and introducing phase transitions. For example, revamping the obsolete educational system requires the adoption of new teaching and training methodologies. Instead of a curriculum based on theory, educational institutions must transition to one grounded in research. The advancement of research and innovations, as well as researchers' capacity to absorb information, is insufficiently bolstered by theoretical knowledge alone. Thus, the quality of education plays a significant role in enhancing the capacity of universities, industries, and the NIS. TRIZ's approach provides a means to alleviate the challenges by emphasizing the significance of E&T for UIRC and the NIS. A detailed and comprehensive description is presented in Fig. 4. Fig. 4 outlines the parameters of the "index of education and educated and skilled workforce," illustrating the proposed solutions to mitigate the challenges related to absorptive capacity. Proposed solutions encompass improving the absorptive capacities of universities and industries, the number of research and innovations, and fortifying the NIS. According to the TRIZ-based framework to enhance the absorptive capacity of the NIS requires a modification in the education index or standard. The inventive principles, such as phase transitions, mechanical system replacements, prior action, and dynamicity, contribute to enhancing the

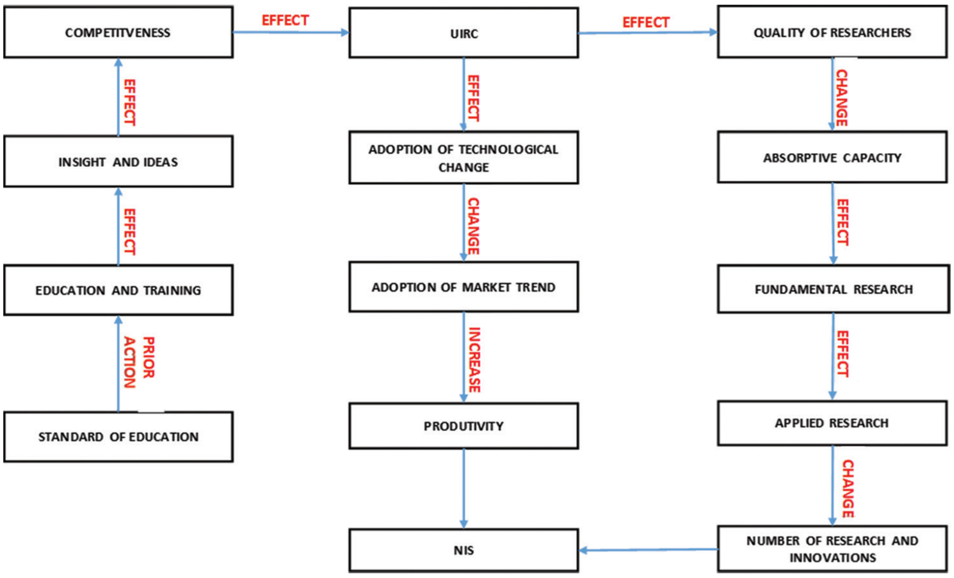


Fig. 4. Theoretical framework using the Theory of Inventive Problem Solving approach
Abbreviations: NIS: National innovative solutions; UIRC: University-industry research collaboration

perception of action shifts. They provide insights into potential strategies that can be employed to improve the absorptive capacity of NIS.

4. Results and Analysis

In this study, both PLS analysis and the Statistical Package for Social Sciences software were utilized for data analysis. It is important to highlight that each variable in this study was employed in a formative manner. Consequently, the evaluation of the formative path model necessitates completing both the measurement and structural model assessments (Labuda, 2015). This approach ensured a comprehensive understanding of the relationships and interactions between the variables, allowing for a thorough analysis of the formative nature of the path model in the study.

To strengthen the connection between TRIZ and PLS, it is essential to clarify the role of PLS as a validation tool rather than an independent analytical framework. While TRIZ provides theoretical solutions based on systematic problem-solving heuristics, PLS analysis empirically tests whether these solutions align with real-world data. If PLS results support TRIZ solutions, it provides quantitative validation that the proposed strategies effectively enhance absorptive capacity within UIRC. However, if PLS results contradict TRIZ, several explanations must be considered. First, a mismatch between theoretical solutions and practical implementation may indicate that while TRIZ suggests effective strategies, they may require additional contextual adaptation before being applicable in specific industry settings. Second, measurement issues within the PLS model could lead to discrepancies if certain constructs do not fully capture the intended aspects of TRIZ-based solutions, requiring adjustments to the model design. Third, a contradiction between TRIZ and PLS could highlight limitations in industry absorptive capacity, where industries may lack the necessary readiness, infrastructure, or policy support to implement TRIZ-based solutions effectively.

By integrating TRIZ with PLS in this manner, a strong correlation is established between theoretical problem-solving and empirical validation, ensuring a logically consistent approach. The study acknowledges that while TRIZ serves as a structured innovation methodology, PLS provides an evidence-based verification mechanism to assess whether proposed strategies yield tangible improvements in absorptive capacity. If discrepancies arise, rather than dismissing TRIZ or PLS findings, further analysis should be conducted to determine the underlying cause of the inconsistency. This approach ensures that TRIZ applications remain logically sound while maintaining a rigorous empirical foundation for validating proposed solutions.

4.1. Assessment of the Measurement Model

The measurement model in this study was designed to elucidate the process of measuring latent constructs. It focuses on assessing the validity of the constructs and their indicators to ensure the accuracy and reliability of the measurements. The goal is to thoroughly evaluate the measurement qualities, ensuring that the indicators effectively capture the intended latent constructs and contribute to the overall robustness of the model. This meticulous examination is crucial for establishing the credibility and effectiveness of the measurement model employed in the study.

4.2. Assessment of Construct Validity

The importance of avoiding redundancy among constructs at the construct level was emphasized to prevent multicollinearity, which is inferred for each construct in this study. Multicollinearity occurs when two or more variables in the model exhibit a significant correlation, leading to potential issues. In the presence of multicollinearity, regression coefficient estimates become unreliable. The study made sufficient efforts to operationalize the five variables effectively, aiming to capture their distinct contributions without introducing redundant or overlapping elements. The goal is to ensure the reliability and accuracy of the regression analysis and subsequent interpretations of the model.

Table 7 encompasses crucial information on constructs, their corresponding indicators, and collinearity statistics, specifically focusing on tolerance and variance inflation factor (VIF). These metrics are fundamental for the meticulous evaluation of the measurement model's quality. The objective of this model is to elucidate the measurement process of latent constructs, aiming to assess the reliability of both the constructs and their associated indicators. The primary emphasis is placed on four key constructs: SE, FESW, E&T, and absorptive capacity. Regarding absorptive capacity, AC_1 and AC_2 exhibited tolerances of 0.332 and 0.410, and VIF values of 3.012 and 1.260, respectively. In the E&T construct, indicators ET_1, ET_2, ET_3, and FW_1 displayed varied tolerance and VIF values, indicating potential collinearity concerns.

Table 7. Assessment of construct validity

Constructs	Indicators	Collinearity statistics	
		Tolerance	Variance inflation factor
Absorptive capacity	AC_1	0.332	3.012
	AC_2	0.410	1.260
Education and training	ET_1	0.831	1.203
	ET_2	0.757	1.320
	ET_3	0.925	3.171
	FW_1	0.811	2.233
Foreign-educated and skilled–workforce	FW_2	0.317	1.162
	FW_3	0.202	1.214
Standard of education	SE_1	0.844	1.125
	SE_2	0.922	1.037

The FESW indicators, FW_2 and FW_3, showed signs of collinearity based on their tolerance and VIF values. The SE construct, represented by SE_1 and SE_2, demonstrated acceptable tolerance and VIF values. Values exceeding five for VIF or falling below 0.2 for tolerance may signify problematic collinearity. These statistics play a pivotal role in evaluating the dependability of the measurement model and guide researchers in addressing potential issues with the indicators of the constructs for a comprehensive assessment of the structural model.

As stated by Cong and Tong (2008), a tolerance level higher than 0.20 and a VIF value not exceeding five are considered acceptable to avoid multicollinearity issues. The VIF test results are presented in Table 3, following the completion of the stepwise regression analysis for each construct. The findings indicate no evidence of multicollinearity, as all tolerance values surpass 0.20, and all VIFs remain below the thresholds. This suggests that the variables within each construct maintain sufficient independence and do not exhibit problematic correlations, meeting the recommended criteria for reliable regression analysis.

4.3. Assessment of Indicators' Validity

The indicator-level analysis addresses the question of whether each indicator effectively contributes to its respective construct by conveying the intended meaning. Strong relationships between constructs and indicators are recommended. In this regard, the weight of each indicator is determined to assess its relevance to the construct (Kim & Park, 2023; Russo & Spreafico, 2015). PLS calculates indicator weights, with a recommended threshold of $p < 1/\sqrt{n}$ to evaluate their contributions. In this research, each construct has a minimum of two and a maximum of three indicators, resulting in *p*-values of 0.709 and 0.578 for constructs with two and three indicators, respectively, as outlined in Table 4. These *p*-values indicate the significance of each indicator's contribution to its corresponding construct. The table illuminates the loadings and weights of the indicators for each construct in the model. These values are crucial for assessing the significance of each indicator's contribution to its corresponding construct. In the absorptive capacity construct, the weights of 0.3450 and 0.2895 for AC_1 and AC_2, respectively, indicate their relative importance. The indicator loadings, represented by 0.6118 and 0.9966, further elucidate the direction and degree of the association between the construct and the indicators. For the E&T construct, ET_1 and ET_2 exhibit weights of 0.4054 and 0.3272, while their loadings of 0.7208 and 0.8116 highlight their contributions. ET_3, with a loading of 0.9906 and a weight of 0.3319, similarly has a considerable impact on the construct. Similarly, FW_1, FW_3, FESW, SE, and SE_1 construct weights emphasize their contributions. Overall, the table facilitates understanding of the relative contributions of each indicator to its particular construct with respect to loading and weight.

The indicators' weights for each connected component are presented in Table 8, showcasing their substantial contribution to the variance in their respective constructs, as indicated by the significant item weights. However, the occurrence of two indicators, specifically "FW (FW_2)" and "SE (SE_2)," varies due to their formulaic values. In this context, when indicator weights are not significant at ($p < 1/\sqrt{n}$), item loadings become essential, as suggested by Hair et al. (2014). Remarkably, all of the constructs' item loadings were deemed significant ($p > 0.50$), underscoring the profound significance and relevance of each construct with its corresponding indicators. Following the establishment of a valid measurement model, PLS analysis was conducted to assess the structural model in the subsequent phase of the study.

4.4. Assessment of the Structural Model

The study investigated three main relationships (H1a, H2) and one reinforcing effect (H1b) proposed for the structural model. The examination of the structural model involved the assessment of R^2 values and pathway coefficients. The findings of the hypotheses are illustrated using the research model, which is presented in two stages in this study. Figs. 5 and 6 provide a visual representation of how FESW impacts UIRC and absorptive capacity, and how E&T influences UIRC and absorptive capacity, respectively. These visualizations offer a clear understanding of the relationships between the inventive principle and the key variables in the study.

Table 8. Assessment of indicators' validity

Constructs	Indicators	Indicators weight (t−V)	Indicators loading (t−V)
Absorptive capacity	AC_1	0.3450	0.9966
	AC_2	0.2895	0.6118
	ET_1	0.4054	0.7208
	ET_2	0.3272	0.8116
Education and training	ET_3	0.3319	0.9906
	FW_1	0.1445	0.6726
Foreign-educated and skilled workforce	FW_2	−0.0476	0.8234
	FW_3	0.0393	0.9667
Standard of education	SE_1	0.6093	0.5882
	SE_2	−0.3969	0.7512

In terms of the structural model, the R^2 values and path coefficients (β) are crucial for explaining the variances in absorptive capacity attributed to the impacts of FESW and UIRC. For instance, the β value of FESW and UIRC, represented as 0.497 and 0.542, respectively, highlights their significant impact on absorptive capacity. Correspondingly, the R^2 values of absorptive capacity, amounting to 11.6%, underscore the substantial contribution of FESW to the variance in absorptive capacity. Thus, Fig. 5 establishes that FESW plays a critical role in enhancing both UIRC and their collaborative absorptive capacity.

Furthermore, Fig. 6 illustrates the impact of E&T on UIRC and absorptive capacity, portraying their significant role on both variables. Path coefficient values of E&T (1.985 and 1.236) indicate a significant influence on absorptive capacity. This depiction in Fig. 6 emphasizes the importance of E&T in fostering UIRC and contributing to their joint impact on absorptive capacity.

The R^2 values for absorptive capacity in the context of E&T stand at 21.3%, indicating a substantial contribution of E&T to the variance in absorptive capacity. Moreover, the path coefficients of E&T (1.985) and UIRC (1.236) were elevated by introducing the reinforcing component SE from 1.985 to 2.708 and UIRC from 1.236 to 2.103, respectively. This reinforcement resulted in an increase in the variations R^2 from E&T to absorptive capacity, escalating from 21.3% to 31.8%. Consequently, Fig. 6 illustrates the significant reinforcing role of SE in enhancing E&T and UIRC efficiency, subsequently improving absorptive capacity.

The analysis conducted in this study underscores the significance of the E&T factor in enhancing the absorptive capabilities of the NIS. While it is widely recognized that the presence of FESW contributes to the development of research and innovations, the study's results reveal an interesting finding. Contrary to the initial hypothesis that FESW would have a more substantial positive impact on NIS's absorptive capacity, the research indicates that E&T, with an R^2 value of 21.3%, has a greater and more significant influence on NIS's absorptive capacity compared to FESW, which has an R^2 value of 11.6%.

The research findings challenge initial assumptions about the perceived significance of FESW in relation to absorptive capacity. Contrary to expectations, the study indicates that E&T emerges as a more influential and desirable element for enhancing the absorptive capacity of NIS compared to FESW. Moreover, the study contributes to existing literature by identifying SE as a reinforcing element. The analysis suggests that while E&T at the national

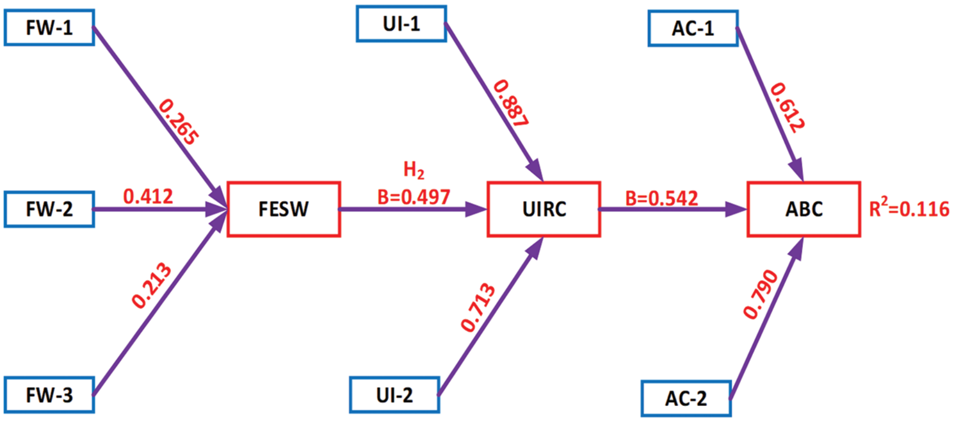


Fig. 5. Effect of FESW on UIRC and ABC
Abbreviations: FESW: Foreign-educated and skilled workforce; UIRC: University-industry research collaboration, ABC: Absorptive capacity

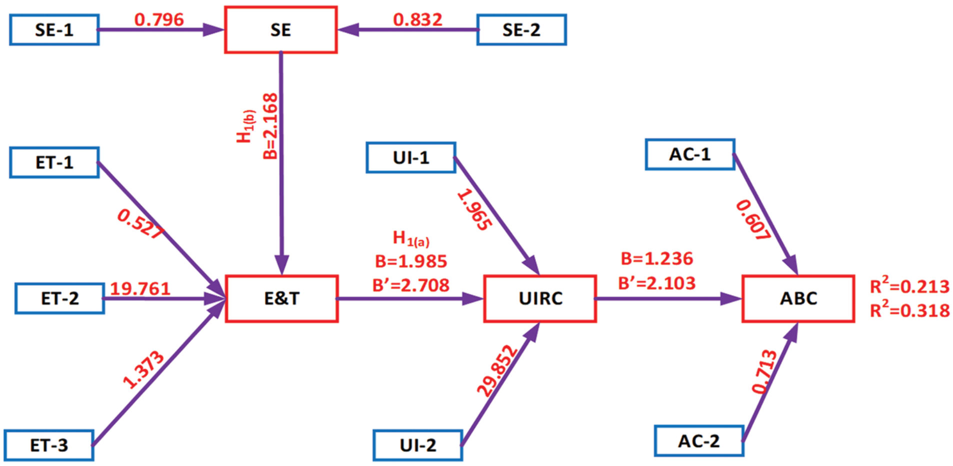


Fig. 6. Effect of E&T on UIRC and ABC
Abbreviations: E&T: Education and training; UIRC: University-industry research collaboration; ABC: Absorptive capacity

level can enhance the competence of UIRC, SE plays a crucial reinforcing role, providing an additional advantage in improving NIS's absorptive capacity. This nuanced understanding contributes valuable insights toward the inventive principles (prior action, dynamicity, replacement of mechanical system and phase transition) that provide a valuable solution to not only mitigate the FESW and employability disorders from the country but also provide a considerable pathway to enhance the absorptive capacities of the researchers and consequently the number of research and innovations in NIS.

5. Conclusion and Future Direction

In conclusion, this research offers enduring pathways for enhancing the absorptive capacity of UIRC and provides profound insights into the realm of NIS. Despite numerous studies focusing on UIRC assessment, the present research stands out as theory-based, filling a critical gap in existing literature. By underscoring the substantial influence of UIRC on the NIS, this article endeavors to bridge the divide between research and innovation organizations and policymakers. The findings aim to inform strategic decisions, fostering a more synergistic relationship between academia, industry, and policymakers in the pursuit of innovation and technological advancement.

The study makes several noteworthy contributions to the ongoing discourse on research and innovation projects involving collaborations between academic institutions and business sectors. First, it provides practical solutions within the framework of TRIZ, shedding light on the limitations and opportunities inherent in such studies. TRIZ introduces a fresh perspective by pinpointing absorptive capacity as a vital factor influencing the inventive performance of UIRC, thus strengthening the NIS.

Second, the research delves into the significant constraints faced by UIRC, with specific emphasis on the role of E&T. This focus underscores the importance of intellectual capital development and the learning processes within UIRC, contributing valuable insights into the challenges and opportunities associated with educational initiatives in the context of collaborative research and innovation, rather than hiring FESW. The study, therefore, extends the understanding of the multifaceted dynamics involved in UIRC, providing a foundation for informed decision-making and strategic planning in this domain. The versatility of the TRIZ framework positions it as an invaluable tool for policymakers seeking practical solutions across various scenarios. The study underscores the effectiveness of TRIZ by generating creative ideas (prior action, dynamicity, replacement of mechanical system, phase transition), emerging as a potent strategy to address limitations in absorptive capacity. This finding adds depth to the existing literature by highlighting the important role of E&T initiatives in enhancing UIRC's capacity for innovation.

Moreover, the research contributes to the scholarly landscape by introducing a reinforcing factor, SE, as a crucial component that fortifies UIRC's E&T capacity. These supportive elements are encompassed in intellectual capital development, fostering robust collaborations between academia and industry, diminishing the dependencies on FESW, and cultivating an independent innovation ecosystem. Recognizing and nurturing these factors becomes paramount for maximizing UIRC's potential to drive innovation. This insight provides policymakers and stakeholders with a nuanced understanding of the intricate elements that contribute to a thriving collaborative research and innovation landscape.

In practical terms, policymakers stand to gain significant insights from the TRIZ approach, as it furnishes them with a comprehensive understanding of how UIRC dynamics influence the NIS. This study aspires to offer a robust framework that enhances UIRC's innovation capacity, providing policymakers with clarity on the intricate interconnections among academia, industry, and the broader innovation ecosystem. By leveraging the TRIZ methodology, legislators can make informed decisions and formulate effective policies that foster a conducive environment for collaborative research, innovation, and the overall advancement of the NIS.

Furthermore, this study suggests intriguing avenues for future research aimed at overcoming the challenges encountered in this study. By expanding the scope beyond Malaysia, conducting comparative analyses among various developed and emerging nations can unveil significant disparities in the enhancement of the NIS's creative potential. Such cross-national analyses have the potential to enhance our understanding of the intricate dynamics propelling innovation within UIRC and its broader implications for NISs. This comparative approach offers the opportunity to glean valuable insights from different contexts, contributing to a more comprehensive and nuanced understanding of the factors influencing innovation on a global scale.

This study essentially functions as a lighthouse, illuminating the path toward a more comprehensive UIRC that integrates the NIS. The insights gained from this research provide a roadmap for policymakers to optimize UIRC, offering guidance to nations in fostering a more innovative and promising technological future. By emphasizing the interconnectedness of UIRC with broader innovation systems and educational standards, this study provides valuable signposts for policymakers seeking to navigate the complexities of enhancing innovation and technological advancement on a national scale.

Acknowledgment

This work was supported by the Research Management Centre (iRMC), i-CATS University College.

References

Ahmed, S.S., Guozhu, J., Mubarik, S., Khan, M., & Khan, E. (2020). Intellectual capital and business performance: The role of dimensions of absorptive capacity. Journal of Intellectual Capital, 21(1), 23–39. https://doi.org/10.1108/JIC-11-2018-0199

Ahn, J.M., Ju, Y., Moon, T.H., Minshall, T., Probert, D., Sohn, S.Y., et al. (2016). Beyond absorptive capacity in open innovation process: The relationships between openness, capacities, and firm performance. *Technology Analysis Strategic Management,* 28(9), 1009–1028. https://doi.org/10.1080/09537325.2016.1181737

Akram, H. (2020). Education governance in Pakistan: A critical analysis of challenges. *Journal Social Science Advancement*, 1(1), 38–41. https://doi.org/10.52223/JSSA20-010105-05

Albort-Morant, G., Leal-Millán, A., & Cepeda-Carrión, G. (2016). The antecedents of green innovation performance: A model of learning and capabilities. *Journal of Business Research,* 69(11), 4912–4917. https://doi.org/10.1016/j.jbusres.2016.04.052

Alexander, C., & Yuriy, H. (2015). Problems and perspectives of performance of higher education institutions in the development of Russian innovative system (regional aspect). *Procedia*

Social Behavioral Science, 166, 497–504. https://doi.org/10.1016/j.sbspro.2014.12.561

Alrowwad, A., Abualoush, S.H., & Masa'deh, R. (2020). Innovation and intellectual capital as intermediary variables among transformational leadership, transactional leadership, and organizational performance. *Journal of Management Development*, 39(2), 196–222. https://doi.org/10.1108/JMD-02-2019-0062

Arshad, M.Z., Arshad, D., Lamsali, H., Alshuaibi, A.S.I., Alshuaibi, M.S.I., Albashar, G., et al. (2023). Strategic resources alignment for sustainability: The impact of innovation capability and intellectual capital on SME's performance. Moderating role of external environment. *Journal Cleaner Production,* 417, 137884. https://doi.org/10.1016/j.jclepro.2023.137884

Carrasco-Carvajal, O., García-Pérez-De-Lema, D., & Castillo-Vergara, M. (2023). Impact of innovation strategy, absorptive capacity, and open innovation on SME performance: A Chilean case study. *Journal of Open Innovation Technology Market and Complexity*, 9(2), 100065. https://doi.org/10.1016/j.joitmc.2023.100065

Chao-Ton Su, Chin-Sen Lin. (2008) A case study on the application of Fuzzy QFD in TRIZ for service quality improvement, *Qual Quant,* 42:563–578, https://doi.org/10.1016/j.joitmc.2023.100065

Chen, Y., Han, J., Xuan, Z., & Gao, W. (2019). Higher education's role in Chinese national innovation system: A perspective of university-industry linkages. Vol. 1. In: *17th International Conference on Scientometrics and Informetrics (ISSI 2019 - Proceedings*), p573–583.

Chryssou, C.E. (2020). University-industry interactions in the Sultanate of Oman: Challenges and opportunities. *Industry and Higher Education*, 34(5), 342–357. https://doi.org/10.1177/0950422219896748

Cong H, Tong L.H. (2008), Grouping of TRIZ inventive principles to facilitate automatic patent classification, *Expert Systems with Applications*, 34(1), 788–795.

Czinki, A., & Hentschel, C. (2016). Solving complex problems and TRIZ. *Procedia CIRP*, 39, 27–32. https://doi.org/10.1016/j.procir.2016.01.161

Davies, D., Jindal-Snape, D., Digby, R., Howe, A., Collier, C., & Hay, P. (2014). The roles and development needs of teachers to promote creativity: A systematic review of literature. *Teaching and Teacher Education*, 41, 34–41. https://doi.org/10.1016/j.tate.2014.03.003

Dong, L., Fu, S., Liu, B., & Yin, B. (2023). Comparison and quantitative assessment of two regional soil erosion survey approaches. *International Soil and Water Conservation Research*, 11(4), 660–668. https://doi.org/10.1016/j.iswcr.2023.04.004

Dooley, L., & Gubbins, C. (2019). Inter-organisational knowledge networks: Synthesising dialectic tensions of university-industry knowledge discovery. *Journal of Knowledge Management*, 23(10), 2113–2134. https://doi.org/10.1108/JKM-06-2018-0343

Ekmekci, I., & Nebati, E.E. (2019). TRIZ methodology and applications. *Procedia Computer Science*, 158, 303–315. https://doi.org/10.1016/j.procs.2019.09.056

Ersin, F. (2009). *Implementation of TRIZ Methodology in Human Capital Master's Thesis*. Bahcesehir University, Istanbul, Turkey.

Farzaneh, M., Wilden, R., Afshari, L., & Mehralian, G. (2022). Dynamic capabilities and innovation ambidexterity: The roles of intellectual capital and innovation orientation. *Journal of Business Research,* 148, 47–59. https://doi.org/10.1016/j.jbusres.2022.04.030

Crown, D., Faggian, A., & Corcoran, J. (2020). Foreign-born graduates and innovation: Evidence from an Australian skilled visa program. *Resarch Policy*, 49(9), 103945. https://doi.org/10.1016/j.respol.2020.103945

Groeneveld, S., Tummers, L., Bronkhorst, B., Ashikali, T., & Van Thiel, S. (2015). Quantitative methods in public administration: Their use and development through time. *International Public Management Journal*, 18(1), 61–86. https://doi.org/10.1080/10967494.2014.972484

Gu, H. (2021). Understanding the migration of highly and less-educated labourers in post-reform China. *Applied Geography,* 137, 102605. https://doi.org/10.1016/j.apgeog.2021.102605

Hammer, J., & Kiesel, M. (2019). A TRIZ and lean-based approach for improving development processes: Creating and managing innovations. In: *Advances in Systematic Creativity.* Springer, Berlin, pp101–114.

Hair, J.F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V.G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review*, 26(2), 106–121. https://doi.org/10.1108/EBR-10-2013-0128

Van Der Heiden, P., Pohl, C., Mansor, S.B., & Van Genderen, J. (2015). The role of education and training in absorptive capacity of international technology transfer in the aerospace sector. *Progress in Aerospace Science*, 76, 42–54. https://doi.org/10.1016/j.paerosci.2015.05.003

Iqbal, A.M., Aslan, A.S., & Khan, A.S. (2010). Research collaboration agreements: A major risk factor between university-industry collaboration. In: *Proceedings of the 3rd International Graduate Conference on Engineering, Science and Humanities (IGCESH 2010)*. Johor Bahru, Malaysia.

Iqbal, A.M., Iqbal, S., Khan, A.S., & Senin, A.A. (2013). A novel cost efficient evaluation model for assessing research-based technology transfer between university and industry. *Journal Tecnology Sciences and Engineering*, 64(2), 24. https://doi.org/10.11113/jt.v64.2242

Iqbal, A.M. (2018). *Influence of National Innovation System on University-Industry Research Collaboration (Doctoral Thesis). Universiti Teknologi Malaysia, Malaysia*.

Iqbal, A.M., Kulathuramaiyer, N., Khan, A.S., & Abdullah, J. (2022). Analyzing the role of human capital in strengthening national innovation system through university-industry research collaboration: A TRIZ-based approach. In: Nowak, R., Chrząszcz, J., & Brad, S., editors. *Systematic Innovation Partnerships with Artificial Intelligence and Information Technology: TFC 2022. IFIP Advances in Information and Communication Technology*. Vol. 655. Springer, Berlin, p417–428.

Iqbal, A.M., Khan, A.S., & Senin, A.A. (2012). Determination of high impact evaluation metrics for evaluating the university-industry technological linkage. *International Journal of Physical and Social Science*, 2(4), 111–122.

Iqbal, A.M., Khan, A.S., Bashir, F., & Senin, A.A. (2015). Evaluating national innovation system of Malaysia based on university-industry research collaboration: A system thinking approach. *Asian Social Science*, 11(13), 45. https://doi.org/10.5539/ass.v11n13p45

Iqbal, A.M., Khan, A.S., & Senin, A.A. (2015). Reinforcing the national innovation system of Malaysia based on university-industry research collaboration: A system thinking approach. *International Journal Management Science Business Research*, 4(1), 6–15.

Iqbal, S., Iqbal, A.M., Khan, A.S., & Senin, A.A. (2013). A modern strategy for the development of academic staff based on university-industry knowledge transfer effectiveness & collaborative research. *Sains Humanika*, 64(3), 35–38. https://doi.org/10.11113/sh.v64n3.64

Iqbal, A.M., Khan, A.S., Iqbal, S., Abdullah, J., Kulathuramaiye, N., & Senin, A.A. (2021). Blended system thinking approach to strengthen the education and training in university-industry research collaboration. *Technology Analysis and Strategic Management,* 34(4), 447–460. https://doi.org/10.1080/09537325.2021.1905790

Iqbal, A.M., Kulathuramaiyer, N., Khan, A.S., Abdullah, J., & Khan, M.A. (2022). Intellectual capital: A system thinking analysis in revamping the exchanging information in university-industry research collaboration. *Sustainability*, 14(11), 6404. https://doi.org/10.3390/su14116404

Ishrat, I., Hasan, M., Khan, F.M., & Javed, M.Y. (2023). Unraveling the structure and trends of TRIZ approach in business and management: Bibliometric synthesis and future research directions. *International Journal of Systematic Innovation*, 7(7). https://doi.org/10.6977/IJoSI.202309_7(7).0002

Patston, T.J., Cropley, D.H., Marrone, R.L., & Kaufman, J.C. (2018). Teacher implicit beliefs of creativity: Is there an arts bias? *Teaching and Teacher Education*, 75, 366–374. https://doi.org/10.1016/j.tate.2018.08.001

Kale, E., Aknar, A., & Başar, O. (2019). Absorptive capacity and firm performance: The mediating role of strategic agility. *International Journal Hospitality Management*, 78, 276–283. https://doi.org/10.1016/j.ijhm.2018.09.010

Kasravi, K. (2010). *Applications of TRIZ to IT: Cases and Lessons Learned*. TRIZCON The Altshuller Institute for TRIZ Studies, Dayton, OH, USA.

Khan, M.S. (2022). Estimating a panel MSK dataset for comparative analyses of national absorptive capacity systems, economic growth, and development in low and middle income countries. *PLoS One*, 17(10), e0274402. https://doi.org/10.1371/journal.pone.0274402

Kim, M., & Park, M.J. (2023). Absorptive capacity in entrepreneurial education: Rethinking the Kolb's experiential learning theory. *The International Journal of Management Education*, 21(3), 100873. https://doi.org/10.1016/j.ijme.2023.100873

Knudsen, M.P., Dalum, B., & Villumsen, G. (2001). Two faces of absorptive capacity creation: Access and utilisation of knowledge. In: *Proceedings of the Nelson and Winter Conference. Danish Research Unit for Industrial Dynamics*, Aalborg, Denmark, p18.

Kobarg, S., Stumpf-Wollersheim, J., & Welpe, I.M. (2018). University-industry collaborations and product innovation performance: The moderating effects of absorptive capacity and innovation competencies. *Journal Technology Transfer*, 43(6), 1696–1724. https://doi.org/10.1007/s10961-017-9583-y

Krejcie, R.V., & Morgan, D.W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30(3), 607–610.

https://doi.org/10.1177/001316447003000308

Labuda, I. (2015). Possibilities of applying TRIZ methodology elements (the 40 inventive principles) in the process of architectural design. *Procedia Engineering*, 131, 476–499. https://doi.org/10.1016/j.proeng.2015.12.443

Li, X., Ma, L., Ruman, A.M., Iqbal, N., & Strielkowski, W. (2023). Impact of natural resource mining on sustainable economic development: The role of education and green innovation in China. *Geoscience Frontiers*, 15(3), 101703. https://doi.org/10.1016/j.gsf.2023.101703

Lin, Y.S., & Chen, M. (2021). Implementing TRIZ with supply chain management in new product development for small and medium enterprises. *Processes*, 9(4), 614. https://doi.org/10.3390/pr9040614

Mallana, M.F.B.A., Iqbal, A.M., Iqbal, S., Khan, A.S., & Senin, A.A. (2013). The critical factors for the successful transformation of technology from developed to developing countries. *Journal Teknology*, 64(3), 105–108. https://doi.org/10.11113/jt.v64.2278

Mann, D., & Spain, J. (2000). Breakthrough thinking with TRIZ for business and management: An overview. In: *Proceedings of TRIZCON2000*.

Miguélez, E., & Moreno, R. (2015). Knowledge flows and the absorptive capacity of regions. *Research Policy*, 44(4), 833–848. https://doi.org/10.1016/j.respol.2015.01.016

Muijs, D. (2010). *Doing Quantitative Research in Education with SPSS*. 2nd ed. Sage Publications, London, Thousand Oaks, CA, New Delhi.

Müller, J.M., Buliga, O., & Voigt, K.I. (2021). The role of absorptive capacity and innovation strategy in the design of Industry 4.0 business models - a comparison between SMEs and large enterprises. *European Management Journal*, 39, 333–343. https://doi.org/10.1016/j.emj.2020.01.002

Murovec, N., & Prodan, I. (2009). Absorptive capacity, its determinants, and influence on innovation output: Cross-cultural validation of the structural model. *Technovation*, 29(12), 859–872. https://doi.org/10.1016/j.technovation.2009.05.010

Naqshbandi, M.M., & Tabche, I. (2018). The interplay of leadership, absorptive capacity, and organizational learning culture in open innovation: Testing a moderated mediation model. *Technological Forecasting and Social Change*, 133, 156–167. https://doi.org/10.1016/j.techfore.2018.03.017

Navas, H.V.G., Tenera, A.M.B.R., & Machado, V.A.C. (2015). Integrating TRIZ in project management processes: An ARIZ contribution. *Procedia Engineering*, 131, 224–231. https://doi.org/10.1016/j.proeng.2015.12.381

Ozgun, A.H., Tarim, M., Delen, D., & Zaim, S. (2022). Social capital and organizational performance: The mediating role of innovation activities and intellectual capital. *Healthcare Analytics*, 2, 100046. https://doi.org/10.1016/j.health.2022.100046

Rehman, W.U., Rehman, C.A., & Sahid, A. (2011). Intellectual capital performance and its impact on corporate performance: An empirical evidence from Modaraba sector of Pakistan. *Australian Journal Business and Management Research*, 1(5), 8–16.

Ruchti, R., Mann, D., & Spain, J. (2001). Using TRIZ to overcome business contradictions: Profitable ecommerce. In: *Proceedings of TRIZCON2001*.

Russo, D., & Spreafico, C. (2015). TRIZ 40 inventive principles classification through FBS ontology. *Procedia Engineering*, 131, 737–746. https://doi.org/10.1016/j.proeng.2015.12.367

Spreafico, C. (2022). Can TRIZ (theory of inventive problem solving) strategies improve material substitution in eco-design? *Sustainable Production and Consumption*, 30, 889–915. https://doi.org/10.1016/j.spc.2022.01.010

Terninko, J., Zusman, A., & Zlotin, B. (1998). *Systematic Innovation: An Introduction to TRIZ (Theory of Inventive Problem Solving)*. CRC Press, United States.

Terstriep, J., & Lüthje, C. (2018). Innovation, knowledge and relations - on the role of clusters for firms' innovativeness. *European Planning Studies*, 26(11), 2167–2199. https://doi.org/10.1080/09654313.2018.1530152

Topkaya, Ö. (2015). Emigration of innovative workforce in the light of patent data. *Procedia Social Behavioral Science*, 195, 42–51. https://doi.org/10.1016/j.sbspro.2015.06.170

Truong, B.T.T., & Nguyen, P.V. (2024). Driving business performance through intellectual capital, absorptive capacity, and innovation: The mediating influence of environmental compliance and innovation. *Asia Pacific Management Review*, 29, 64–75. https://doi.org/10.1016/j.apmrv.2023.06.004

Valjak, F., & Bojcetic, N. (2022). Functional modelling through function class method: A case from DfAM domain. *Alexandria Engineering Journal*, 66. https://doi.org/10.1016/j.aej.2022.12.001

Walter, L. (2005). TRIZ-tools in innovation management: The use of the inventive principles and separation principles for resolving contradictions. *Industrie Management*, 21(3), 13–16.

Wang, M., Zhang, D., & Zhang, L. (2015). Introduction of TRIZ theory for conflict-solving in the building energy and environment management system innovation. *Procedia Engineering*, 121, 2232–2239. https://doi.org/10.1016/j.proeng.2015.09.201

Ye, W., & Wang, Y. (2019). Exploring the triple helix synergy in Chinese national system of innovation. *Sustainability*, 11(23), 6678. https://doi.org/10.3390/su11236678

Yu, H., Zhang, J., Zhang, M., & Fan, F. (2022). Cross-national knowledge transfer, absorptive capacity, and total factor productivity: The intermediary effect test of international technology spillover. *Technology Analysis Strategic Management*, 34(6), 625–640. https://doi.org/10.1080/09537325.2021.1915476

Zahra, S.A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. *Academy Management Review*, 27(2), 185–203. https://doi.org/10.2307/4134351

AUTHOR BIOGRAPHIES

Dr. Abeda Muhammad Iqbal earned her Bachelor's degree in Commerce from the University of Karachi in 2005, followed by a Master's and Ph.D. in Human Resource Development from Universiti Teknologi Malaysia in 2012 and 2018, respectively. She previously served as a Research Fellow and later as a Postdoctoral Research Fellow at the Institute of Social Informatics and Technological Innovation (ISITI), Universiti Malaysia Sarawak. At present, she is a Senior Lecturer in the Department of Business Administration, Faculty of Economics, Business and Accounting (FEBA), i-CATS University College, Kuching, Sarawak. Her research interests include technology management, smart contracts, blockchain technology, organizational behavior, and entrepreneurship, and she has contributed to several reputable publications in these areas.

Professor Dr. Narayanan Kulathuramaiyer is a pioneer academic serving as a Professor at the Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak (UNIMAS) for the past 24 years. He earned his Ph.D. in Computer Science from Graz University of Technology, Austria. He has published several peer-reviewed high-impact articles in reputable venues, such as technology-enhanced learning, data analytics, and e-learning in the digital age. Prof. Narayanan has led numerous initiatives in digital education, including UNIMAS' Virtual Campus Programme, its first MOOC initiative, and the creation of the eLearning blueprint in 2001. Nationally, he contributed to the NITC Strategic Taskforce on eLearning (2000–2001). His work impacts marginalized communities in Malaysia, Indonesia, and the Philippines through innovative MOOC technology. He holds multiple registered Intellectual Properties and has received national and international award for projects, such as eBario, e-Toro, and Semantic Clustering Toolkit.

Adnan Shahid Khan (Senior Member, IEEE) received the B.Sc. degree (Hons.) in computer science from the University of the Punjab, Lahore, Pakistan, in 2005, and his Master's, Ph.D., and Postdoctoral degrees in networks and information security from the Universiti Technology Malaysia, Johor Bahru, Malaysia, in 2008, 2012, and 2013, respectively. He is currently an Associate Professor at the Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak (UNIMAS). His research interests include cybersecurity in wireless communication, cloud computing, the Internet of Things, software-defined networking, cryptography, networks, and information security.

Dr. Johari Abdullah is currently serving as an Associate Professor at the Faculty of Computer Science & IT, UNIMAS, Sarawak. He received his PhD (Computing Science) from Newcastle University (UK), his Master of IT from Queensland University of Technology, Brisbane, Australia, and a Bachelor of Computer Science (Networking) degree from Universiti Putra Malaysia. His interest in ICT includes deep learning, generative AI, cybersecurity, trusted systems, blockchain technology, web system design and development, system architecture, problem solving using tools, such as TRIZ, ICT education for children and youth through computational thinking, scratch, and computer science unplugged, and also open-source systems and software.

DOI: 10.6977/IJoSI.202508_9(4).0003
L. Muflikhah, A.G. Nurfansepta, etc./Int. J. Systematic Innovation, 9(4), 32-41 (2025)

Detection of lung cancer mutation based on clinical and morphological features using adaptive boosting method

Lailil Muflikhah*, Amira G. Nurfansepta, Edy Santoso, Agus Wahyu Widodo

Department of Informatics Engineering, Faculty of Computer Science, Brawijaya University, Malang, East Java, Indonesia

*Corresponding author E-mail: lailil@ub.ac.id

(Received 20 November 2024; Final version received 19 April 2025; Accepted 17 June 2025)

Abstract

Lung cancer is a leading cause of cancer-related mortality worldwide, and accurate detection of epidermal growth factor receptor mutations is essential for personalized treatment. However, non-invasive identification of these mutations remains challenging due to the complexity of clinical and morphological patterns. This study develops an adaptive boosting (AdaBoost)-based machine learning model for detecting lung cancer mutations using clinical and morphological data. The dataset consists of clinical and morphological attributes from 80 patients, which processed through comprehensive preprocessing steps, including imputation, outlier removal, and feature selection. One-hot encoding increased the feature count beyond the original 28, and analysis of variance was employed to retain the most relevant 33 features. AdaBoost was trained with optimized hyperparameters, including learning rate and the number of estimators, which were tuned using grid search to ensure robustness. The model's performance was evaluated using an 80/20 train-test split and k-fold cross-validation to assess generalization capability. Experimental results demonstrated that AdaBoost outperformed other models, achieving an accuracy of 83% and an area under the curve of 0.90 after feature selection. The model maintained superior cross-validation scores compared to Naive Bayes, decision tree, K-nearest neighbors, and support vector machine, reinforcing its reliability in mutation detection. The study highlights the significance of preprocessing steps in improving classification performance and suggests that AdaBoost can serve as an effective, non-invasive tool for assisting clinical decision-making in lung cancer mutation detection.

Keywords: Adaptive Boosting, Analysis of Variance, Lung Cancer, Machine Learning, Mutation

1. Introduction

Lung cancer is a leading cause of cancer-related mortality, with early and accurate detection of genetic mutations playing a crucial role in optimizing treatment strategies (Rakesh & Baskar, 2024). Among these mutations, epidermal growth factor receptor (EGFR) mutations are particularly significant for targeted therapies (Wang et al., 2019). Traditional detection methods rely on invasive procedures such as tissue biopsies, which pose risks to patients and may not always be feasible (Kanan et al., 2024).

Advancements in machine learning (ML) offer promising non-invasive alternatives for mutation detection using clinical and morphological data (Yu et al., 2019). Various ML models, including support vector machines (SVM), decision trees, and K-nearest neighbors (KNN), have been applied in cancer diagnosis, but their performance often depends heavily on extensive feature engineering and preprocessing (Jain et al., 2024). Boosting algorithms, particularly adaptive boosting (AdaBoost), have demonstrated superior performance by combining multiple weak learners into a robust predictive model (Bushara et al., 2023).

This study leverages the AdaBoost algorithm to enhance the accuracy and sensitivity of EGFR mutation detection based on clinical and morphological features. Through rigorous preprocessing, including outlier removal, feature encoding, and analysis of variance (ANOVA)-based feature selection, we optimize the dataset for improved classification performance. The effectiveness of AdaBoost is evaluated in comparison to other ML models, emphasizing its potential as a reliable, non-invasive alternative for assisting clinical decision-making in lung cancer mutation detection.

2. Related Work

The application of ML in lung cancer detection has been extensively explored, with various studies highlighting the effectiveness of traditional ML models in identifying cancerous mutations. Previous research has applied ML approaches such as SVM, decision trees, KNN, and ensemble techniques such as Random Forest in cancer classification (Maurya et al., 2024). These methods have demonstrated success in distinguishing cancer subtypes and predicting patient outcomes, but often require extensive feature engineering to enhance model performance (Li, 2023).

Boosting methods, particularly AdaBoost, have been increasingly recognized for their ability to enhance classification accuracy in medical applications by combining multiple weak learners into a strong predictive model (Gautam et al., 2024). Unlike deep learning techniques, which require large datasets and substantial computational resources, AdaBoost provides a more interpretable and computationally efficient approach, making it suitable for clinical applications with limited data availability (Jain et al., 2024).

Several studies have applied ML techniques in lung cancer detection using clinical and morphological data. For instance, Kwon et al. (2023) demonstrated that integrating multiple blood markers and clinico-pathological features significantly improved classification performance (Kwon et al., 2023). Similarly, Wang et al. (2019) investigated the use of imaging and clinical attributes for mutation detection, showing that feature selection played a crucial role in model optimization (Wang et al., 2019). However, most prior research has focused on deep learning-based models, such as convolutional neural networks (CNNs), for lung cancer diagnosis (Le et al., 2021). While CNNs have demonstrated high accuracy in radiomics-based studies, their black-box nature and high computational requirements limit their practical use in clinical settings (Kanan et al., 2024, p. 20).

In contrast, this study focuses on leveraging AdaBoost to detect EGFR mutations based on clinical and morphological features. The rationale for using AdaBoost over deep learning approaches lies in its ability to handle smaller datasets while maintaining high classification performance. In addition, AdaBoost enables easier interpretation of feature importance, which is crucial in clinical decision-making (Sachdeva et al., 2024). By incorporating feature selection techniques such as ANOVA, this study aims to further enhance model robustness and accuracy, addressing gaps in existing literature that often overlook the impact of feature selection on ML performance.

3. Research Methods

This research was designed as a predictive study aimed at detecting lung cancer mutations, specifically EGFR mutations, using ML on clinical and morphological data. The study involved several phases, as shown in Fig. 1, including data acquisition, preprocessing, feature selection, model development, and performance evaluation. By employing the AdaBoost algorithm, this study sought to improve mutation detection accuracy, leveraging its capacity to enhance predictive accuracy through boosting weak learners into a stronger predictive model.

This study employs the AdaBoost algorithm, an ensemble learning method that iteratively combines weak classifiers to create a more robust predictive model. As illustrated in Fig. 1, AdaBoost operates by assigning weights to training instances and adjusting them iteratively based on model performance. A properly referenced workflow diagram depicting this process is included to enhance comprehension.

3.1. Data Sets

The data set used in this study consists of clinical and morphological data collected from 80 lung cancer patients, initially comprising 28 features. To ensure data quality, several preprocessing steps were performed. Missing values in numerical features were addressed using KNN imputation, whereas categorical features underwent mode imputation to maintain data completeness. Outlier detection and removal were conducted using the interquartile range (IQR) method, minimizing the impact of extreme values that could potentially bias the model. In addition, categorical variables such as lobe location and emphysema type were transformed using one-hot encoding, which increased the total number of features beyond the initial 28. Feature selection was then performed using the ANOVA method, refining the feature set to 33 relevant predictors.

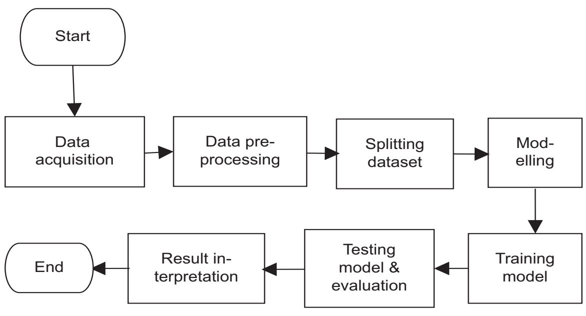


Fig. 1. General proposed method

The boxplots in Fig. 2 depict the distribution of various features related to diabetic nephropathy, such as age, calcification, tumor location, and metastasis. “Age” is well-distributed with a median in the mid-60s, whereas features such as “dimension” and “density” show minimal variability, suggesting limited diagnostic relevance. “Calcification” and “tumor location” exhibit greater variability, indicating potential significance in disease progression. Metastasis-related features, such as “liver” and “bone metastasis,” show rare occurrences with occasional outliers. Overall, the plots highlight patterns and anomalies that may aid in understanding the variability of clinical features associated with diabetic nephropathy.

3.2. Preprocessing

Data preprocessing is a crucial step in the data analysis workflow, aimed at preparing raw data into a cleaner and more usable form for analysis models or ML. This process includes various techniques, such as data cleaning to remove missing values, handling outliers, data transformation – which may involve converting data types or encoding categorical variables into numeric formats – and feature selection. The main goal of preprocessing is to enhance data quality so that the model used can produce more accurate results, while also reducing the risk of bias and errors in further analysis (Benhar et al., 2020).

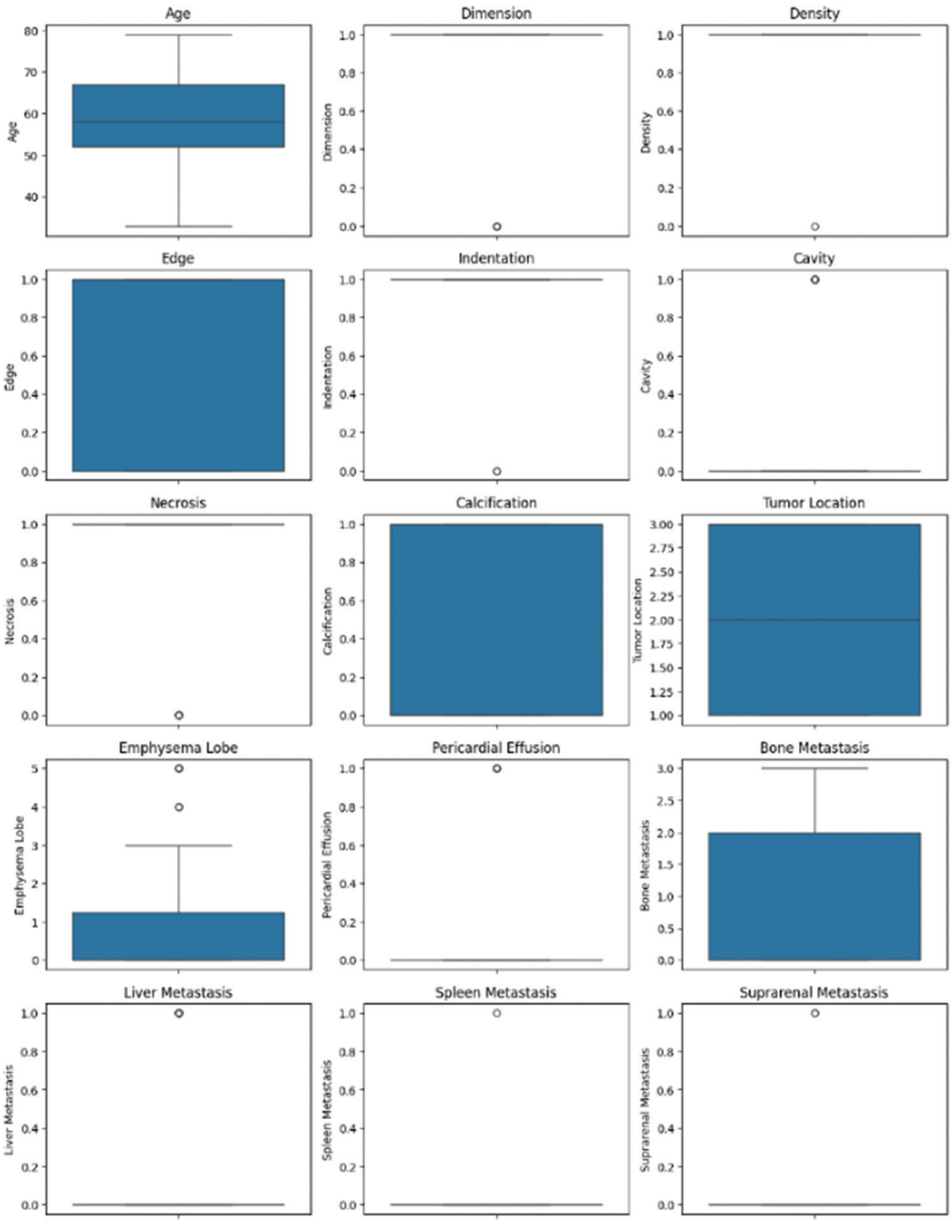


Fig. 2. Boxplot of the distribution of several features in data sets

3.2.1. Data Cleansing

The data were first checked for duplicates, and no duplicate entries were found. Next, an assessment of missing values revealed that several features contained missing data, requiring either removal or imputation to ensure data quality. To streamline the dataset for modeling, the features "no" and "test number" were removed, as they were not relevant to the modeling process. In addition, patient records with missing values across nearly all features were excluded, resulting in the removal of four patient records due to the high proportion of missing data.

3.2.2. Outlier Detection

Outlier detection using the IQR method, also known as the Tukey method, is a highly effective technique for identifying and removing extreme values from a dataset. This method is based on the IQR measurement, which is the range between the first quartile (*Q1*) and the third quartile (*Q3*), encompassing the central portion of the data distribution (Berger & Kiefer, 2021). This process helps produce more accurate models and analysis results, especially in situations where data distributions are non-normal or highly variable. The IQR formula, along with the Tukey method for outlier identification, is shown in Eq. (1), Eq. (2), and Eq. (3).

$$IQR = Q_3 - Q_1 \tag{1}$$

$$Outlier > Q_3 + 1.5 \times IQR \tag{2}$$

$$Outlier > Q_1 + 1.5 \times IQR \tag{3}$$

3.2.3. One-Hot Encoding

Performing one-hot encoding for several features, such as lobe location, emphysema type, emphysema location, lymphadenopathy, pulmonary nodule, and pleural effusion, can yield new features that are more relevant and have a higher correlation with EGFR mutation. Results are improved when one-hot encoding is applied. To handle categorical variables, we applied one-hot encoding, which converts each categorical feature into multiple binary variables, ensuring compatibility with ML models. In this study, categorical features such as lobe location and emphysema type were transformed using this method.

3.2.4. Feature Selection

Feature selection plays a crucial role in improving model performance by eliminating irrelevant or redundant features. In this study, categorical variables such as lobe location and emphysema type were transformed using one-hot encoding, which expanded the feature space. To reduce dimensionality and retain only the most relevant predictors, the ANOVA method was applied. ANOVA evaluates the statistical significance of each feature in relation to the target variable, ensuring that only features with strong discriminative power are selected. As a result, the feature set was reduced to 33 features, which were subsequently used for model training and evaluation.

Feature selection using the ANOVA method is a technique that identifies features that have a significant impact on the target variable in a dataset. ANOVA is applied to compare the means of groups generated by different features to determine if these differences are substantial enough to influence the target variable (Nasiri & Alavi, 2022). Features showing significant differences are considered important and are retained in the model, whereas non-significant features may be removed to simplify the model and reduce the risk of overfitting. This method is particularly useful in regression or classification analysis, where selecting the right features can significantly enhance model accuracy and computational efficiency. The ANOVA formula involves the total sum of squares (*SST*) and sum of squares between (*SSB*) as shown in Eq. (4) and Eq. (5).

$$SST = \sum_{i=1}^{N} \left(X_i - \bar{X} \right)^2 \tag{4}$$

Remark:
X_i: i^{th} data point
$\bar{X}$: Means of all data

N: Total number of observations for all groups

$$SSB = \sum_{j=1}^{k} n_j \left(\bar{X}_j - \bar{X} \right)^2 \tag{5}$$

Remark:
n_j: Number of observations (data points) in group j
k: Number of groups
$n_1, n_2, \ldots n_k$: Sample size of each group

Furthermore, the *F*-statistics in ANOVA is a measure used to compare variances, and it is calculated based on the SSB and the Within-Group Sum of Squares (SSW) as shown in Eq. (8). Specifically, the *F*-value is obtained by dividing the mean square between groups (MSB) by the mean square within groups (MSW) as shown in Eq. (6) and Eq. (7).

$$MSB = \frac{SSB}{k-1} \tag{6}$$

$$MSW = \frac{SSW}{N-k} \tag{7}$$

$$F = \frac{MSB}{MSW} \tag{8}$$

3.3. Adaptive Boosting Algorithm

ML methods typically assume that data are well-distributed, though in practice, this ideal condition is rarely met. In real-world classification tasks, class imbalance is a common challenge, where certain classes have significantly fewer samples than others. This imbalance can negatively impact model performance, as standard classification methods may be biased toward the majority class, leading to poor generalization on minority classes. Unlike traditional ML, which builds a single model from a dataset that ensembles learning methods and combines multiple models to enhance predictive performance (Zhou, 2012).

Ensemble methods aim to reduce model errors and improve accuracy by leveraging the strengths of multiple classifiers. Several techniques are used in ensemble learning: Stacking integrates outputs from different models, where a meta-model predicts the final outcomes based on the outputs of base models, while bagging (e.g., random forest) improves stability by training models on bootstrapped subsets of data. Boosting is a powerful ensemble learning technique designed to improve prediction accuracy by sequentially combining multiple weak learners into a single strong learner (Rincy & Gupta, 2020).

The boosting process works iteratively, where each new model is trained to focus on the mistakes made by its predecessors. Specifically, misclassified instances are assigned to higher weights, making them more influential in training subsequent models. This iterative process continues, with each new weak learner refining the overall prediction by correcting previous errors (González et al., 2020). Finally, the predictions of all models are aggregated – using weighted voting for classification or weighted summation for regression – to generate the final output. A workflow diagram of the boosting process is shown in Fig. 3, illustrating how multiple models contribute to building a more robust predictor.

This study leverages the AdaBoost algorithm, a robust ensemble method, particularly effective in handling class imbalance and enhancing weak classifiers. AdaBoost assigns higher weights to misclassified instances, ensuring that hard-to-classify cases receive more focus in subsequent iterations. The AdaBoost algorithm follows these key steps, and a detailed illustration of the AdaBoost process is provided in Fig. 3, depicting how misclassified samples influence model training at each iteration.

1. Initializing sample weights: Initially, all training samples x_i are assigned equal weights ω_i, ensuring a uniform distribution across the dataset. This step is mathematically represented in Eq. (9).

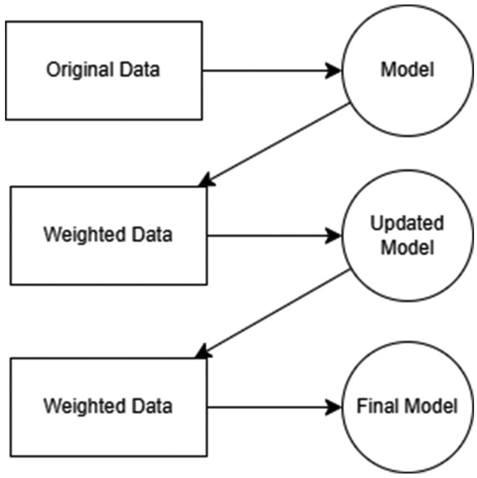


Fig. 3. The stages of the AdaBoost process

$$\omega_i = \frac{1}{N} \tag{9}$$

2. Training weak learners: At each boosting iteration t, a weak classifier $h_t(x)$ is trained using the weighted dataset. The classification error $\in_t$, which quantifies the misclassification rate of the weak learner, is computed as defined in Eq. (10).

$$\in_t = \sum_{i=1}^{N} \omega_i . I\left(h_t\left(x_i\right) \neq y_i\right) \tag{10}$$

3. Updating classifier weight: The importance of each weak classifier is determined based on its classification error. A higher weight is assigned to more accurate classifiers, as shown in Eq. (11), where α_t is calculated as a function of $\in_t$.

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\in_t}{\in_t}\right) \tag{11}$$

2. Updating sample weights: The weights of misclassified samples are increased to ensure they receive more attention in subsequent iterations. The new sample weight distribution is determined using Eq. (12), ensuring that harder-to-classify instances influence future classifiers more significantly.

$$\omega_i \leftarrow \omega_i . exp\left(-\alpha . y_i h_i\left(x_i\right)\right) \tag{12}$$

3. Normalization of weights: To maintain a valid probability distribution, the sample weights are normalized, as represented in Eq. (13).

$$\omega_i \leftarrow \frac{\omega_i}{\sum_{j=1}^{N} \omega_i} \tag{13}$$

4. Final model (strong learner): The final strong classifier H_x is obtained by combining all weak classifiers, weighed according to their performance, as stated in Eq. (14).

$$H_x = sign\left(\sum_{t=1}^{T}\alpha_t . h_t(x)\right) \quad (14)$$

4. Results and Discussion

4.1. Experimental Result

To evaluate model performance, five ML models, KNN, Naive Bayes, SVM, decision tree, and AdaBoost, were tested under different preprocessing scenarios. The dataset was initially split into 80% training and 20% testing, ensuring that the models were trained on a substantial portion of the data while preserving a separate test set for final evaluation.

To optimize hyperparameters and assess model generalization, k-fold cross-validation (with k = 5 or 10, as specified per experiment) was applied exclusively to the training set. This procedure ensured that model selection was based on performance across multiple validation splits, preventing overfitting to a single subset. After cross-validation, the best-performing hyperparameters were used to train the final model, which was then evaluated on the unseen test set to provide an independent measure of performance. Hyperparameter tuning was conducted using a grid search approach, systematically exploring multiple parameter values for each model. For KNN, the number of neighbors was tested with values (3, 5, 7, and 9). Naive Bayes was implemented using the Gaussian Naive Bayes approach, assuming a normal distribution for continuous features. For SVM, the radial basis function kernel was applied, with the penalty parameter tested over the range (0.1, 1, 10, and 100). Decision tree models were optimized by varying the maximum depth between (5, 10, 15, and 20), and the minimum number of samples per leaf was tested at (1, 5, and 10). For AdaBoost, the number of estimators was set to (50, 100, and 200), whereas the learning rate was tuned within the range (0.01, 0.1, and 1).

The final hyperparameter configurations were determined based on the highest cross-validation accuracy. Once optimized, the models were evaluated on the test set, and their performance was measured using key classification metrics: accuracy, precision, recall, F1 score, and AUC-receiver operating characteristics (ROC). The results, presented in Tables 1-3, highlight the impact of different preprocessing strategies on model performance. These findings demonstrate that AdaBoost consistently outperformed other models across various preprocessing scenarios, specifically after applying ANOVA-based feature selection.

The results in Table 1 highlight the performance of various models for lung cancer mutation detection without feature selection or outlier removal. KNN performs best with an accuracy of 68.8% and a high AUC-ROC of 0.746. It shows stable performance even without tuning. Naive Bayes improves significantly after tuning, matching KNN's performance and achieving high precision (81.8%). SVM and AdaBoost initially performed poorly but improved with tuning, with SVM reaching 68.8% accuracy and AdaBoost improving its F1-score to 0.613. The decision tree shows moderate performance with minimal gains from tuning. Overall, the results indicate that preprocessing steps such as feature selection and outlier removal are crucial for improving model performance.

The results in Table 2 demonstrate the significant improvement in performance for lung cancer mutation detection when outliers are removed, even without feature selection. AdaBoost achieves perfect scores across all metrics, indicating exceptional model performance with complete alignment between

Table 1. Comparison of performance results (without feature selection and outlier removal)

Model	Accuracy	F1-score	Precision	Recall	AUC
KNN	0.688	0.689	0.695	0.688	0.746
KNN (tuning)	0.688	0.689	0.695	0.688	0.698
Naive bayes	0.625	0.588	0.798	0.625	0.698
Naive Bayes (Tuning)	0.688	0.689	0.818	0.688	0.695
SVM	0.563	0.405	0.316	0.563	0.240
SVM (tuning)	0.688	0.684	0.731	0.688	0.238
Decision tree	0.563	0.557	0.556	0.563	0.546
Decision tree (tuning)	0.563	0.564	0.570	0.563	0.516
AdaBoost	0.500	0.450	0.577	0.500	0.399
AdaBoost (tuning)	0.625	0.613	0.689	0.625	0.508

Abbreviations: AdaBoost: Adaptive boosting; AUC: Area under the curve; KNN: K-nearest neighbors; SVM: Support vector machine.

precision, recall, and AUC. KNN shows remarkable improvement with tuning, reaching 87.5% accuracy and strong F1, precision, and recall scores. Naive Bayes also performs consistently well, achieving high accuracy (87.5%) and an excellent AUC of 0.933, both with and without tuning. The decision tree delivers strong performance, with tuned and untuned versions achieving 87.5% accuracy and a high AUC of 0.900. SVM exhibits moderate results, maintaining consistent scores before and after tuning. Overall, the removal of outliers significantly enhances the models' robustness and effectiveness, particularly boosting tuned algorithms such as AdaBoost and KNN.

The results in Table 3 demonstrate the effect of combining feature selection and outlier removal on lung cancer mutation detection. AdaBoost and its tuned version maintain perfect scores across all metrics, achieving 100% accuracy, precision, recall, F1-score, and AUC, showcasing its effectiveness in handling the refined dataset. KNN and Naive Bayes perform consistently well, with tuning enhancing their performance to 87.5% accuracy and achieving an AUC of 0.933. SVM delivers moderate results, maintaining a balanced performance with 75% across all metrics, showing that feature selection and outlier removal have a limited impact on this algorithm. The decision tree shows significant improvement with tuning, increasing accuracy and recall to 87.5% and achieving an AUC of 0.906. Overall, combining feature selection with outlier removal enhances model robustness, especially for tuned models such as AdaBoost, KNN, and decision trees, leading to better classification performance and improved detection capability.

4.2. Discussion

The results of this study demonstrate the effectiveness of AdaBoost in detecting EGFR mutations using clinical and morphological features. Compared to other models, AdaBoost consistently achieved the highest classification performance across different preprocessing scenarios. These findings align with previous research by Kwon et al. (2023), who reported improved cancer classification by integrating

Table 2. Comparison of performance results (without feature selection but with outlier removal)

Model	Accuracy	F1-score	Precision	Recall	AUC
KNN	0.625	0.631	0.656	0.625	0.500
KNN (tuning)	0.875	0.868	0.896	0.875	0.500
Naive Bayes	0.875	0.868	0.896	0.875	0.933
Naive Bayes (tuning)	0.875	0.868	0.896	0.875	0.933
SVM	0.750	0.750	0.750	0.750	0.667
SVM (tuning)	0.750	0.750	0.750	0.750	0.667
Decision tree	0.875	0.877	0.906	0.875	0.900
Decision tree (tuning)	0.875	0.877	0.906	0.875	0.900
AdaBoost	1.000	1.000	1.000	1.000	1.000
AdaBoost (tuning)	1.000	1.000	1.000	1.000	1.000

Abbreviations: AdaBoost: Adaptive boosting; AUC: Area under the curve; KNN: K-nearest neighbors; SVM: Support vector machine.

Table 3. Comparison of performance results (with both feature selection and outlier removal)

Model	Accuracy	F1-score	Precision	Recall	AUC
KNN	0.625	0.631	0.656	0.625	0.667
KNN (tuning)	0.875	0.868	0.896	0.875	0.933
Naive Bayes	0.875	0.868	0.896	0.875	0.933
Naive Bayes (tuning)	0.875	0.868	0.896	0.875	0.933
SVM	0.750	0.750	0.750	0.750	0.750
SVM (tuning)	0.750	0.750	0.750	0.750	0.750
Decision tree	0.750	0.750	0.850	0.750	0.800
Decision tree (tuning)	0.875	0.877	0.906	0.875	0.906
AdaBoost	1.000	1.000	1.000	1.000	1.000
AdaBoost (tuning)	1.000	1.000	1.000	1.000	1.000

Abbreviations: AdaBoost: Adaptive boosting; AUC: Area under the curve; KNN: K-nearest neighbors; SVM: Support vector machine.

multiple clinical biomarkers. Similarly, Wang et al. (2019) emphasized the importance of feature selection, showing that carefully curated features significantly enhance predictive accuracy (Wang et al., 2019).

The role of preprocessing was particularly notable in this study. The removal of outliers and the application of ANOVA-based feature selection resulted in improved model performance, which highlighted the impact of feature selection in medical ML applications (Jain et al., 2024). Moreover, the superiority of ensemble methods in handling complex, heterogeneous data has been previously established by Gautam et al. (2024), supporting the efficacy of AdaBoost in this study (Gautam et al., 2024).

While deep learning approaches such as CNNs have been extensively used in lung cancer detection. Their application is often constrained by high computational demands and the need for large datasets (Le et al., 2021). The findings of this study further validate the viability of traditional ML models, particularly ensemble methods such as AdaBoost, as practical alternatives in scenarios where data availability and interpretability are crucial factors.

Despite these promising results, certain limitations remain. The dataset used in this study was relatively small, which may affect the model's generalizability. Future studies should explore external validation on larger datasets and incorporate additional features, such as genetic and radiomic data, to enhance model robustness. In addition, further comparisons with deep learning models could provide deeper insights into the trade-offs between interpretability and predictive performance.

5. Conclusion

This study demonstrated the effectiveness of the AdaBoost algorithm in detecting EGFR mutations in lung cancer patients using clinical and morphological features. Compared to other ML models such as SVM, decision tree, and KNN, AdaBoost achieved superior classification performance, emphasizing its potential as a non-invasive diagnostic tool. The preprocessing steps, including outlier removal, feature encoding, and ANOVA-based feature selection, played a crucial role in optimizing the dataset and improving model accuracy. Furthermore, hyperparameter tuning using grid search ensured optimal model performance, highlighting the importance of systematic parameter selection in ML-based medical applications.

6. Future Work

Despite these promising findings, this study has some limitations. The dataset was relatively small, consisting of 80 patient records, which may impact the generalizability of the results. Therefore, to address these limitations and build upon the findings of this study, several future research directions are proposed. First, the model should be validated on larger and more diverse datasets to assess its robustness and generalizability. Second, incorporating additional features, such as genetic markers and radiomic data, may enhance classification performance and provide a more comprehensive assessment of mutation status.

Acknowledgment

This research is supported by the Faculty of Computer Science, Brawijaya University, under DIPA Research Grant with contract number: 4559/UN10.Fl 5/PN/2024.

References

Benhar, H., Idri, A., & Fernández-Alemán, J.L. (2020). Data preprocessing for heart disease classification: A systematic literature review. *Computer Methods and Programs in Biomedicine*, 195, 105635. https://doi.org/10.1016/j.cmpb.2020.105635

Berger, A., & Kiefer, M. (2021). Comparison of different response time outlier exclusion methods: A simulation study. *Frontiers in Psychology*, 12, 675558. https://doi.org/10.3389/fpsyg.2021.675558

Bushara, A.R., Vinod Kumar, R.S., & Kumar, S.S. (2023). An ensemble method for the detection and classification of lung cancer using computed tomography images utilizing a capsule network with visual geometry group. *Biomedical Signal Processing and Control*, 85, 104930. https://doi.org/10.1016/j.bspc.2023.104930

Gautam, N., Basu, A., & Sarkar, R. (2024). Lung cancer detection from thoracic CT scans using an ensemble of deep learning models. *Neural Computing and Applications*, 36(5), 2459–2477. https://doi.org/10.1007/s00521-023-09130-7

González, S., García, S., Del Ser, J., Rokach, L., & Herrera, F. (2020). A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. *Information Fusion*, 64, 205–237. https://doi.org/10.1016/j.inffus.2020.07.007

Jain, R., Singh, P., Abdelkader, M., & Boulila, W. (2024). Efficient lung cancer detection using computational intelligence and ensemble learning. *PLOS ONE*, 19(9), e0310882. https://doi.org/10.1371/journal.pone.0310882

Kanan, M., Alharbi, H., Alotaibi, N., Almasuood, L., Aljoaid, S., Alharbi, T., et al. (2024). AI-driven models for diagnosing and predicting outcomes in lung cancer: A systematic review and meta-analysis. *Cancers (Basel)*, 16(3), 674. https://doi.org/10.3390/cancers16030674

Kwon, H.J., Park, U.H., Goh, C.J., Park, D., Lim, Y.G., Lee, I.K., et al. (2023). Enhancing lung cancer classification through integration of liquid biopsy multi-omics data with machine learning techniques. *Cancers (Basel)*, 15(18), 4556. https://doi.org/10.3390/cancers15184556

Le, N.Q.K., Kha, Q.H., Nguyen, V.H., Chen, Y.C., Cheng, S.J., & Chen, C.Y. (2021). Machine learning-based radiomics signatures for EGFR and KRAS mutations prediction in non-small-cell lung cancer. *International Journal of Molecular Sciences*, 22(17), 9254. https://doi.org/10.3390/ijms22179254

Li, X. (2023). Lung cancer risk prediction and feature importance analysis with machine learning algorithm. *Applied and Computational Engineering*, 19, 205–210. https://doi.org/10.54254/2755-2721/19/20231034

Maurya, S.P., Sisodia, P.S., Mishra, R., & Singh, D.P. (2024). Performance of machine learning algorithms for lung cancer prediction: A comparative approach. *Scientific Reports*, 14(1), 18562. https://doi.org/10.1038/s41598-024-58345-8

Rakesh, M., & Baskar, R. (2024). A support vector machine for lung cancer detection with classification and compared with KNN for better accuracy. *AIP Conference Proceedings*, 2853(1), 020067. https://doi.org/10.1063/5.0198176

Rincy, T.N., & Gupta, R. (2020). Ensemble Learning Techniques and its Efficiency in Machine Learning: A Survey. *2nd International Conference on Data, Engineering and Applications (IDEA)*. p1–6. https://doi.org/10.1109/IDEA49133.2020.9170675

Sachdeva, R.K., Bathla, P., Rani, P., Lamba, R., Ghantasala, G.S.P., & Nassar, I.F. (2024). A novel K-nearest neighbor classifier for lung cancer disease diagnosis. *Neural Computing and Applications*. 36, 22403-22416. https://doi.org/10.1007/s00521-024-10235-w

Wang, S., Shi, J., Ye, Z., Dong, D., Yu, D., Zhou, M., et al. (2019). Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *European Respiratory Journal*. 53, 1800986. https://doi.org/10.1183/13993003.00986-2018

Yu, L., Tao, G., Zhu, L., Wang, G., Li, Z., Ye, J., et al. (2019). Prediction of pathologic stage in non-small cell lung cancer using machine learning algorithm based on CT image feature analysis. *BMC Cancer*, 19(1), 464. https://doi.org/10.1186/s12885-019-5646-9

Zhou, Z.H. (2012). *Ensemble Methods: Foundations and Algorithms*. 1st ed. Chapman and Hall/CRC, Boca Raton.

AUTHOR BIOGRAPHIES

Lailil Muflikhah is a professor in the field of machine learning at the Faculty of Computer Science, Brawijaya University. She received a B.Sc. degree in computer science from the Institut Teknologi Sepuluh Nopember (ITS). She holds an M.Sc. degree in Computer Science from Universiti Teknologi Petronas (UTP), Malaysia, and a Ph.D. degree in Bioinformatics from Brawijaya University. Her research interests include soft computing, machine learning, and intelligent systems. She can be contacted at email: lailil@ub.ac.id.

Amira Ghina Nurfansepta is a student of Informatics Engineering at the Faculty of Computer Science, Brawijaya University. She has experience working as a research assistant, and her research interests include machine learning. She can be contacted at amiragn25@student.ub.ac.id

Edy Santoso is a lecturer in Informatics Engineering, Faculty of Computer Science, Brawijaya University. He is a member of the Intelligent Computing research group. He holds a Bachelor's degree in Mathematics with a focus on Computer Science from Brawijaya University, and a Master's degree in Informatics Engineering from the Sepuluh Nopember Institute, Surabaya, Indonesia.

Agus Wahyu Widodo is a senior lecturer and currently serves as the Vice Dean for General Administration and Finance at the Faculty of Computer Science, Universitas Brawijaya (FILKOM UB). He earned his Bachelor's degree in Electrical Engineering from Universitas Brawijaya and a master's degree in Computer Science from Universitas Gadjah Mada. His teaching portfolio includes fundamental and advanced courses in programming, data mining, machine learning, and evolutionary algorithms. His research interests encompass artificial intelligence, machine learning applications, data science, and intelligent systems, as reflected in his extensive publication record in both national and international journals and conferences. He has also contributed to society through various community service programs, particularly in digital literacy, smart village systems, and AI-based education and health-care initiatives. In addition, he has authored academic books and is actively involved in curriculum development and academic leadership.

DOI: 10.6977/IJoSI.202508_9(4).0004
J. Jayanth, H. K. Ravikiran, etc./Int. J. Systematic Innovation, 9(4), 42-53 (2025)

Single-frame super-resolution with deep residual network–generative adversarial networks

J. Jayanth[1], H. K. Ravikiran[2]*, T. Yuvaraju[1], R. Dileep[1]

[1]Department of Electronics and Communication Engineering, GSSS Institute of Engineering and Technology for Women, Visvesvaraya Technological University, Belagavi, Karnataka, India

[2]Department of Electronics and Communication Engineering, Navkis College of Engineering, Visvesvaraya Technological University, Belagavi, Karnataka, India

*Corresponding author E-mail: hkr@navkisce.ac.in

(Received 14 February, 2025; Final version received 24 April, 2025; Accepted 16 May, 2025)

Abstract

Developing and evaluating a deep learning-based method to enhance satellite image resolution has emerged as a promising approach to address challenges posed by motion, imaging blur, and noise without modifying existing optical systems. This study utilized an enhanced super-resolution generative adversarial network (SRGAN) with ResNet-50 as the generator and a modified VGG-19 in the discriminator. The model was trained on remote sensing images from the Linear Imaging Self-Scanning imagery and compared with very deep super resolution, SRGAN, and enhanced SRGAN methods using the structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR) as evaluation metrics. Utilizing an enhanced SRGAN with ResNet-50 and modified VGG-19 significantly improved satellite image resolution. The proposed method consistently outperformed conventional convolutional neural network- and generative adversarial network-based super-resolution techniques. Across three test datasets, the method achieved SSIM scores as high as 0.862 and PSNR scores of 33.256, 32.886, and 34.885, demonstrating its superior ability to preserve image properties and enhance resolution. The incorporation of perceptual loss alongside pixel loss contributed to improved visual quality, making the approach particularly effective in maintaining fine details and naturalistic high-frequency characteristics.

Keywords: Generative Adversarial Network, Linear Imaging Self-Scanning Image, Peak Signal-to-Noise Ratio, ResNet-50, Structural Similarity Index Measure, Super Resolution

1. Introduction

Super-resolution (SR) reconstruction is a technology that generates high-resolution (HR) images from a sequence of low-resolution (LR) images using a particular algorithm, all without altering the settings of the imaging hardware (Adarsh et al., 2020). Ultra-high-definition television, emergency event monitoring, target identification and localization, military precision targeting, battlefield environment surveillance, and medical image diagnosis are just a few of the many applications of image SR reconstruction technology in both civilian and military sectors (Wang et al., 2022).

Spatial resolution, the lowest discernible unit size or dimension in remote sensing images, is a measure of the image's ability to identify ground target features (Han et al., 2023). Finer target identification is made possible by higher spatial resolution, suggesting that remote sensing photographs include a greater amount of information about ground objects. However, attaining higher spatial resolution solely through hardware advancements is difficult due to the downsampling effects of imaging sensors and various factors that cause deterioration in satellite image processing. These difficulties result in high development costs and lengthy hardware iteration cycles.

Nonetheless, by reconstructing HR images from LR but easily accessible images, image SR technology offers an affordable means to acquire HR images (Sui et al., 2023). Conventional SR reconstruction approaches, including interpolation and prior information-based reconstruction, possess limitations. Interpolation techniques, such as bilinear and bicubic interpolation, perform well in real time but suffer from observable edge effects and poor detail recovery performance (Sui et al., 2023; Wang et al., 2023; Zhang et al., 2022). Reconstruction techniques based on prior knowledge often rely on constraints such as maximal a posteriori probability and iterative back-projection; however, these methods are computationally costly, have limited applicability, and exhibit poor generalizability (Bu et al., 2024; Dong et al., 2022; Haut et al., 2018).

As deep learning techniques advance, single-image SR techniques based on them outperform conventional methods in remote sensing applications and demonstrate wide-ranging utility (Khan et al., 2023; Pang et al., 2023). For example, the SR convolutional neural network (SRCNN) model applies non-linear mapping to extract LR image features for reconstruction (Wang et al., 2023). However, SRCNN's limitations include its single-scale applicability, slow training convergence, and shortcomings in handling various image contents. To overcome these challenges, W. This innovative approach significantly enhances the model's adaptability, training speed, and capacity to capture intricate image nuances, addressing the shortcomings of SRCNN effectively.

The advent of VGG networks has led to the widespread use of deeper network models (Shi et al., 2023). Residual networks (ResNet) (Wang et al., 2023), offer a solution to the gradient-vanishing issues in deep networks. A very deep SR (VDSR) model is used by incorporating ResNet to extract high-frequency information residuals and capture more image detail information (Frizza et al., 2022). Further innovations, such as multiscale ResNet (Zhang et al., 2022), cascading ResNet, and enhanced ResNet (Liu et al., 2022), fully utilize low-level features with multiscale residual blocks for feature extraction and fusion.

The residual channel attention network (Tang et al., 2022) is introduced as a mean of integrating attention processes into the SR field, and the generative adversarial networks (GANs) (Wang et al., 2021) are developed through advancements in deep learning. To address issues with excessively smooth reconstructed images, a SR generative adversarial network (SRGAN), which introduced perceptual loss and utilized the GAN framework for adversarial training, was introduced (Zhang et al., 2021). The more realistic textures produced by enhanced SRGAN (ESRGAN) outperformed those of SRGAN (Wang et al., 2024). Several algorithms, including SRGAN (Min et al., 2024), extended regularized adversarial GAN (Lei et al. 2017), and transferred GAN (He et al., 2022; Jiang et al., 2018; Li et al., 2022; Wang et al., 2022; Veganzones et al., 2016), have been developed for use in remote sensing applications. These algorithms combine multi-loss training techniques, semantic segmentation, and attention mechanisms to enhance the reconstruction of remote sensing images. Furthermore, research on resolution enhancement extends to other types of remote sensing images, including multisource image fusion (Dileep et al., 2024; Jayanth et al., 2025) and multispectral imaging (Ravikiran et al., 2024).

Recent advancements in unsupervised hyperspectral image SR have significantly enhanced the reconstruction of HR images without requiring paired training data. Li et al. (2025) proposed an enhanced deep image prior network that leverages network structure as an implicit prior to guide reconstruction. The model-informed multistage unsupervised network integrates degradation models into a multistage architecture for progressive refinement (Li et al., 2024). Meanwhile, the X-shaped interactive autoencoders utilize cross-modality mutual learning between spectral and spatial domains, while the model-guided coarse-to-fine fusion network applies a coarse-to-fine reconstruction strategy guided by physical priors. These approaches contribute to improved spatial-spectral fidelity in real-world applications.

Despite the remarkable achievements of GANs in image reconstruction and style transfer, challenges persist in their training process, such as mode collapse and gradient vanishing. Moreover, the majority of current techniques use pixel-level loss functions, such as mean squared error (MSE), potentially resulting in excessively smooth reconstructed images devoid of high-frequency information. Given the intricacy of scenes and the variety of target attributes in remote sensing photographs, the qualities of actual remote sensing datasets must be considered throughout the reconstruction process.

The generator in our single-frame SRGAN framework is designed using the ResNet-50 architecture, chosen for its strength in training deep networks via residual learning. This helps mitigate issues such as vanishing gradients and supports the learning of complex, high-level features. In this implementation, the generator receives an LR satellite image and a random noise vector as inputs. These are concatenated to form a rich input space that combines deterministic and stochastic components. This fusion enables the network to produce more realistic and varied HR images (Dileep et al., 2025). The core of the generator consists of several residual blocks, each containing convolutional layers, batch normalization, and rectified linear unit (ReLU) activations. The skip connections in these blocks facilitate the retention of low-level features while capturing high-level abstractions, helping to preserve spatial and textural details (Li et al., 2023).

Following feature extraction, the generator uses upsampling layers—typically transposed convolutions—to increase the spatial resolution of the feature maps. The final output is produced through convolutional layers that refine these upsampled features into a coherent HR image. On the other hand, a discriminator is built upon a modified VGG-19 architecture, which is well-regarded for its effectiveness in perceptual feature extraction. The VGG-19 network is pre-trained and adapted to serve as a binary classifier within the GAN framework, distinguishing between real HR images and those generated by the network. Its deep yet straightforward structure makes it ideal for identifying fine-grained visual differences, thereby enhancing the adversarial training process and helping the generator improve the realism of its outputs (Li et al., 2023).

In this work, the input image passes through a series of convolutional layers with ReLU activations, interleaved with max pooling operations to progressively extract higher-level features. The deep convolutional structure helps the discriminator capture fine-grained texture patterns and global structures, both of which are essential for assessing image realism. After convolutional feature extraction, the output is flattened and passed through multiple dense layers with *LeakyReLU* activations, allowing the network to model more complex decision boundaries. The final layer uses a sigmoid function to produce a probability score indicating whether the input image is real or generated.

By combining the perceptual strength of VGG-19 with a GAN setup, the discriminator plays a crucial role in guiding the generator to produce visually convincing HR images. This adversarial training setup encourages the generator to create outputs that are not only structurally accurate but also perceptually indistinguishable from real images.

In conventional image SR tasks, loss functions, such as MSE and mean absolute error, are commonly used to measure the pixel-wise differences between the generated and ground truth images. However, while these losses are mathematically straightforward and encourage the generated image to be numerically close to the target, they often lead to overly smooth outputs that lack high-frequency details and perceptual sharpness, especially in complex images, such as satellite imagery, where texture and structure are critical.

To address this limitation, perceptual loss is introduced. Unlike conventional pixel-level loss functions, perceptual loss evaluates the difference between high-level feature representations of the generated and ground truth images, extracted from a pre-trained deep convolutional network (typically a network like VGG-19). These feature maps capture semantic content, texture, and structural patterns that align more closely with how humans perceive image quality.

By minimizing perceptual loss, the generator is guided to produce images that are not only numerically similar but also visually and structurally closer to real HR images. This results in outputs that retain finer details, sharper edges, and more natural textures.

Furthermore, current SR techniques may not generalize well to new targets and situations, necessitating training methods and model designs that explicitly improve resilience and generalization. To address these challenges, modifications to the GAN were implemented, specifically by adding ResNet-50 in the generator network and VGG-19 in the discriminator model. The modifications are as follows:

- To enhance channel efficiency, the generator network was modified by switching out its basic residual blocks with bottleneck blocks from ResNet-50 and extracting the features that are important for SR
- Improved skip connections make information flow between layers more effective and aid in the reconstruction of HR features. In order to improve feature blending and guarantee reliable classifier performance, an additional fully connected (FC) layer was implemented
- Effective discriminative feature extraction was improved by fine-tuning VGG-19 within the discriminator.

The article's structure is as follows: Section 2 outlines the methodology, detailing the GAN; Section 3 presents the results and discussion along with a comparison with other algorithms using different datasets; finally, Section 4 provides a brief conclusion.

2. Methodology

In SR, GANs have become a dominant approach, providing a strong framework for converting LR images into HR images. The SRGAN architecture typically consists of a generator and a discriminator, both are essential components in the creation and assessment of super-resolved images.

2.1. General SRGAN Architecture

The algorithm for the SRGAN is as follows:

(i) Input: LR image
(ii) Architecture: Series of convolutional layers without multiscale residual blocks, followed by batch normalization and activation functions
(iii) Upsampling: The generator aims to learn the mapping from an LR image to an SR image

(iv) Output: SR image with dimensions rW × rH × C, where *r* is the upsampling factor
(v) The complete generator function is:

G(LR image) = SR image.

When the generator network transitioned to ResNet-50, the bottleneck blocks in this method generator model took the place of the basic residual blocks. The discriminator model evolved concurrently into a condensed version of the VGG-19 network, retaining the first 16 layers and including an FC layer for enhanced functionality. Bottleneck blocks were used in the architecture, which used ResNet-50, to improve performance. Lowering the number of channels and increasing processing performance in deeper networks were made possible in large part by the bottleneck blocks. This design decision was crucial for handling the computing power requirements of deep network architectures.

The generator based on ResNet-50 comprised an input layer, ResNet blocks for feature extraction, upsampling layers to improve spatial resolution, and a final output layer that generates the HR image. Batch normalization, ReLU activation, and convolutional layers were combined to form the ResNet blocks. This study used bottleneck blocks to improve speed while generating SR using ResNet-50. Remarkably, when the stride was set to 1, neither the basic nor the bottleneck blocks performed downsampling. However, when the stride was set to 2, downsampling occurred prior to the addition operation. The main purpose of the convolution kernel was to reduce the number of channels that are handled later in the module. With fewer parameters in the channel, this reduction maximized its use.

A pooling layer was layered following a conventional convolutional layer in the first construction layer. The three remaining bottleneck blocks were then integrated into the second construction tier. Downsampling residual bottleneck blocks was the first step in the third, fourth, and fifth construction layers, followed respectively by three, five, and two residual bottleneck blocks. After processing 16 bottleneck blocks for feature extraction, pixel-shuffle layers were used to improve the resolution twice. Crucially, this optimization happened deliberately at the end of the network layers, helping to maximize resolution while reducing the amount of processing resources used.

The architecture of the generator was carefully designed to learn the complex mapping from LR images to SR images. The generator was made up of a sequence of convolutional layers without multiscale residual blocks. It also included activation functions and batch normalization. Increasing the spatial resolution of LR images was the main goal, and an SR image with dimensions rW × rH × C—where *r* denotes the upsampling factor—is the result. The discriminator functioned to separate the generated SR images from genuine HR images at the same time. Convolutional layers were used in conjunction with batch normalization and activation functions to generate a binary classification that indicates the likelihood that an image is a true HR representation.

A wide range of loss functions is necessary for the SRGAN to perform well. By measuring the generator's realism, the adversarial loss (*advLadv*) creates a competitive environment between the generator and discriminator. Perceptual dissimilarity between produced and HR images was measured by the perceptual loss (*Lperceptual*), ensuring that important features are retained. Furthermore, the pixel-by-pixel variations were quantified by the MSE, which offers a more detailed evaluation of fidelity.

A complex feature extraction technique was integrated into the generator's design using ResNet-50. The LR satellite image was represented by *Ilow* in the input layer of the latent space, while *z* provided a random noise vector. The first input, *x0*, was formed by concatenating *Ilow* and *z*. To extract features, a set of ResNet blocks—each consisting of convolutional layers, batch normalization, and ReLU activation—was combined. The ResNet blocks' skip connections promoted gradient flow, reducing the effects of disappearing gradients. The next step involved an upsampling layer that used transposed convolutions or other methods to improve spatial resolution. The final output layer, which was composed of convolutional layers with a suitable activation function (*tanh*), yielded the HR image *I^high*. This output resulted from the upsampled features, represented as *xupsampled*.

The residual learning paradigm was embodied by the ResNet block, which was defined as *x + Convolution(ReLU(BatchNorm(Convolution(x))))*. The trainable parameters *þG*, representing the weights of the convolutional layers, were encapsulated in the generator function as a whole, represented as *G(Ilow; θG) = I^high*. Essentially, the proposed SRGAN architecture incorporates ResNet-50's ability to smoothly capture complex features, offering a strong foundation for producing realistic and detailed HR satellite images. The design advances the state-of-the-art in SR image synthesis by achieving a delicate balance through the complex interplay of the generator, discriminator, and the stated loss functions.

2.2. Generator Architecture with ResNet-50

The algorithm for the generator architecture is as follows:

(i) Input layer (latent space): *Ilow* is the input LR satellite image, and *z* is a random noise vector.
  - $x_0 = Concatenate(Ilow, z)$

(ii) ResNet blocks:
- Integrate several ResNet blocks for feature extraction.
- Each ResNet block consists of convolutional layers, batch normalization, and ReLU activation.
- The skip connections help in gradient flow.
- $X_i = ResNetBlock(x_i - 1)$, $i = 1, 2, \ldots, n$.

(iii) Upsampling layers:
- Upsample the features to increase spatial resolution.
- Use transposed convolutions or other upsampling techniques.
- *Xupsampled = Upsample(xn)*

(iv) Output layer: Produce the final HR image using convolutional layers with a suitable activation function (e.g., *tanh*).
- *I^high = Convolution(xupsampled)*

(v) Each ResNet block can be defined as follows: *ResNetBlock(x) = x + Convolution(ReLU(Batch Norm(Convolution(x))))*

where *Convolution* refers to a 3 × 3 convolutional layer, *BatchNorm* is batch normalization, and *ReLU* is the ReLU activation function.

(vi) Complete generator function: $G(Ilow; \theta_G) = I$^*high*

Where θ_G represents the weights of the convolutional layers in the generator.

This architecture leverages ResNet-50's ability to capture intricate features, helping the generator produce more realistic and detailed HR satellite images.

2.3. Discriminator with VGG-19

The algorithm for the discriminator with VGG-19 is as follows:

(i) Input: Real HR images (I_{HR}) and generated SR images (I_{SR}).
(ii) Architecture: Convolutional layers followed by batch normalization and activation functions.
(iii) Objective: Discriminate between real and generated images.
(iv) Output: Binary classification indicating the probability of being a real HR image.
(v) Complete discriminator function:

(high)→Binary ClassificationD(Ihigh; θD)→ Binary Classification

With a VGG-19 architecture, the discriminator functions as a discriminating agent in the GAN, especially for SR tasks. This discriminator is equipped with a complex architecture inspired by VGG-19 and incorporates an advanced method to discriminate between generated SR and genuine HR images.

The discriminator performs feature extraction, discriminative analysis, and classification, leveraging the VGG-19 architecture as its base. The discriminator consists of six block structures, each composed of convolutional and FC layers, enabling feature extraction and classification across 19 layers. Most importantly, it uses the first 16 layers to generate feature maps that capture important details from the input images.

After an image is input into the discriminator, it undergoes a transformation process involving discriminative analysis and feature extraction. Initially, the image passes through the first convolution block, which utilizes a 3 × 3 kernel size and a stride. To reduce computational complexity and prevent overfitting, the convolved image is downsampled using max pooling procedures.

This process progresses through subsequent blocks corresponding to the second, third, fourth, and fifth phases of the network, where deeper convolutions hone and enhance the characteristics that have been recovered. When the image reaches the FC layers, it experiences a significant metamorphosis into a vector. The next step involves passing this vector across thick layers, each with 1,024 neural units. Here, the discriminator's capacity to distinguish minute variations between generated and actual images is improved by the addition of non-linearity to its decision-making process using the *LeakyReLU* activation function. As the image propagates through the discriminator, it is subjected to a number of adjustments to extract complex characteristics and identify minute distinctions between generated SR and genuine HR images. The discriminator examines textures, structures, and general appearances using max pooling and convolutional procedures to cultivate a robust discriminative capability.

When classification reaches its final stage, the discriminator thoroughly assesses the converted vector. The last dense layer has an output dimension of two, denoting true or false depending on the classification outcomes. A sigmoid activation function is applied to produce a probability score for each image, classifying the image as either HR or SR. Notably, the sigmoid layer is preceded by an extra-thick FC7 layer that improves the discrimination between SR images and ground truth images, as well as enhances the discriminator's capacity to aggregate features from HR images.

2.4. Discriminator with VGG-19

The algorithm for the discriminator with VGG-19 is as follows:

(i) Input: LR or SR image (*I*).
(ii) Initialization:

- Load pre-trained modified VGG-19 model weights.
- Extract the convolutional and FC layers for feature extraction and classification.

(iii) Feature extraction:
- Pass the input image through the first convolution block:
  (a) *F1 = Conv(I, W1)*
  (b) *A1 = ReLU(F1)*
- Perform max pooling:
  P1 = MaxPool(A1)
- Iterate through subsequent convolution blocks.
- Apply convolutions and activations:
  (a) *Fi = Conv(Pi − 1, Wi)*
  (b) *Ai = ReLU(Fi)*
- Utilize max pooling:
  Pi = MaxPool(Ai)

(iv) Dense layer transformation:
- Flatten the feature maps into a vector:
  V = Flatten(Pn)
- Pass the vector through dense layers:
  (a) *Z1 = V·Wfc1 + bfc1*
  (b) *Afc1 = LeakyReLU(Z1)*
  (c) *Z2 = Afc1·Wfc2 + bfc2*
  (d) *Afc2 = LeakyReLU(Z2)*

(v) Classification: Pass the transformed vector through the final dense layer:
  (a) *Z3 = Afc2·Wfc3 + bfc3*
  (b) *Afc3 = Sigmoid(Z3)*

(vi) Output: Probability score indicating the authenticity of the input image as either LR or SR.

2.5. Loss Function

The two primary components of the loss for single-image SR (*lSSR*) are the regularization loss and the perceptual loss.

(i) Perceptual loss (*lperceptual*): In a perceptual space, perceptual loss quantifies the difference between the generated SR images and the HR ground truth images. It is frequently calculated using a feature similarity metric between the feature representations of the HR and SR images taken from a pre-trained deep neural network, such as the MSE or cosine similarity. The perceptual loss may be mathematically represented as follows:

$$lperceptual = \frac{1}{N}\sum_{i=1}^{N}\left\|\Phi(I_{HR}^{i}) - \Phi\left(G\left(I_{LR}^{i}\right)\right)\right\|_2^2 \quad (1)$$

Where N is the total number of training samples, G stands for the SR generator model, ϕ for the feature extractor, and $I_{HR}^{(i)}$ and $I_{LR}^{(i)}$ for the *i-th* HR and LR input images, respectively.

(ii) Regularization loss (*lregularization*): Overfitting is avoided, and favorable features in the resulting SR images are encouraged by regularization loss. Typical regularization methods include adversarial loss to impose realism, total variation regularization to enhance smoothness, and L1 or L2 regularization on the generator's parameters. The regularization loss can be expressed mathematically as follows:

$$lregularization = \lambda_{reg} = \sum_{\theta}\left\|\theta\right\|_{P} \quad (2)$$

Where θ stands for the generator model's parameters, P is the regularization norm (e.g., P = 1 for L1 regularization and P = 2 for L2 regularization), and *lregularization* is the regularization parameter.

The total loss for the single-image SR is then computed as the weighted sum of the perceptual loss and regularization loss:

lSSR = λperceptual·lperceptual + λregularization ·lregularization

Where λperceptual and λregularization are hyperparameters that control the relative importance of the perceptual and regularization terms, respectively.

3. Results and Discussion

3.1. Dataset

In this work, a new dataset is introduced, consisting exclusively of images captured by the Linear Imaging Self-Scanning (LISS) satellite sensor, as shown in Fig. 1. The LISS sensor provides fine-grained spatial resolutions of 24 m and 5 m, while capturing images across a 140-km orbital sweep. Every 24 days, its operational cycle is repeated. In contrast, the Advanced Wide Field Sensor (AWiFS) offers a slightly poorer spatial resolution of 56 m but spans an even larger orbital sweep of 740 km. AWiFS also features a more frequent revisit cycle of 5 days.

The original LISS images had a pixel resolution of 1956 × 1983. These images were randomly cropped to 86 × 86 to generate training and validation datasets. After that, the cropped images underwent image degradation and were downsampled by a factor of four to produce an LR image set. This LR set was paired with the original image to form LR–HR training pairs. Finally, data augmentation techniques, such as rotation and staggering, were used to augment the dataset.

3.2. Training Process

The proposed architecture was trained in an experimental setup over 40 training epochs with a batch size of 32 and a learning rate of 0.001. Training GAN-type networks presents inherent challenges,

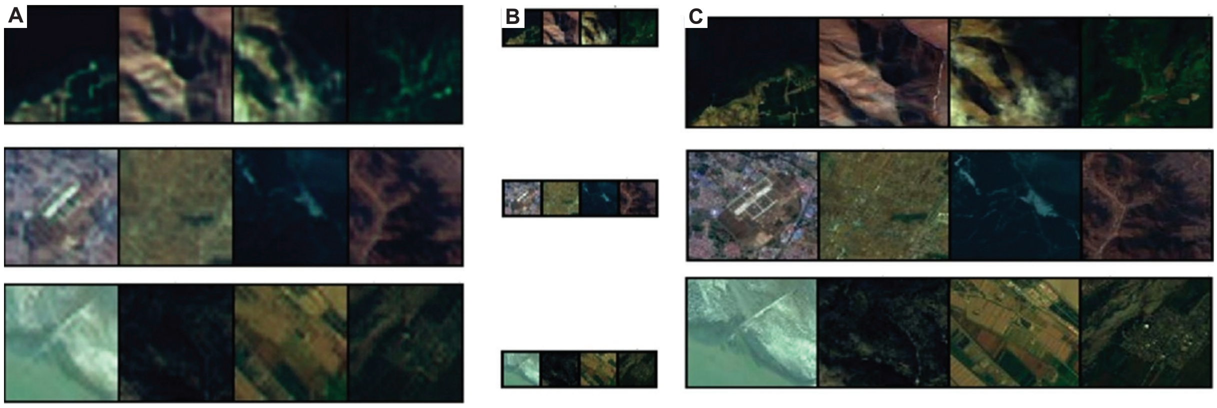


Fig. 1. Original images. (A) Ground truth images. (B) Low-resolution images. (C) High-resolution images

such as instability and mode collapse. To address these issues, a halting standard was introduced. In particular, if a predefined loss threshold of 0.001 was not achieved after 400 epochs, training was terminated. The model loss was monitored after each epoch, and the checkpoint with the lowest loss was preserved. By acting as a benchmark for subsequent epochs, this checkpoint ensured the model maintained optimal performance. The goal of this approach was to reduce instability, prevent premature convergence, and enhance model performance.

Despite some fluctuations, the loss trend exhibited a steadily dropping trajectory, suggesting that the model was improving and converging. Notably, the proposed model showed improved stability and resilience compared to conventional GAN networks, where variations in the loss function could reflect training collapse and intrinsic instability. To produce LR images, each HR image was downscaled by a factor of four. The LR images were 32 × 32 pixels in size, while the HR images were 128 × 128 pixels. Training on smaller image patches enabled the model was able to concentrate on fine-grained local textures, structural components, and object characteristics in remote sensing images. This method facilitated the extraction of crucial information and patterns needed for accurate SR reconstruction. Additionally, using smaller-sized images reduced memory utilization and computational complexity, contributing to a more efficient training process.

3.3. Dataset Splitting Strategy

Three datasets were used, covering varied remote sensing scenarios. Each dataset was split into 70% for training, 15% for validation, and 15% for testing, ensuring the model was evaluated on unseen data to assess generalization performance.

The training details are as follows:

- Epochs: Trained for 40 epochs (with early stopping after 400 if loss <0.001 not achieved)
- Batch size: 32
- Learning rate: 0.001
- Loss monitoring: Checkpoints were saved based on the minimum validation loss to prevent overfitting and mode collapse
- Downscaling: HR images (128 × 128) were downscaled to 32 × 32 for training, reducing memory usage while retaining crucial local features.

In the objective evaluation shown in Table 1, two metrics, including structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR), were chosen to quantitatively assess the quality of the super-resolved images.

3.4. Qualitative Analysis

The proposed method was compared to a CNN-based method (referred to as VDSR) and two GAN-based methods (SRGAN and ESRGAN), which integrate perceptual loss through fusion methods to enhance visual quality (Fig. 2). These SR algorithms were all meticulously optimized on the training set for fair comparison. The proposed method primarily relies on a CNN-based approach with a generator network, trained using pixel loss to independently reconstruct HR images from LR images. It comprises both generator and discriminator networks within a GAN model, incorporating perceptual loss for improved visual quality. However, it may lack human perception due to its sole reliance on pixel loss.

The evaluation compared the proposed method against conventional single-image SR algorithms across three test sets. Two different networks (ResNet-50 and modified VGG-19) were used in the deep ResNet–GANs approaches, ensuring a balanced analysis. The proposed method generates realistic and visually appealing results closely resembling natural images, attributed to innovative algorithm design

Table 1. Different objective evaluation metrics

Parameter	**Equation**	**Explanation**
Peak signal-to-noise ratio (PSNR)	$PSNR=10\ log_{10}\ (MAX^2/MSE)$ where MAX is the maximum pixel value of the original image data, and mean squared error (MSE) is the average of the squared deviations between the comparable pixel values of the original and super-resolved images.	A key statistic for assessing the quality of super-resolved images in relation to their original counterparts is the PSNR. In remote sensing, the PSNR is a vital measure of the performance of super-resolution techniques in preserving image information and reducing reconstruction mistakes. A higher PSNR value indicates a better fidelity and less distortion between the original and super-resolved images.
Structural similarity index (SSIM)	$SSIM\ (x, y) = (2\mu x\mu y+c1)\ (2\sigma xy+c2)/ (\mu x2 + \mu y2+c1)\ (\sigma x2 + \sigma y2+c2)$ where x and y represent the original and super-resolved images, respectively; μx and μy denote the mean intensity of x and y images; σx and σy are the standard deviations of x and y images; σxy represents the covariance of x and y images; and $c1$ and $c2$ are constants added to avoid instability when the denominator approaches zero.	The SSIM metric compares the brightness, contrast, and structure of two images to determine their similarity. It yields a number between 0 and 1, where 0 represents total dissimilarity and 1 represents perfect similarity. As SSIM accounts for aspects of human visual perception, unlike standard metrics, such as MSE or PSNR, it is more appropriate for evaluating image quality changes that impact human perception. This makes it especially pertinent to remote sensing applications.

Fig. 2. Comparison of reconstructed details in distinct small targets (i and ii): (A) original image; (B) images reconstructed using VDSR; (C) images reconstructed using SRGAN; (D) images reconstructed using ESRGAN; and (E) images reconstructed using the proposed algorithm in this study
Abbreviations: ESRGAN: Enhanced super-resolution generative adversarial network; SRGAN: Super-resolution generative adversarial network; VDSR: Very deep super-resolution

techniques that balance visual quality optimization and artifact minimization.

3.5. Quantitative Analysis

In this study, two metrics were employed to quantitatively evaluate the SR results: PSNR and SSIM.

A comparison of the PSNR scores for several SR techniques—VDSR, SRGAN, ESRGAN, and the suggested method—across three different datasets is shown in Table 2.

(i) Interpretation of the PSNR metric: A quantitative indicator of image quality; a greater PSNR value denotes a smaller disparity between the original and reconstructed images
(ii) Superiority of the proposed technique: The PSNR scores obtained by the proposed approach are consistently higher than those obtained by VDSR, SRGAN, and ESRGAN on all three datasets
(iii) Dataset-specific performance: Extensive PSNR values for every dataset demonstrate the enhanced image quality preservation performance of the proposed approach

Table 2. Average PSNR/dB for different algorithms

Dataset	**VDSR**	**SRGAN**	**ESRGAN**	**Proposed method**
1st dataset	28.345	26.484	32.896	33.256
2nd dataset	29.275	26.784	31.569	32.886
3rd dataset	29.568	29.725	32.451	34.885

Abbreviations: ESRGAN: Enhanced super-resolution generative adversarial network; PSNR: Peak signal-to-noise ratio; SRGAN: Super-resolution generative adversarial network; VDSR: Very deep super-resolution.

(iv) Low difference in the first dataset: The proposed approach produces a PSNR of 33.256 in the first dataset, demonstrating a small difference between the reconstructed and original images, and a higher level of image quality than the other methods

(v) Consistent performance in the second dataset: The proposed method consistently reduces disparities in image reconstruction, as seen by its high PSNR across datasets, with a score of 32.886 in the second dataset

(vi) Highest PSNR in the third dataset: The proposed method outperforms other methods in producing better image quality and fidelity, as evidenced by its PSNR value of 34.885, the highest in the third dataset

(vii) Implication of results: The higher PSNR values of the proposed method across all tested datasets indicate its usefulness and resilience in a range of settings, suggesting that it excels in reducing disparities between reconstructed and real images for improved image quality.

The average SSIM scores for the proposed method, VDSR, SRGAN, ESRGAN, and other SR methods across several datasets are shown in Table 3.

i. The SSIM compares the brightness, contrast, and structure of the reconstructed and original images; higher values within the range of [0,1] indicate better preservation of these features. Out of all the methods tested, the proposed method recorded the highest SSIM scores, demonstrating its superior ability to maintain image properties, including structural similarities, brightness, and contrast, during the SR process.

ii. In the first dataset, for example, the proposed method achieves an SSIM score of 0.862, indicating that it can maintain image properties more effectively than other methods. It also demonstrates outstanding performance in other datasets, as evidenced by the scores as high as 0.833 in the second dataset and 0.899 in the third.

iii. The higher SSIM values of the proposed method across several datasets highlight its extraordinary performance in preserving image features, indicating its potential for producing high-quality SR outcomes with improved visual integrity.

Table 3. Average SSIM for different algorithms

Dataset	VDSR	SRGAN	ESRGAN	Proposed method
1st dataset	0.745	0.771	0.772	0.862
2nd dataset	0.725	0.774	0.702	0.833
3rd dataset	0.812	0.698	0.806	0.899

Abbreviations: ESRGAN: Enhanced super-resolution generative adversarial network; SRGAN: Super-resolution generative adversarial network; SSIM: Structural similarity index measure; VDSR: Very deep super-resolution.

4. Conclusion

In conclusion, this work introduces a novel approach utilizing ResNet-50 in the generator network of the stable SRGAN, alongside a modified VGG-19 network in the discriminator. Through employing bypass branches and shortcuts in ResNet-50, the network effectively learns residuals and mitigates information loss during transmission. This design, combined with the enhanced feature extraction capabilities of the modified VGG-19 network, improves the overall performance of the SR process through preserving image properties and discerning between real and generated images.

The proposed method was compared against conventional single-image SR algorithms, including VDSR, SRGAN, and ESRGAN, across three datasets. The results consistently demonstrate the superiority of the proposed approach, as evidenced by higher SSIM and PSNR scores. For example, achieving an SSIM score of 0.862 in the first dataset signifies superior preservation of image properties compared to other methods. Furthermore, notable PSNR scores, such as 33.256, 32.886, and 34.885 across the datasets, highlight the capability of the proposed approach to maintain image quality and fidelity.

While the proposed approach demonstrates strong performance in terms of quantitative metrics and architectural innovation, there are a few areas that present opportunities for further enhancement. The use of deep networks, such as ResNet-50 and a modified VGG-19, while beneficial for feature extraction and residual learning, may introduce increased computational demands. The evaluation primarily relies on PSNR and SSIM scores; incorporating perceptual or user-centered assessments in future studies could provide a more comprehensive understanding of visual quality. Lastly, as with many GAN-based models, ensuring stable and efficient training across diverse conditions may benefit from further optimization and experimentation.

Future work should focus on enhancing both the efficiency and generalizability of the proposed approach. This includes exploring adaptive or hybrid loss functions—combining perceptual, adversarial, and contextual losses—to improve visual quality and structural fidelity. To support real-world applicability, evaluations should be extended to remote sensing tasks, such as land cover classification and object detection. Additionally, efforts should be made to develop lightweight architectures for resource-constrained environments and to incorporate domain adaptation techniques, enabling robust performance across varying sensors, regions, and seasonal conditions. These directions aim to refine model effectiveness while broadening its scope and practical relevance.

Acknowledgement

We thank the management and staff of GSSSIETW for providing timely support for completing the work. This research was funded by the "Consolidated University Research for Innovation and Excellence CORE Grant for Women PG College" (DST-CURIE), Department of Science and Technology, Govt. of India, New Delhi, India (Grant no.: DST/CURIE-PG/2022/71).

References

Adarsh, P., Rathi, P., & Kumar, M. (2020). YOLO v3-tiny: Object detection and recognition using one stage improved model. *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India*. IEEE, United States, p687–694. https://doi.org/10.1109/ICACCS48705.2020.9074315

Bu, L., Zhang, J., Zhang, Z., Yang, Y., & Deng, M. (2024). Deep learning for integrated speckle reduction and super-resolution in multi-temporal SAR. *Remote Sensing*, 16(1), 18. https://doi.org/10.3390/rs16010018

Dileep, R., Jayanth, J., Choodarathnakar, A.L., & Ravikiran, H.K. (2025). Walrus optimization algorithm for panchromatic and multispectral image fusion. *Remote Sensing Applications: Society and Environment*, 38, 101562. https://doi.org/10.1016/j.rsase.2025.101562

Dileep, R., Jayanth, J., Ravikiran, H.K., Vedha, A.S., Shyamala, C., & Yuvaraju, T. (2024). Enhancing image fusion via optimized BAT algorithm in Brovey transform for remote sensing images. In: *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*. p1–6. https://doi.org/10.1109/ICKECS61492.2024.10617149

Dong, R., Zhang, L., & Fu, H. (2022). RRSGAN: Reference-based super-resolution for remote sensing image. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–17. https://doi.org/10.1109/TGRS.2020.3046045

Frizza, T., Dansereau, D.G., Seresht, N.M., & Bewley, M. (2022). Semantically accurate super-resolution generative adversarial networks. *Computer Vision and Image Understanding,* 221, 103464. https://doi.org/10.1016/j.cviu.2022.103464

Han, R, Mei, B., Huang, X., Xue, H., Jiang, X., & Yang, S. (2023). Remote sensing image super-resolution adversarial network based on reverse feature fusion and residual feature dilation. *IEEE Access*, 11, 85259–85267. https://doi.org/10.1109/ACCESS.2023.3304050

Haut, J.M., Fernandez-Beltran, R., Paoletti, M.E., Plaza, J., Plaza, A., & Pla, F. (2018). A new deep generative network for unsupervised remote sensing single-image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing,* 56(11), 6792–6810.

He, Z., Li, J., Liu, L., He, D., & Xiao, M. (2022). Multiframe video satellite image super-resolution via attention-based residual learning. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–17. https://doi.org/10.1109/TGRS.2021.3072381 https://doi.org/10.1109/TGRS.2018.2843525

Jayanth, J., Ravikiran, H.K., Shyamala, C., & Dileep, R. (2025). Detecting the stages of ragi crop diseases using satellite data in villages of Nanjangud Taluk. In: Singh, S., Sood, V., Srivastav, A.L., & Ampatzidis, Y., editors. *Hyperautomation in Precision Agriculture.* Academic Press, United States, p131–145. https://doi.org/10.1016/B978-0-443-24139-0.00011-7

Jiang, K., Wang, Z., Yi, P., Jiang, J., Xiao, J., & Yao, Y. (2018). Deep distillation recursive network for remote sensing imagery super-resolution. *Remote Sensing*, 10(11), 1700. https://doi.org/10.3390/rs10111700

Khan, B., Mumtaz, A., Zafar, Z., Sedkey, M., Benkhelifa, E., & Moazam, M. (2023). CGA-Net: Channel-wise gated attention network for improved super-resolution in remote sensing imagery. *Machine Vision and Applications*, 34, 128. https://doi.org/10.1007/s00138-023-01477-0

Lei, S., Shi, Z., & Zou, Z. (2017). Super-resolution for remote sensing images via local-global combined network. *IEEE Geoscience and Remote Sensing Letters*, 14(8), 1243–1247. https://doi.org/10.1109/LGRS.2017.2705985

Li, J., Zheng, K., Gao, L., Han, Z., Li, Z., & Chanussot, J. (2025). Enhanced deep image prior for unsupervised hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–18. https://doi.org/10.1109/TGRS.2025.3531646

Li, J., Zheng, K., Gao, L., Ni, L., Huang, M., & Chanussot, J. (2024). Model-informed multistage unsupervised network for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–17. https://doi.org/10.1109/TGRS.2024.3391014

Li, J., Zheng, K., Gao, L., Ni, L., Xiao, M., & Chanussot, J. (2023). X-shaped interactive autoencoders with cross-modality mutual learning for unsupervised hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–17. https://doi.org/10.1109/TGRS.2023.3300043

Li, J., Zheng, K., Liu, W., Li, Z., Yu, H., & Ni, L. (2023). Model-guided coarse-to-fine fusion network for unsupervised hyperspectral image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 20, 1–5. https://doi.org/10.1109/LGRS.2023.3309854

Li, Q., Yang, R., Xiao, F., Bhanu, B., & Zhang, F. (2022). Attention-based anomaly detection in multi-view surveillance videos. *Knowledge-Based Systems,* 252, 109348. https://doi.org/10.1016/j.knosys.2022.109348

Liu, J., Yuan, Z., Pan, Z., Fu, Y., Liu, L., & Lu, B. (2022). Diffusion model with detail complement for super-resolution of remote sensing. *Remote Sensing*, 14(19), 4834. https://doi.org/10.3390/rs14194834

Min, J., Lee, Y., Kim, D., & Yoo, J. (2024). Bridging the domain gap: A simple domain matching method for reference-based image super-resolution in remote sensing. *IEEE Geoscience and Remote Sensing Letters*, 21, 1–5. https://doi.org/10.1109/LGRS.2024.3345678

Pang, B., Zhao, S., & Liu, Y. (2023). The use of a stable super-resolution generative adversarial network (SSRGAN) on remote sensing images. *Remote Sensing*, 15(20), 5064. https://doi.org/10.3390/rs15205064

Ravikiran, H.K., Jayanth, J., Sathisha, M.S., Yogeesh, G.H., & Dileep, R. (2024). Classification of crop across heterogeneous landscape through experienced artificial bee colony. *SN Computer Science*, 5, 428–435. https://doi.org/10.1007/s42979-024-02790-9

Shi, Y., Jiang, C., Liu, C., Li, W., & Wu, Z. (2023). A super-resolution reconstruction network of space target images based on dual regression and deformable convolutional attention mechanism. *Electronics*, 12(13), 2995. https://doi.org/10.3390/electronics12132995

Sui, J., Ma, X., Zhang, X., & Pun, M.O. (2023). GCRDN: Global context-driven residual dense network for remote sensing image superresolution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 4457–4468. https://doi.org/10.1109/JSTARS.2023.3273081

Tang, E., Wang, L., Wang, Y., Yu, Y., & Zeng, X. (2022). Lightweight frequency-based attention network for image super-resolution. *Journal of Electronic Imaging*, 31(5), 053014. https://doi.org/10.1117/1.JEI.31.5.053014

Veganzones, M.A., Simões, M., Licciardi, G., Yokoya, N., Bioucas-Dias, J.M., & Chanussot, J. (2016). Hyperspectral super-resolution of locally low rank images from complementary multisource data. *IEEE Transactions on Image Processing*, 25(1), 274–288. https://doi.org/10.1109/TIP.2015.2496263

Wang, C., Zhang, X., Yang, W., Li, X., Lu, B., & Wang, J. (2023). MSAGAN: A new super-resolution algorithm for multispectral remote sensing image based on a multiscale attention GAN network. *IEEE Geoscience and Remote Sensing Letters*, 20, 1–5. https://doi.org/10.1109/LGRS.2023.3258965

Wang, J., Gao, K., Zhang, Z., Ni, C., Hu, Z., Chen, D., & Wu, Q. (2021). Multisensor remote sensing imagery super-resolution with conditional GAN. *Journal of Remote Sensing*, 2021, 9829706. https://doi.org/10.34133/2021/9829706

Wang, P., Bayram, B., & Sertel, E. (2022). A comprehensive review on deep learning based remote sensing image super-resolution methods. *Earth-Science Reviews*, 232, 104110. https://doi.org/10.1016/j.earscirev.2022.104110

Wang, X., Sun, L., Chehri, A., & Song, Y. (2023). A review of GAN-based super-resolution reconstruction for optical remote sensing images. *Remote Sensing*, 15(20), 5062. https://doi.org/10.3390/rs15205062

Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., *et al*. (2022). A review of image super-resolution approaches based on deep learning and applications in remote sensing. *Remote Sensing*, 14, 5423. https://doi.org/10.3390/rs1421542

Wang, Y., Lyu, B., Shi, C., & Hu, Y. (2023). Non-parametric simulation of random field samples from incomplete measurements using generative adversarial networks. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 18, 60–84. https://doi.org/10.1080/17499518.2023.2222383

Wang, Y., Shao, Z., Lu, T., Wang, J., Cheng, G., Zuo, X., & Dang, C. (2024). Remote sensing pan-sharpening via cross-spectral-spatial fusion network. *IEEE Geoscience and Remote Sensing Letters*, 21, 37844. https://doi.org/10.1109/LGRS.2023.3337844

Zhang, L., Dong, R., Yuan, S., Li, W., Zheng, J., & Fu, H. (2021). Making low-resolution satellite images reborn: A deep learning approach for super-resolution building extraction. *Remote Sensing*, 13(15), 2872. https://doi.org/10.3390/rs13152872

Zhang, T., Chen, H., Chen, S., & Bian, C. (2022). Edge-enhanced efficient network for remote sensing image super-resolution. *International Journal of Remote Sensing*, 43(14), 5324–5348. https://doi.org/10.1080/01431161.2022.2076217

DOI: 10.6977/IJoSI.202508_9(4).0004
J. Jayanth, H. K. Ravikiran, etc./Int. J. Systematic Innovation, 9(4), 42-53 (2025)

Zhang, Z., Gao, K., Wang, J., Min, L., Ji, S., Ni, C., & Chen, D. (2022). Gradient enhanced dual regression network: Perception-preserving super-resolution for multi-sensor remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5. https://doi.org/10.1109/LGRS.2021.3134798

AUTHOR BIOGRAPHIES

J. Jayanth is a Professor in the Department of Electronics and Communication Engineering, GSSS Institute of Engineering and Technology for Women, Mysuru, and is a distinguished academic with over 15 years of teaching and research experience. He is an active researcher and has completed funded projects and consultancy projects. JJ completed his B.E., M.Tech., and Ph.D. in Electronics and Communication Engineering from Visveswarya Technological University, Belagavi. His research focuses on image processing and remote sensing. As a prolific author, he has published more than 25 research articles in reputed international journals indexed in the SCI and Scopus databases, as well as in prestigious IEEE conferences. His expertise spans artificial intelligence, machine learning, and antenna design and its analysis. As an educator, he has mentored numerous students, integrating theoretical and practical knowledge to prepare them for successful careers.

H. K. Ravikiran was born in Karnataka, India, in April 1987. He received his B.E. degree in Electronics and Communication Engineering from S.J.C. Institute of Technology, Chikballapur, Karnataka, India, in 2008, and his M.Tech. in Digital Electronics and Communication Systems from Malnad College of Engineering, Hassan, Karnataka, India, in 2010. He completed his Ph.D. in Electrical & Electronics Engineering Science from Visvesvaraya Technological University (VTU), Karnataka, India, in 2020. RK is currently serving as a Professor in the Department of Electronics and Communication Engineering at Navkis College of Engineering, Hassan, with 15 years of teaching and research experience. He has made significant contributions to research, publishing 32 papers across Scopus-indexed journals, SCI journals, international journals, and conferences. He has authored two books on the Internet of Things (IoT) and holds three granted patents and a copyright. He is also an active reviewer for prestigious journals and conferences, contributing to the global research community with his expertise and insights. His research interests include image processing, deep learning, and soft computing techniques, where he continues to pioneer innovative methodologies and impactful solutions.

T. Yuvarju is an Assistant Professor in the Department of Electronics and Communication Engineering, GSSS Institute of Engineering and Technology for Women, Mysuru, with over 14 years of teaching and research experience. He completed his B.E. in Electronics and Communication from KVG Sullia and M.Tech. in Digital Electronics and Communication from M.V.J. College of Engineering, Bangalore. He is currently pursuing a Ph.D. in remote sensing from Visveswarya Technological University, Belagavi. With a strong academic background and a passion for research, YT has made significant contributions to the fields of image processing and remote sensing. His research interests include data analytics, machine learning, and artificial intelligence. He has published research articles in renowned journals and IEEE conferences, establishing himself as a prolific researcher in his field. YT is also actively involved in academic and professional development initiatives, participating in international workshops and conferences to disseminate his findings and collaborate with experts worldwide.

R. Dileep is an Assistant Professor in the Department of Electronics and Communication Engineering, GSSS Institute of Engineering and Technology for Women, Mysuru, with over 14 years of teaching and research experience. He completed his B.E. in Electronics and Communication from RNSIT Bangalore and M.Tech. in Digital Electronics and Communication from Malnad College of Engineering, Hassan. He is currently pursuing a Ph.D. in remote sensing from Visveswarya Technological University, Belagavi. With a strong academic background and a passion for research, DR has made significant contributions to the fields of image processing and remote sensing. His research interests include data analytics, machine learning, and artificial intelligence. He has published research articles in renowned SCI-indexed journals and IEEE conferences, establishing himself as a prolific researcher in his field. DR is also actively involved in academic and professional development initiatives, participating in international workshops and conferences to disseminate his findings and collaborate with experts worldwide.

DOI: 10.6977/IJoSI.202508_9(4).0005
S. Jafri, etc./Int. J. Systematic Innovation, 9(4), 54-70 (2025)

Harnessing mobile multimedia for entrepreneurial innovation and sustainable business growth

Samreen Jafri[1], Kamal Upreti[2]*, Rishu Bhardwaj[3], Balraj Verma[3], Rituraj Jain[4]

[1]Department of Administrative Sciences, Applied College, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

[2]Department of Computer Science, School of Sciences, CHRIST University, Delhi NCR, Ghaziabad, India

[3]Chitkara Business School, Chitkara University, Rajpura, Punjab, India

[4]Department of Information Technology, Faculty of Engineering and Technology, Marwadi University, Rajkot, Gujarat, India

*Corresponding author E-mail: kamal.upreti@christuniversity.in

(Received 3 March 2025; Final version received 27 April 2025; Accepted 16 May 2025)

Abstract

This research investigates the role of mobile multimedia platforms and artificial intelligence (AI) in driving innovation and ensuring the sustainability of entrepreneurial businesses, focusing particularly on technology acquisition, integration, and infrastructure. For data collection, the study employed a quantitative research design and surveyed 150 Indian technology firms that had adopted mobile multimedia applications. Structural equation modeling was used to analyze the data, supported by descriptive statistics, correlation, regression analysis, and mixed methods to understand the adoption and use of digital technologies for innovation activities. The results show that AI-driven applications, when combined with multimedia content and real-time analytics, significantly enhance entrepreneurial innovation by improving operational efficiency, increasing customer engagement, and facilitating expansion into new international markets. Companies utilizing mobile multimedia platforms gain a competitive advantage, translating into long-term business growth and sustainability. This research contributes to the literature on AI and entrepreneurship in the context of digital transformation, highlighting the need for startups to invest in AI-enabled mobile technologies. It equally serves policymakers by informing the regulation of an environment that promotes innovation and business sustainability through digital initiatives. This research addresses a significant gap in the literature by providing evidence on how AI acts as a driver of change and provides insight into the adoption of new technologies in the context of entrepreneurship, an area that remains largely underexplored.

Keywords: Artificial Intelligence, Digital Transformation, Innovation Management, Mobile Multimedia Systems, Sustainable Business Models

1. Introduction

Investing in technology is perhaps one of the most complicated strategic choices business owners encounter in today's highly competitive world (Ashton & Stacey, 1995; Quinn et al., 2015). Businesses seek to respond to these changes by fostering new processes and advancing the ongoing race for innovation. It is evident that companies are aware of the accelerating pace of change in modern society, which has made environmental shifts, particularly technological advancements, a major driver of innovation and transformation. Businesses that want to maintain a competitive edge in such volatile and fast-changing circumstances must remain agile (Snihur et al., 2021).

Mobile technology skills stand out as some of the most significant factors driving entrepreneurial success today, due to their profound influence on business growth. Technology, as defined by an organization, facilitates a cost-efficient approach that improves productivity, performance, and results (Lee et al., 2019; Obschonka

& Audretsch, 2020). The development of artificial intelligence (AI) and mobile multimedia technologies has transformed business activities, paving the way for creativity, productivity, and deeper interconnectivity. Entrepreneurs use these emerging technologies to innovate business processes, increase operational productivity, and enter new markets (Maoning, 2022). Multimedia communication technologies and mobile platforms enable timely interaction and improve the quality of customer service, facilitating real-time, mobile multimedia personalized communication and decision-making. These efforts highlight the significant role of mobile multimedia technology in achieving business success in a highly competitive global market (Shinde & Prasad, 2021). Moreover, together with big data, mobile networks, and multimedia technologies create limitless opportunities for entrepreneurs to understand consumers and the market, thereby making informed decisions (Slyusar et al., 2024). In addition, with the advancement of extended reality technologies, more sophisticated customer engagement and services are being offered, allowing for better business development opportunities. Entrepreneurs can leverage mobile computing along with AI to create intelligent, responsive systems that automatically meet user needs and desires, thereby improving customer satisfaction and brand loyalty (Rajesh et al., 2023).

When an organization shifts toward technology-centric approaches, strong backing from top management is important. The increasing capabilities of society are prompting corporations to recognize the need to stay ahead of the competition, which has led to an increased dependence on leadership for employee training and technology adaptation (Chege & Wang, 2020; Shah et al., 2013). This involves understanding the gap in competencies and ensuring that educators and practitioners stay current in their knowledge to meet the evolving needs of the industry (Li & Chan, 2019).

The rapid growth of information and communications technologies has greatly impacted businesses in terms of operations, required skills, organizational culture, and market trends (Leischnig et al., 2017; Martínez-Caro et al., 2020). While these regions have developed technologically, innovation is primarily driven by business needs and the adoption of new technologies by many modern businesses (Park & Mithas, 2020). In this regard, communication and information have become essential resources for entrepreneurial activity. Virtually no aspect of life today remains untouched by information technology (IT), making it crucial for business success. Entrepreneurship involves the creation of economic activity through innovation, taking financial and intellectual risks, and mobilizing resources for self-reliance (Steininger, 2019; Uhlenbruck et al., 2006). With these new features and characteristics, economic growth and job opportunities have significantly increased, and the digital form of entrepreneurship, largely driven by technology, is positively changing the economy (Jafari-Sadeghi et al. 2021).

Novel entrepreneurs substantially expand the economy, create employment opportunities, and enhance the quality of social services. The transformation from an industrial economy to an information society highlights an important aspect of economic development – entrepreneurial activity (Rehman et al., 2020; Uhlenbruck et al., 2006). Without knowledge, determining market gaps and providing innovative solutions becomes impossible. Therefore, information and skills are vital to entrepreneurial activities, just as communication is fundamental. The dissemination of information, as well as the efficiency with which it is communicated, and, in turn, the level of professional interaction embraced, have greatly increased due to technological development (Sihite & Prihandini, 2019). These inventions, in turn, have radically affected education by creating favorable conditions for launching and nurturing entrepreneurial and new venture opportunities (Kumar, et al., 2022; Kumar et al., 2022). The impact of modern technology is global, sociocultural, and economic, making it a crucial aspect of human and community life. Over the past 20 years, technology-intensive developed industrial countries have become increasingly important drivers of innovation and economic advancement.

This research investigates the effects of mobile multimedia technologies, such as AI-driven multimedia platforms, and other digital communicative tools on innovative entrepreneurship and organizational performance. The study sheds light on technology acquisition, integration, and infrastructure by examining the mobile multimedia technologies ecosystem. The results have real-world implications for companies looking to leverage mobile and multimedia technologies to achieve sustainable growth or gain a competitive edge.

The rest of the paper is organized as follows: Section 2 provides a critical review of the research on mobile multimedia technologies and their contributions to entrepreneurship. Section 3 discusses the study's methodological approach, including how the data were collected, the sample of interest, and the methods of analysis used. Section 4 presents the results related to technology acquisition, integration, and infrastructure in entrepreneurial innovation. Section 5 concludes by articulating the major findings, highlighting the practical significance of the results, and identifying areas for further research.

2. Literature Review

(i) Mobile Multimedia and Entrepreneurial Innovation

Jiao et al. (2016) researched entrepreneurial potential and its relationship to innovation, specifically focusing on how an owning entity affects that relationship. The study, which examined 788 publicly traded firms, found that strategic management skills enhance technological innovation. In addition, the research highlighted that entrepreneurship involves the integration of technological savviness along with a variety of other skills.

Poudel et al. (2019) investigated the impact of technical expertise, client perception, and entrepreneurial initiative on business performance. Using structured modeling approaches, Poudel et al. conducted a randomized survey of selected small and medium-sized enterprises (SMEs) in a metropolitan area in the southeastern United States. The results showed that the technological capabilities of innovative, people-centered firms are a prerequisite for business expansion and economic growth.

Kheni (2017) classified successful entrepreneurship as targeting niche markets, which result from the fusion of innovation and technology. Similarly, Lin et al. (2021) and Martin-Rojas et al. (2019) examined the impact of integrating infrastructure and technology on businessmen environments, as well as how technological capabilities affect entrepreneurial activities in new ventures.

Usai et al. (2021) sought to understand the relationship between implementing digital technologies and the innovative performance of startups. The study, which focused on the adoption of information and communication technologies (ICT) in European businesses and their performance, used principal component analysis and multiparameter statistical analysis. They found that while research and development (R&D) spending is the most promising predictor of innovation, technological processes received limited attention in relation to process innovation.

(ii) AI-Driven Applications in Business Growth

Amouri et al. (2021) focused on exploring the factors that may enable or inhibit young entrepreneurs from embarking on new social business ventures. Their study emphasized that an entrepreneurial and psychosocial orientation toward technology is favorable for entrepreneurs seeking to start social business ventures. Moreover, Abubakre et al. (2022) analyzed the contribution of IT culture and individual IT creativity to the achievement of digital entrepreneurship. The results of their study indicated a significant correlation between the cultures of effective software engineering and marketing technology.

Jafari-Sadeghi et al. (2021) investigated the impact of technology on entrepreneurial activities, also considering the effects of digital transformation on value creation. Their study, based on data from 28 European countries, examined the impact of digitalization on entrepreneurship from a unique perspective. The conclusions advanced the concept of digital entrepreneurship as the combination of IT and business, which has particular and profound consequences for business ventures and entrepreneurs.

Matejun (2016) argued that the core aim of a firm's technological entrepreneurship is to increase its ability to innovate through the synergy of internal resources and external possibilities. These external possibilities are provided by R&D institutions and entities in high-technology industries. The study also highlighted the need to absorb technological and innovative outputs from academia into the commercial sector.

Oyero & Oyedele (2022) examined the relationship between techno-entrepreneurship and the sustained growth of SMEs from the perspective of intellectual property rights, R&D, and innovation. Employing linear regression analysis based on a randomly selected sample of 126 agrobusiness survey responses, the study found that techno-enterpreneurship positively and significantly affects the revenues and profits of SMEs.

Kilintzis et al. (2023) adopted employment in the high-technology sector as an indicator of technological entrepreneurship and analyzed several macroeconomic variables at the country level. This research evaluated the statistically significant region-specific quantitative variables that explain differences in technological entrepreneurship within the European Union. The analysis confirmed the importance of education, gross domestic product, and venture capital funding to the distribution of high-tech jobs in Europe. These results have significant policy implications for promoting technological entrepreneurship in the region.

(iii) Technology Integration and Organizational Performance

Sonika et al. (2023) explored how strategic and marketing planning influence the financial outcomes of manufacturing businesses. Based on a sample of 137 employees at Nestlé Foods Nigeria Plc, the study found that strategic planning, combined with the use of technology, positively impacts overall organizational performance. The researchers suggested that, to achieve greater competitiveness, the firm should focus on innovation and increase investment in employee training to enhance productivity.

Romano et al. (2016) argued that the current economic landscape reflects an entrepreneurial economy characterized by turbulence and uncertainty. They claimed that an entrepreneurial approach, particularly one that is technology-centric, has emerged and is effectively driving business success while simultaneously promoting economic growth. Rather than being solely market-oriented, modern entrepreneurship is becoming increasingly knowledge-based and technology-oriented, as new technologies enable the conception, execution, and renewal of business models. Such technological entrepreneurs act as active learning resources who contribute to the creation, dissemination, and absorption of knowledge within a multifaceted innovation environment.

Roemintoyo et al. (2022) analyzed the use of interactive multimedia in educational processes. While students strongly supported the adoption of ICT, teachers faced challenges in its implementation. The research suggested recommended greater emphasis on output quality and highlighted the importance of interactive multimedia in education. In addition, new forms of digital entrepreneurship, such as the use of social media and smartphone apps, are being employed to engage customers, broaden public perception of artists, increase willingness to collaborate with them, and deepen both professional and audience engagement in entrepreneurship (Psomadaki et al., 2022).

2.1. Research Gap

Entrepreneurial activities, management skills, and customer attitudes have consistently been key indicators in assessing the factors that contribute to business success. The moderating role of ownership on entrepreneurial intention has been studied, along with its correlation with technological development, identified as a primary driver of business investments in new digital processes, and its overall impact on effectiveness and growth. Furthermore, aspects like infrastructure and technological investments, their acquisition, and various combinations thereof in relation to business performance have been thoroughly explored. The intricate relationship between IT investment and the electronic business operations of enterprises has also been analyzed. Other studies have focused on factors that either facilitate or hinder young entrepreneurs in launching inclusive business ventures, the impact of digital competencies on a firm's dynamic capabilities, and the role of an IT culture that fosters continuous individual learning as a foundation for successful digital businesses. Notably absent from the literature, however, is a focused analysis of the relationship between advancements in IT and their impact on entrepreneurial innovation. From this perspective, the present research aims to fill that gap by determining how various attributes of IT affect the innovation potential of entrepreneurs.

3. Technology Adoption and Its Impact on Entrepreneurial Innovation

3.1. Theoretical Foundations of Technology Adoption

Established theoretical frameworks like the Innovation Diffusion Theory (IDT) and the Technology-Organization-Environment (TOE) framework help explain the impact of technology acquisition, technology integration, and technology infrastructure in fostering entrepreneurial innovation. These frameworks provide insight into how businesses incorporate technology to gain a competitive advantage.

The IDT focuses on the spread of technological innovations within an industry and the factors that influence adoption, such as relative advantage, compatibility, complexity, trialability, and observability. In the context of this study, moderate acquisition and assimilation of new technologies support sustained innovation and aggressive market expansion strategies.

Similarly, the TOE framework posits that the adoption of new technologies is influenced by specific technological, organizational, and environmental factors. The relevance of technology acquisition, integration, and infrastructure in this study aligns with themes of digital transformation and entrepreneurship. Adopting AI-based solutions and multimedia applications is less of a challenge for companies with highly developed technological infrastructures, leading to greater operational efficiency and growth.

3.2. Key Technological Drivers of Entrepreneurial Innovation

The adoption and integration of new technologies like AI, mobile multimedia platforms, and digital automation have become key drivers for entrepreneurial innovation. A company's ability to leverage the latest technologies determines the extent of its market expansion, the efficiency of its operations, and its long-term sustainability. This study identified three key components of technology integration for innovation: acquisition, integration, and overall infrastructure.

3.2.1. Technology Acquisition as a Catalyst for Market Expansion

Technological acquisition is important for entrepreneurial innovation, as it enables the procurement and implementation of advanced digital solutions. Companies that adopt AI applications, big data analytics, and mobile multimedia technologies tend to be more responsive to market opportunities and evolving consumer demands. Companies that implement a technology-centered acquisition strategy experience higher rates of innovation and competitive success as a result of advanced digital tool implementation.

Furthermore, the use of AI-driven multimedia platforms allows a company to automate customer service processes, thereby enhancing user satisfaction and brand loyalty. Businesses that adopt more advanced technologies gain a strategic advantage over competitors slower to undergo digital transformation, thus enabling them to distinguish themselves in the market.

3.2.2. Technology Integration for Organizational Agility

It is not enough for firms to obtain modern technologies; they need to incorporate them into their operational frameworks. This process is referred to as technology integration. This involves embedding technology into business systems to enable process automation, coordination, and real-time decision-making. Businesses that rely on AI, cloud computing, and mobile applications for core functions tend to respond more effectively to market and industry changes or disruptions.

Strategic agility, resource optimization, and prompt responses to emerging business challenges are attainable due to enhanced interdepartmental coordination and efficient workflows. For example, AI-powered analytics tools help companies track market and consumer behavioral trends, enabling them to adjust operations in real time. By fostering a culture of digital innovation, organizations can mitigate competition and seize new opportunities.

3.2.3. Strengthening Innovation through Robust Technology Infrastructure

The effective use of technology is a primary requirement for long-term entrepreneurship success. Weak IT infrastructures create scaling challenges, increase security threats, and hinder seamless business operations. Companies with strong IT foundations, especially those leveraging cloud computing, high-speed internet, and AI-powered automation, tend to maintain innovation over longer periods.

A well-established technology framework allows companies to adopt new digital technologies with minimal interference, ensuring continuous process optimization and business responsiveness to market demands. In addition, businesses with robust IT frameworks can utilize new technologies such as Extended Reality, blockchain, and Internet of Things in their product offerings, thereby creating unique customer experiences.

Moreover, a strong technology infrastructure enhances entrepreneurs' access to capital, promotes professional networking, and facilitates participation in the innovation ecosystem, contributing to more sustainable business ventures. Firms that build digital infrastructure stand a higher chance of receiving venture capital funding and forming strategic business partnerships, which drive growth and enable sustainable competition in the long run.

4. Methodology

4.1. Research Motivation and Objectives

Entrepreneurs and IT are two renowned and important aspects that currently affect the world economy. The rapid advancement of IT industries and their integration into society, as well as the correlating interaction between entrepreneurs and innovation, have highlighted the necessity of detailed research at the industrial level. Furthermore, the ever-increasing demand for both foundational and applied research, carefully crafted to produce cutting-edge results in innovation and entrepreneurship within the scope of IT, is evident.

The scope of IT-driven innovation and entrepreneurship now includes the latest research on phenomena, processes, and practices performed at an international level, focusing on the conception, proliferation, diffusion, and application of IT. Moreover, it offers modern, thoroughly researched, and practical insights into business strategy, model development, and execution, valuable for academic researchers, entrepreneurs, and business practitioners alike.

This research aims to achieve the following primary objectives:

(i) To understand how IT drives advancements in entrepreneurship led by technology.
(ii) To determine the effectiveness and success of strategic IT applications in supporting startup operations.
(iii) To examine the relationship between innovation and business enterprises in relation to their competitive edge.

4.2. Hypothesis Development

The hypotheses that accompany the set objectives and questions of this study are presented below. Fig. 1 illustrates the conceptual framework for the proposed hypotheses.

Entrepreneurs utilize previously acquired knowledge by adopting relevant technologies that allow firms to communicate effectively and exploit future market opportunities. Technology acquisition enables new market entrants to develop and promote emerging technologies. The use of technology by entrepreneurs enables them to capitalize on market trends, leading to

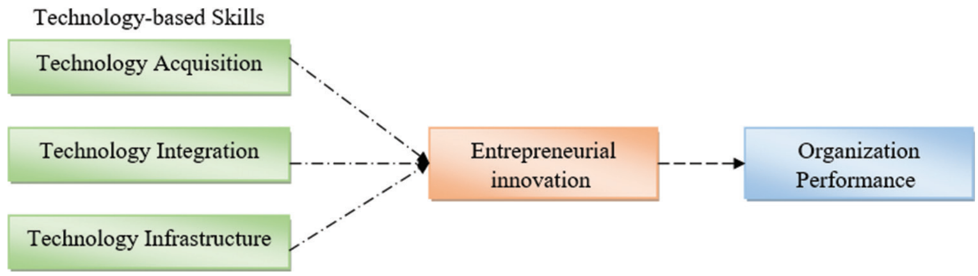


Fig. 1. Framework—research hypothesis

both technological and business growth. The process of recognizing new ideas, technologies, and concepts lies at the core of entrepreneurship. Without the ability to identify and act on opportunities, effective entrepreneurship cannot occur. While the recognition of opportunities may be distributed across a wide spectrum of individuals, productive entrepreneurs are those who can perceive market opportunities and strategically position themselves to benefit from them. Opportunities must be attractive and feasible in the context of the marketplace to be successfully exploited. Firms also benefit from entrepreneurial activities through networking, utilizing both internal and external resources to achieve competitiveness. Based on the above discussion, the following hypotheses are formulated:

- H1: Technology acquisition has a significant positive impact on organizational performance.
- H2: Technology integration mediates the relationship between technology acquisition and organizational performance.
- H3: Technology infrastructure strengthens the relationship between technology integration and organizational performance.
- H4: Mobile multimedia adoption positively influences sustainable business growth.

4.3. Research Design, Conceptual Framework, and Data Collection

This part of the project examined the core structures of influence that mobile multimedia technologies and AI-powered tools have on entrepreneurial innovation. These emerging technologies tend to alter business models by improving engagement, decision-making, and customer interactions. AI-empowered multimedia applications assist in the automation of business processes, foster more flexible approaches to the marketplace, and help devise strategies for long-term sustainability, elements that are pertinent to contemporary entrepreneurs.

This research focused on Indian technology companies that provide R&D services and utilize mobile multimedia technologies. These organizations have embraced a workplace culture grounded in socio-technological motivation, where digital integration of values enables innovation and sustainable competitiveness.

The conceptual framework for this study integrates elements from two widely accepted theoretical models: IDT and the TOE framework, to systematically explain the drivers of entrepreneurial innovation and sustainable business performance.

- “Technology Acquisition” corresponds to the “Relative Advantage” and “Compatibility” components of IDT, as organizations perceive the adoption of new technologies as beneficial and aligned with their existing processes. It also relates to the “Technological Context” of the TOE framework, capturing the extent to which firms acquire innovative mobile multimedia and AI-driven technologies.
- “Technology Integration” aligns with the “Complexity” and “Trialability” dimensions of IDT. Integration efforts may involve overcoming perceived difficulties (complexity) in embedding technologies into operations, while pilot implementations (trialability) enable firms to assess their effectiveness. Within the TOE framework, “Technology Integration” also belongs to the “Technological Context,” emphasizing how internal technological capabilities facilitate innovation.
- “Technology Infrastructure” is mapped to “Observability” in IDT, reflecting how the outcomes of a strong technological foundation (such as enhanced performance and innovation) are visible to internal and external stakeholders. Within the TOE framework, it falls under the “Organizational Context,” representing the internal readiness, skills, and technological resources required to support innovation and entrepreneurial growth.

Each conceptual construct is, therefore, explicitly grounded in recognized theoretical components, ensuring that the framework is well supported by established theories on innovation and technology adoption.

A standardized questionnaire was formulated and distributed to 150 entrepreneurs and managers leading the departments in information management, business performance evaluation, and digital transformation business strategies in their respective organizations. The primary objective of this survey was to examine how mobile multimedia technologies contribute to achieving sustainability in entrepreneurship.

The survey instrument was designed based on established frameworks adapted from prior validated studies related to technology adoption, integration, infrastructure development, and organizational performance. Each construct was measured using multiple items on a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Respondents evaluated statements regarding their organization's practices and capabilities. Sample items included: "The organization effectively integrates mobile multimedia technologies across its operations" (for "Technology Integration") and "The existing technological infrastructure adequately supports business activities" (for "Technology Infrastructure"). The internal consistency of each construct was verified using Cronbach's alpha, and confirmatory factor analysis was employed to validate the measurement model. Detailed survey items and reliability statistics are presented in Table 1.

To provide a clear overview of the constructs used in this study, Table 1 summarizes the number of measurement items for each construct, a sample survey statement, and the corresponding reliability coefficient (Cronbach's alpha). The strong internal consistency values indicate that the measurement instruments were both reliable and appropriate for subsequent analysis.

4.4. Data Analysis and Model Evaluation

To understand the effect of IT on entrepreneurial innovations, a statistical approach was adopted. Initially, descriptive statistics were employed to determine the mean and standard deviation of the responses, which explained the central tendency and variability of the data. Subsequently, a correlation analysis was performed to assess the relationships between key variables and entrepreneurial innovation, using the most significant associations defined in previous steps.

The study employed structural equation modeling (SEM) to evaluate the relationships between technology acquisition, integration, infrastructure, and organizational performance. The SEM procedure involved: (i) Model specification based on the hypothesized framework, (ii) confirmatory factor analysis to validate the measurement models, (iii) evaluation of model fit indices (e.g., Comprehensive Fit Index, Root Mean Square Error of Approximation, standardized root mean square residual [RMR]), and (iv) estimation of structural paths. Subsequently, multiple regression analysis was performed to validate significant predictors and assess the robustness of the SEM findings. Additional regression analysis was conducted to evaluate the hypotheses and determine the significance and magnitude of the relevant statistical coefficients. The study aimed to interpret the relationships that technology acquisition, integration, and infrastructure have with entrepreneurial innovation and organizational performance, with emphasis placed on identifying which relationships were statistically significant.

5. Result and Discussion

5.1. Profile of Demographic Respondents

The demographic profile data are summarized in Table 2, which presents the survey outcomes. The data reveal that the sample consisted of 41.9% females and 58.1% males. The respondents' ages were categorized as follows: 21.6% were between 18 and 25 years of age, 26.4% were within the 26 to 35 age range, and 31.2% in the 36 to 45 age bracket. Interestingly, there were no respondents older than 45 years.

With respect to marital status, the majority of respondents (66.5%) reported being married, while 33.5% were single. Regarding educational attainment, 33.5% had completed an undergraduate degree, while 25.2% held a postgraduate degree. High school dropouts accounted for 19.3%, and 23.3% had no formal education.

Table 1. Summary of constructs, sample items, and reliability measures

Construct	Number of items	Sample survey item	Cronbach's alpha (α)
Technology Acquisition	4	"The organization adopts the latest mobile technologies effectively."	0.82
Technology Integration	5	"Mobile multimedia technologies are integrated across business units."	0.85
Technology Infrastructure	4	"The organization's technological infrastructure supports digital growth."	0.80
Organizational Performance	5	"The organization's use of mobile technologies improves overall performance."	0.87

Table 2. Respondents' demographic analysis

Respondent characteristic	*n*	%
Gender		
Male	187	58.1
Female	135	41.9
Total	322	100.0
Age		
18–25 years	97	30.1
26–35 years	80	24.8
36–45 years	145	45.0
Total	322	100.0
Marital status		
Married	214	66.5
Unmarried	108	33.5
Total	322	100.0
Educational qualification		
Higher secondary	62	19.3
Graduate	108	33.5
Post graduate	77	23.9
Others	75	23.3
Total	322	100.0
Monthly income		
<40,000	70	21.7
40,001–70,000	84	26.1
70,001–100,000	70	21.7
Above 100,000	98	30.4
Total	322	100.0

Note: *n* denotes frequency and % indicates percentage.

Respondents were also categorized based on their monthly income: 21.7% earned below the 10,000 mark, 26.1% earned between 10,001 and 20,000, 21.7% earned between 20,001 and 30,000, and 30.4% earned above 30,000. This information helps to understand the overall income status and contributes to building a comprehensive demographic profile of the respondents in relation to their socioeconomic status.

5.2. Evaluation of Descriptive Test

The metrics for the study's variables are captured in the analysis, with a focus on their means and standard deviations, as presented in Table 3. The results show that the mean scores associated with entrepreneurial innovation range from 3.16 and 3.96. The highest mean (3.96) was observed for the item related to searching for and retrieving new information for business purposes, which had a standard deviation of 1.054.

With regard to technology acquisition, the mean scores ranged from a low of 3.19 to a high of 4.06, with developing strategic options recording the highest mean of 4.04 and a standard deviation of 0.910. Similarly, the "Technology Integration" dimension had means ranging from 3.49 to 3.55, where potential for growth had the highest mean of 3.55 with a standard deviation of 0.960.

For "Technology Infrastructure," the mean scores varied between 3.23 and 3.92. The item developing and launching solutions to the market with speed recorded

Table 3. Descriptive analysis of study variables

Description of variables	**Mean**	**SD**
Entrepreneurial innovation		
Identify business opportunities	3.44	1.096
Focus on how to take advantage of opportunities	3.61	1.075
Improve the situation as industry changes	3.96	1.054
Take the time to look for fresh knowledge and apply it to your work	3.16	1.253
Analyze developing markets and decide how to profit from them	3.51	1.195
Look for changes in the market and create strategies for utilizing such developments for financial advantage	3.50	1.097
Analyze the potential for achievement and failure carefully	3.48	1.045
Technology acquisition		
Enhance the market share	3.59	1.099
Develop strategic options	4.04	0.910
Gain efficiency improvements	3.19	1.235
Technology integration		
Successful implementation of business strategies	3.49	1.208
More effective communication	3.49	1.106
Potential for growth	3.55	0.960
Technology infrastructure		
To provide customers with a good experience, the website should be accessible without interruption	3.47	1.233
Rapidly create and introduce solutions to the market	3.92	1.082
Gather timely data to aid in decision-making	3.23	1.201
Organization performance		
Improve employee productivity	3.49	1.203
Reduce business and other risks	3.48	1.111
Increase efficiency with minimum resources	3.42	1.097
Establish innovation ideas	3.60	1.098
Assist employees to accomplish goals	3.98	1.027

the highest mean of 3.92, with a standard deviation of 1.082. Finally, the means for organizational performance ranged from 3.42 to 3.98, with the highest mean of 3.98 corresponding to assisting employees in accomplishing goals. The standard deviation for this category was 1.027.

Fig. 2 illustrates the mean scores and standard deviations of variables measuring "Entrepreneurial Innovation," "Technology Acquisition," "Technology Integration," "Technology Infrastructure," and their impact on "Organizational Performance." These metrics help to clarify the importance of various factors in business success, technological adoption, and organizational efficiency. In addition, the indices are plotted to demonstrate the goodness of fit (GIF) of the framework, thereby offering confidence in its validity and reliability.

5.3. Analysis of Structural Equation Modeling

Structural equation modeling is an advanced statistical tool used for the measurement and analysis of the relationships between measurable and unmeasurable variables. Unlike simpler approaches such as the coefficient of determination, SEM considers linear causal relationships along with estimation errors, making it more powerful than traditional methods. SEM can integrate multiple dependent and independent variables, providing deeper insight into the interactions among various factors within a dataset.

5.3.1. Indices of Fit

To evaluate measurement error and ensure the accuracy of model estimation, model fit must be confirmed before analyzing the results. This process includes an analysis of additional indices beyond factor loadings and descriptive statistics derived from experimental data. Taken collectively, all indices demonstrate that the model successfully establishes and represents the relationships among parameters, producing a robust and meaningful model fit.

The GFI and adjusted GFI (AGFI) indices tend to increase with larger sample sizes and can, in some cases, exceed the set cut-off points. However, when the number of indicators per factor or the total number of factors to be included is high, especially in studies with small sample sizes, GFI and AGFI values often decrease. The optimal cut-off points for GFI, AGFI, and RMR can vary significantly depending on sample size and model specification. Generally, GFI and AGFI are considered acceptable when their values exceed 0.

An acceptable model fit is indicated by the Normed Fit Index and the other related indices, as shown in Table 4, with the Normed Fit Index scoring 0.933. Similarly, the Suggestive Index Fulfilment Fit scored 0.666, and the Incremental Fit Index scored 0.946, suggesting a substantial fit. In addition, the model's strength is further reinforced by the Comprehensive Fit Index, scoring 0.980. The Parsimony Normed Fit Index, which scored 0.274, also supports the sufficiency of the model parameters in the economy claim.

Most of the indices listed in Table 4 fall within acceptable ranges, confirming the overall validity of the model estimation. The scope of these indices strengthens the proposed model and demonstrates its congruence with the sample data. The fit values presented reflect the degree of alignment between the proposed model and the empirical data, confirming the model's validity for more in-depth analysis.

These indices indicate how well the conceptual model aligns with the observed data. The pie chart shown in Fig. 3 categorizes them according to their fitness levels, including Acceptable, Good, and Middle.

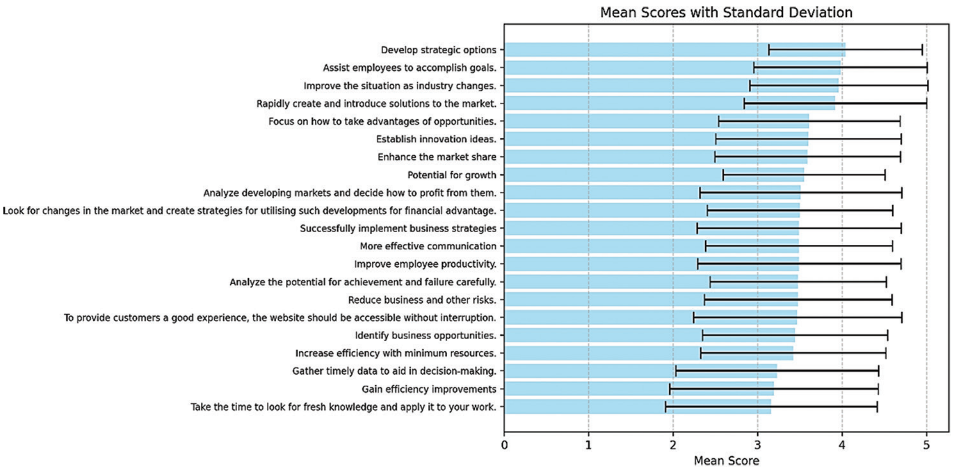


Fig. 2. Entrepreneurial innovation and technology adoption: analyzing key factors for organizational performance and strategic growth

Table 4. Summary of model fit indices and their obtained fitness values

Criteria of goodness of fit	Indices of fit	Obtained fitness value	Fitness condition
Goodness of Fit Index	GFI	0.902	Acceptable
Root Mean Square Residual	RMR	0.024	Acceptable
Adjusted Goodness of Fit Index	AGFI	0.508	-
Parsimony Goodness of Fit Index	PGFI	0.180	Middle
Normed Fit Index	NFI	0.913	Acceptable
Relative Fit Index	RFI	0.910	Acceptable
Incremental Fit Index	IFI	0.915	Acceptable
Comparative Fit Index	CFI	0.915	Acceptable
Parsimony Normed Fixed Index	PNFI	0.274	Good

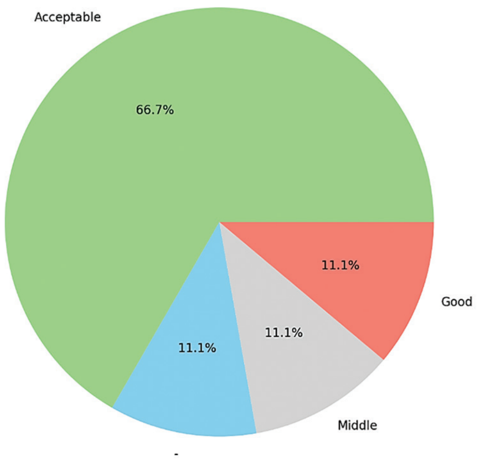


Fig. 3. Distribution of model fitness conditions

5.3.2. Hypothesis Testing

The process of hypothesis testing provides insight into the role IT systems play in enabling entrepreneurial technological development. In this context, the *p*-value defines the statistical relevance of the relationships being analyzed. The threshold for statistical significance is a $p<0.05$.

As shown in Table 5, the $p=0.001$ for technology acquisition is considerably lower than the cut-off. This suggests that technology acquisition has a significant positive effect on entrepreneurial innovation. Similarly, the *p*-value for technology integration is also 0.001, indicating that integration of technology likewise has a significant positive effect on entrepreneurial innovation.

The *p*-value for "Technology Infrastructure" is also 0.001, which supports the assumption that it has a substantial impact on entrepreneurial innovation. The final factor considered was organizational performance, which also had a $p=0.001$. Therefore, it can be concluded that organizational performance has the most significant impact on entrepreneurial innovation.

Overall, all the above-mentioned factors—"Technology Acquisition," "Technology Integration," "Technology Infrastructure," and "Organizational Performance"—make significant contributions to entrepreneurial innovation. This highlights the importance of IT systems in driving business growth and innovation.

The statistical analysis confirms that all four formulated hypotheses (H_1, H_2, H_3, and H_4) are supported and validated, based on the significant *p*-values obtained (<0.05) alongside the robust regression coefficients, which affirm the effects of technology adoption, integration, and infrastructure on entrepreneurial innovation and organizational performance.

5.4. Correlation Analysis

Table 6 presents the relationships analyzed, with an emphasis on one dependent variable, entrepreneurial innovation, measured using economic indicators, and four independent variables: "Technology Acquisition," "Technology Integration," "Technology Infrastructure," and "Organizational Performance." This table depicts the connections between technology usage and entrepreneurial innovation, showing a correlation among the dependent variables in the study. Thus, the research outcomes would remain valid if any independent factor were replaced with entrepreneurial innovation.

Moreover, the increasing correlation between "Technological Infrastructure," "Technology Integration," "Technology Acquisition," and "Organizational Performance" demonstrate their strong effects on entrepreneurial innovation. These findings support the technological ecosystem hypothesis, which suggests that an increase in innovation and business development activities is achieved by a well-developed economy.

As depicted in the heatmap in Fig. 4, the correlations between "Technology Acquisition," "Technology Integration," "Technology Infrastructure," and "Organizational Performance" are evaluated.

Table 5. Hypothesis testing of information technology and entrepreneurial innovation

Hypothesis	**Relationship**	***p*-value**	**Impact**
H_1	Entrepreneurial Innovation→Technology Acquisition	0.001	Significant impact
H_2	Entrepreneurial Innovation→Technology Integration	0.001	Significant impact
H_3	Entrepreneurial Innovation→Technology Infrastructure	0.001	Significant impact
H_4	Entrepreneurial Innovation→Organizational Performance	0.001	Significant impact

Table 6. Correlation analysis between information technology and entrepreneurial innovation

Control variables	**Technology Acquisition**	**Technology Integration**	**Technology Infrastructure**	**Organization Performance**
Entrepreneurial Innovation				
Technology Acquisition	1.000	0.395	0.596	0.083
Technology Integration	0.395	1.000	0.363	0.489
Technology Infrastructure	0.596	0.363	1.000	0.008
Organization Performance	0.083	0.489	0.008	1.000

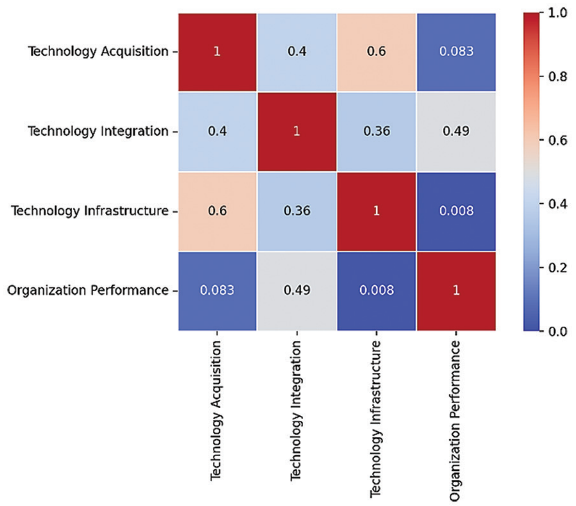


Fig. 4. Correlation heatmap of control variables

Understanding such relationships is important to assess how distinct technological elements relate to and impact organizational outcomes.

5.5. Analysis of Regression

The focus of this study on entrepreneurial innovation centers on technological systems, their acquisition, and the effects of these two aspects. The relationships between the dependent variable "Entrepreneurial Innovation" and the independent variables, e.g., "Technology Integration," "Technology Acquisition," and "Technology Infrastructure," were analyzed through regression analysis.

The findings reveal that "Technology Acquisition" has a coefficient of 1.278, with a corresponding $R^2 = 0.482$. This implies that 48.2% of the variation in entrepreneurial innovation is explained by "Technology Acquisition." A p=0.000 indicates a relationship that is both practically and statistically significant, suggesting that investment in technology by entrepreneurs positively affects innovation and business growth.

Likewise, "Technology Integration" recorded the second highest $R^2 = 0.556$, suggestive of an even greater impact on entrepreneurial innovation. Its statistical significance is supported by a p=0.000. This suggests that innovation through integrated technology is attainable, especially during business operations to improve efficiency, decision-making, and responsiveness to market competition.

For "Technology Infrastructure," the $R^2 = 0.369$ reflects moderate contemporary innovations affecting entrepreneurial innovation, and this relationship was found to be statistically significant, as indicated by the p=0.000. This emphasizes the need to develop sound technological infrastructure to support sustainable business growth and innovation in the long term.

To verify the reliability of these results, adjusted R^2 values were used to confirm that the regression model adequately captures the relationships between the entrepreneurial innovation and technological factors while minimizing the risk of overfitting.

Table 7 summarizes the results of the regression analysis, quantifying the association of "Technology Acquisition," "Technology Integration," and "Technology Infrastructure" with entrepreneurial innovation. The table presents essential regression parameters such as coefficient values, R^2, adjusted R^2, standard error, residual sum of squares, F-statistics, and p-values, which help assess the relevance of the model.

A technology integration coefficient of 0.632 (p=0.000) shows a strong positive relationship and

Table 7. Regression analysis between technology acquisition and entrepreneurial innovation

Model	C	R^2	Adj. R^2	Standard error	SS residual	F	Sign (F)
Technology Acquisition	1.278	0.48	0.48	0.38	46.2	298.17	0
Technology Integration	1.304	0.56	0.55	0.352	39.65	400.29	0
Technology Infrastructure	1.852	0.37	0.37	0.4195	56.31	187.18	0

Table 8. Regression analysis between entrepreneurial innovation and organizational performance

Factor	Unstandardized Coefficients (B)	Standard Error	Standardized Beta (β)	t-value	Significance (p-value)
Constant	1.482	0.108	-	13.743	0.000
Organizational Performance	0.583	0.030	0.732	19.240	0.000

is statistically significant at the 5% level. In addition, the adjusted R^2 = 0.554 suggests that 55.4% of the variance in the dependent variable is explained by the independent variables, further confirming the validity and reliability of the results.

The finding of a technology infrastructure coefficient of 0.472 (statistically significant at the 5% level, with a p=0.000) also indicates a positive correlation. The adjusted R^2 = 0.367 implies that over 36.7% of the variability in the dependent variable can be attributed to changes in the independent variables. These findings further enhance the credibility of the model used to determine the effect of technological factors on entrepreneurial innovation. Regression analysis results showed that "Technology Integration" had a significant positive effect on organizational performance (β = 0.482, p<0.01), while "Technology Infrastructure" demonstrated a moderate effect (β = 0.315, p<0.05). Detailed path coefficients and model fit indices are presented in Table 7.

Table 8 shows a sufficient positive correlation, with the Integrated Performance Coefficient standing at 0.583, significant at the 5% level with a p=0.000. From this, it can be concluded that the results demonstrate considerable strength, suggesting that 53.5% or more of the dependent variable can be explained, given the adjusted R^2 is at the aforementioned value. This further confirms the relationship between organizational performance and entrepreneurial innovation.

The findings of this study align with prior research emphasizing the critical role of technology in fostering entrepreneurial innovation and business growth. For instance, the strong positive impact of technology acquisition on organizational performance observed in this study corroborates the findings of Poudel et al. (2019), who highlighted technological capabilities as key prerequisites for business expansion and economic growth. Similarly, the importance of technology integration in driving entrepreneurial activities is consistent with the work of Lin et al. (2021) and Martin-Rojas et al. (2019), who demonstrated that technological capabilities significantly influence new venture development.

Furthermore, the results regarding the influence of AI-driven applications on business innovation align with the conclusions drawn by Jafari-Sadeghi et al. (2021), who found that digital transformation profoundly shapes entrepreneurial activities across Europe. In line with Matejun (2016), our study reinforces the view that technological entrepreneurship depends heavily on leveraging both internal resources and external technological advancements. Finally, the significance of technological infrastructure in enhancing organizational performance is supported by the findings of Sonika et al. (2023), who emphasized that strategic planning and technology usage are crucial drivers of firm competitiveness and productivity.

While it was anticipated that technology integration would have a strong and direct influence on entrepreneurial innovation, the magnitude of this effect in the present study was relatively moderate compared to expectations. This contrasts slightly with the assertions of Kheni (2017), who emphasized technology–innovation fusion as a defining trait of successful niche entrepreneurship. One potential explanation for this deviation could be the presence of organizational inertia or cultural resistance to technological change, which were not directly assessed in this study and could moderate the observed relationships.

6. Conclusion

6.1. Implications

The role of social technologies in knowledge acquisition appears to surpass their contribution to knowledge creation, indicating broader phenomena such as the use of outsourcing, the division of labor, and the increasing sophistication of technology markets. It is advisable for organizations to incorporate such advances in social network technologies and knowledge technologies to optimize intellectual capital value added throughout the social process control system, from innovation to social knowledge utilization. This can greatly improve organizational productivity and contribute to both economic and social development.

The integration of social networks with IT for knowledge management can be improved by examining the global best practices outlined in this study. Furthermore, these findings serve as an essential source of information to inform foster policy, specifically around enhancing social capital, increasing IT literacy skills, and strategically using social networking technologies at both the national and regional levels.

6.2. Limitations

This study is limited in scope largely due to the omission of several interacting factors concerning entrepreneurial attributes and attitudes that affect business performance. Wherever innovation and entrepreneurial zeal intersect, there is tremendous potential for IT infusion, and the resulting interceding variables invite more in-depth investigation in the long run.

Furthermore, a comparative study with entrepreneurs from Rajasthan would aid in understanding the operational variations among entrepreneurs in different geographical regions with common entrepreneurial attributes. This would open new horizons in examining the role of technological innovations in entrepreneurial ingenuity and performance in the developing world.

While the present study provides valuable insights into the influence of mobile multimedia technologies and AI-driven applications on entrepreneurial innovation and sustainable business growth, its findings are based solely on data collected from technology firms operating in India. As such, caution should be exercised when generalizing the results to other geographical regions or industrial sectors, where technological adoption patterns, organizational cultures, and market conditions may differ. Future research could extend the model to diverse country contexts and sectors, enabling cross-comparative validation and broader applicability of the results.

6.3. Principal takeaways

The adoption and use of mobile multimedia technologies have transformed entrepreneurial success by automating processes, improving communication, and enabling better decision-making, thereby altering traditional business models. While these technologies present both advantages and disadvantages, their overall effect on entrepreneurial innovation and sustainable business model development is positive, especially in the areas of efficiency and growth potential.

This research confirms that all hypotheses proposed in this study were statistically supported. The results show that the processes of technology acquisition, integration, and infrastructural development significantly enhance entrepreneurial innovativeness, which results in improved organizational performance.

Mobile multimedia platforms improve business effectiveness by automating customer interactions, enhancing decision-making, and increasing labor productivity and revenue flow. Furthermore, their ability to automate processes further enhances operational efficiency and business growth. This study suggests that the use of mobile multimedia technologies facilitates entrepreneurial development through more effective technology adoption and integration, infrastructure improvements, and increased organizational performance.

Furthermore, mobile multimedia platforms support entrepreneurship by expanding job opportunities, facilitating learning, and boosting employment through remote work capabilities. Specifically, mobile multimedia communication systems and electronic marketing technologies empower entrepreneurs in the development of innovative and competitive business models within fast-evolving business and technology environments.

Future research could explore the extension of the proposed framework to non-technology sectors, such as healthcare, agriculture, and retail, to assess its broader applicability. Comparative cross-country studies involving both emerging and developed economies may provide insights into regional variations in technology adoption and entrepreneurial innovation. Moreover, incorporating moderating variables such as organizational culture, leadership dynamics, and external market turbulence could yield a more nuanced understanding of the technology–entrepreneurship relationship. Longitudinal studies would also be valuable to examine how technological impact on entrepreneurial performance evolves over time.

References

Abubakre, M., Zhou, Y., & Zhou, Z. (2022). The impact of information technology culture and personal innovativeness in information technology on digital entrepreneurship success. *Information*

Technology and People, 35(1), 204–231. https://doi.org/10.1108/ITP-01-2020-0002

Amouri, A., Festa, G., Shams, S.M.R., Sakka, G., & Rossi, M. (2021). Technological propensity, financial constraints, and entrepreneurial limits in young entrepreneurs' social business enterprises: The tunisian experience. *Technological Forecasting and Social Change*, 173, 121126. https://doi.org/10.1016/j.techfore.2021.121126

Ashton, W. B., & Stacey, G. (1995). Technical intelligence in business: understanding technology threats and opportunities. *International Journal of Technology Management,* 10(1), pp 79-104 https://doi.org/10.1504/IJTM.1995.025615

Chege, S.M., & Wang, D. (2020). Information technology innovation and its impact on job creation by SMEs in developing countries: an analysis of the literature review. *Technology Analysis and Strategic Management*, 32(3), 256–271. https://doi.org/10.1080/09537325.2019.1651263

Jafari-Sadeghi, V., Garcia-Perez, A., Candelo, E., & Couturier, J. (2021). Exploring the impact of digital transformation on technology entrepreneurship and technological market expansion: The role of technology readiness, exploration and exploitation. *Journal of Business Research*, 124, 100–111. https://doi.org/10.1016/j.jbusres.2020.11.020

Jiao, H., Yang, D., Gao, M., Xie, P., & Wu, Y. (2016). Entrepreneurial ability and technological innovation: Evidence from publicly listed companies in an emerging economy. *Technological Forecasting and Social Change*, 112, 164–170. https://doi.org/10.1016/j.techfore.2016.08.003

Kheni, S. (2017). Role of Innovation in Entrepreneurship Development. *International Journal of Management and Social Sciences Research*, 3(2), pp 1-5.

Kilintzis, P., Avlogiaris, G., Samara, E., & Bakouros, Y. (2023). Technology entrepreneurship: A model for the European Case. *Journal of the Knowledge Economy*, 14(2), 879–904. https://doi.org/10.1007/s13132-022-00950-x

Kumar, N., Upreti, K., & Mohan, D. (2022). Blockchain adoption for provenance and traceability in the retail food supply chain. *International Journal of E-Business Research*, 18(2), 1–17. https://doi.org/10.4018/IJEBR.294110

Kumar, N., Upreti, K., Jafri, S., Arora, I., Bhardwaj, R., Phogat, M., et al. (2022). Sustainable computing: A determinant of industry 4.0 for sustainable information society. *Journal of Nanomaterials*, 2022(1), 1–10.

Lee, J., Suh, T., Roy, D., & Baucus, M. (2019). Emerging technology and business model innovation: The case of artificial intelligence. *Journal of Open Innovation: Technology, Market, and Complexity*, 5(3), 44. https://doi.org/10.3390/joitmc5030044

Leischnig, A., Wölfl, S., Ivens, B., & Hein, D.W.E. (2017). From digital business strategy to market performance: Insights into key concepts and processes. In: *Proceedings of the 38th International Conference on Information Systems*.

Li, T. (Carol), & Chan, Y.E. (2019). Dynamic information technology capability: Concept definition and framework development. *The Journal of Strategic Information Systems*, 28(4), 101575. https://doi.org/10.1016/j.jsis.2019.101575

Lin, W.T., Chen, Y.H., & Chou, C.C. (2021). Assessing the business values of e-commerce and information technology separately and jointly and their impacts upon US firms' performance as measured by productive efficiency. *International Journal of Production Economics*, 241, 108269. https://doi.org/10.1016/j.ijpe.2021.108269

Maoning, L. . (2022). Research on Innovation and Entrepreneurship Approach in Universities Based on Large Data: Innovation and Entrepreneurship. *Journal of Mobile Multimedia*, 19(02), 547–566. https://doi.org/10.13052/jmm1550-4646.1929

Martínez-Caro, E., Cegarra-Navarro, J.G., & Alfonso-Ruiz, F.J. (2020). Digital technologies and firm performance: The role of digital organisational culture. *Technological Forecasting and Social Change*, 154, 119962. https://doi.org/10.1016/j.techfore.2020.119962

Martin-Rojas, R., Garcia-Morales, V.J., & Gonzalez-Alvarez, N. (2019). Technological antecedents of entrepreneurship and its consequences for organizational performance. *Technological Forecasting and Social Change*, 147, 22–35. https://doi.org/10.1016/j.techfore.2019.06.018

Matejun, M. (2016). Role of technology entrepreneurship in the development of innovativeness of small and medium-sized enterprises. *Management*, 20(1), 167–183. https://doi.org/10.1515/manment-2015-0032

Obschonka, M., & Audretsch, D.B. (2020). Artificial intelligence and big data in entrepreneurship: A new era has begun. *Small Business Economics*, 55(3), 529–539. https://doi.org/10.1007/s11187-019-00202-4

Oyero, M.A., & Oyedele, O.O. (2022). Techno-entrepreneurship - pathway to sustainable business performance: Empirical evidence from SMEs in Ogun state, Nigeria. *World Review of Entrepreneurship, Management and Sustainable Development*, 18(5/6), 545. https://doi.org/10.1504/WREMSD.2022.10049636

Park, Y., & Mithas, S. (2020). Organized complexity of digital business strategy: A configurational perspective. *MIS Quarterly*, 44(1), 85–127. https://doi.org/10.25300/MISQ/2020/14477

Poudel, K.P., Carter, R., & Lonial, S. (2019). The impact of entrepreneurial orientation, technological capability, and consumer attitude on firm performance: A multi□theory perspective. *Journal of Small Business Management*, 57(sup2), 268–295. https://doi.org/10.1111/jsbm.12471

Psomadaki, O., Matsiola, M., Dimoulas, C.A., & Kalliris, G.M. (2022). The significance of digital network platforms to enforce musicians' entrepreneurial role: Assessing musicians' satisfaction in using mobile applications. *Sustainability*, 14(10), 5975. https://doi.org/10.3390/su14105975

Quinn, B.L., Seibold, E., & Hayman, L. (2015). Pain assessment in children with special needs. *Exceptional Children*, 82(1), 44–57. https://doi.org/10.1177/0014402915585480

Rajesh, M. ., Vengatesan, K. ., Sitharthan, R. ., Dhanabalan, S. S. ., & Gawali, M. B. . (2023). Enhancing Mobile Multimedia Trustworthiness through Federated AI-based Content Authentication: Enhancing Mobile Multimedia. *Journal of Mobile Multimedia*, 19(06), 1415–1438. https://doi.org/10.13052/jmm1550-4646.1963

Rehman, N., Razaq, S., Farooq, A., Zohaib, N.M., & Nazri, M. (2020). Information technology and firm performance: Mediation role of absorptive capacity and corporate entrepreneurship in manufacturing SMEs. *Technology Analysis and Strategic Management*, 32(9), 1049–1065. https://doi.org/10.1080/09537325.2020.1740192

Roemintoyo, R., Miyono, N., Murniati, N.A.N., & Budiarto, M.K. (2022). Optimising the utilisation of computer-based technology through interactive multimedia for entrepreneurship learning. *Cypriot Journal of Educational Sciences*, 17(1), 105–119. https://doi.org/10.18844/cjes.v17i1.6686

Romano, A., Passiante, G., & Del Vecchio, P. (2016). The technology-driven entrepreneurship in the knowledge economy. In: *Creating Technology-Driven Entrepreneurship*. Palgrave Macmillan, United Kingdom. p21–48. https://doi.org/10.1057/978-1-137-59156-2_2

Shah, S.F.H., Nazir, T., Zaman, K., & Shabir, M. (2013). Factors affecting the growth of enterprises: A survey of the literature from the perspective of small- and medium-sized enterprises. *Journal of Enterprise Transformation*, 3(2), 53–75. https://doi.org/10.1080/19488289.2011.650282

Shinde, D.D., & Prasad, R. (2021). Total productive education: Model for higher technical education. *Journal of Mobile Multimedia*, 17(1–3), 1–26. https://doi.org/10.13052/jmm1550-4646.17131

Sihite, B.J., & Prihandini, A. (2019). Information technology in supporting education world to become an entrepreneur. *IOP Conference Series: Materials Science and Engineering*, 662(3), 032039. https://doi.org/10.1088/1757-899X/662/3/032039

Slyusar, V., Kondratenko, Y., Shevchenko, A., & Yeroshenko, T. (2024). Some aspects of artificial intelligence development strategy for mobile technologies. *Journal of Mobile Multimedia*, 20, 525–554. https://doi.org/10.13052/jmm1550-4646.2031

Snihur, Y., Lamine, W., & Wright, M. (2021). Educating engineers to develop new business models: Exploiting entrepreneurial opportunities in technology-based firms. *Technological Forecasting and Social Change*, 164, 119518. https://doi.org/10.1016/j.techfore.2018.11.011

Sonika, B., Dayal, M., Singh, D., Upreti, S., Upreti, K., & Kumar, J. (2023). Block-hash signature (BHS) for transaction validation in smart contracts for security and privacy using blockchain. *Journal of Mobile Multimedia*, 19, 935–962.

Steininger, D.M. (2019). Linking information systems and entrepreneurship: A review and agenda for IT□associated and digital entrepreneurship research. *Information Systems Journal*, 29(2), 363–407. https://doi.org/10.1111/isj.12206

Uhlenbruck, K., Hitt, M.A., & Semadeni, M. (2006). Market value effects of acquisitions involving internet firms: A resource-based analysis. *Strategic Management Journal*, 27(10), 899-913.

Usai, A., Fiano, F., Messeni Petruzzelli, A., Paoloni, P., Farina Briamonte, M., & Orlando, B. (2021). Unveiling the impact of the adoption of digital technologies on firms' innovation performance. *Journal of Business Research*, 133, 327–336. https://doi.org/10.1016/j.jbusres.2021.04.035

AUTHOR BIOGRAPHIES

Dr. Samreen Jafri is an accomplished and multifaceted professional with an illustrious career spanning over 17 years in academia, research, and curriculum development. As an editor, she embodies intellectual rigor, meticulous attention to detail, and a passion for advancing knowledge in her field. Her career reflects a deep commitment to fostering innovation and excellence in human resource and supply chain management. Dr. Jafri's collaborative nature, coupled with her intellectual curiosity, makes her an invaluable contributor to multidisciplinary projects. Holding a Ph.D. in Supply Chain Management and backed by a robust academic foundation, Dr. Jafri has consistently pushed the boundaries of her discipline. A prolific researcher, she has published extensively in SCOPUS-indexed journals and other prestigious venues, with expertise spanning topics from cyber security to organizational behavior. Her editorial acumen is evident in her role as guest editor for SUSCOM-2024's special issue on sustainable security, when she adeptly curates influential research and upholds the highest standards of publication quality. Dr. Jafri's leadership, as an editor, researcher, and educator, is driven by her commitment to building a more sustainable and interconnected academic and professional community.

Dr. Kamal Upreti is currently serving as an Associate Professor and Research Coordinator in the Department of Computer Science, CHRIST University, Delhi-NCR, Ghaziabad, Uttar Pradesh. He completed his B. Tech (Hons) from UPTU, M. Tech (Gold Medalist), PGDM (Executive) from IMT Ghaziabad, and a PhD in Computer Science and Engineering. He also completed his postdoctoral research at the National Taipei University of Business, Taiwan, funded by the MHRD. In 2024, Dr. Upreti was recognized among the top 2% of scientists globally by Stanford University, California. He has published over 50 patents, 45 books, 32 magazine issues, and 130 research papers in reputed journals and international conferences. His areas of interest include artificial intelligence, machine learning, data analytics, cybersecurity, machine learning, healthcare, embedded systems, and cloud computing. He has authored and edited more than 45 books with renowned publishers such as CRC Press, IGI Global, Oxford Press, and Arihant Publication.

Dr. Rishu Bhardwaj has been working as an Assistant Professor at Chitkara Business School, Chitkara University, since January 2022. She has over 15 years of experience as a faculty member and trainer in International Business & Trade, Economics, and Entrepreneurship. Her expertise includes Business Economics, International Trade & Export-Import Management, and Entrepreneurship Development. She has worked with various reputed business schools, including NIILM New Delhi and DIT Business School. Notably, she has collaborated with IED-UP, where she delivered expert and dynamic sessions on Import-Export Procedures & Planning, statutory requirements to start EXIM businesses, venture capitalization, and new business strategies. Dr. Bhardwaj has contributed to the "One District One Product" scheme under the Uttar Pradesh Government. She has conducted workshops at various esteemed institutions, including MCM DAV College (Punjab University, Chandigarh), Raipur; MET Mumbai; Matrusri Engineering College, Hyderabad; Hindu College, Amritsar; Dronacharya College of Engineering, Gurgaon; KCG Chennai; BMI, Bangalore; Raja Rajeshwar College, Bangalore; Apeejay; IIITM, Gwalior; ITM University; Acropolis Engineering College, Indore; Daulat Ram College, Delhi; and Kalandi College, Delhi. She holds three published patents, two copyrights, and has authored various Scopus-indexed papers, as well as numerous UGC Care and IEEE conference papers.

Dr. Balraj Verma is an Associate Professor at the Doctoral Research Centre, Chitkara Business School, Chitkara University, Punjab. He earned his Ph.D. from Jaypee University of Information Technology (JUIT), Waknaghat, and holds an MBA with a strong foundation in business strategy and marketing. With over 17 years of combined academic and industry experience, he has developed a diverse academic portfolio characterized by scholarly rigor, practical relevance, and a commitment to educational excellence.

Dr. Verma has taught a wide range of subjects, including Marketing Management, Strategic Management, Business Statistics, and Research Methodology. He has successfully supervised multiple doctoral scholars to completion and continues to mentor Ph.D. candidates across interdisciplinary domains. His scholarly output includes numerous articles published in high-impact journals indexed in SCI, Scopus, and ABDC, as well as edited books and book chapters with leading international publishers. His research interests are interdisciplinary and evolving, focusing primarily on E-commerce, retail and aggregator business models, technology adoption in emerging markets, blockchain for supply chains, robo-advisory services, and the behavioral aspects of digital transformation.

Rituraj Jain is currently serving in the Department of Information Technology at Marwadi University, Rajkot, India. He has an extensive academic and research background, with numerous publications in reputed international journals and conferences. Additionally, he is the author of four published books in the field of computer science and information technology. He holds professional certification as an Oracle Database SQL Certified Expert and has completed several specialized certifications through online platforms, including Hadoop Fundamentals I – Version 2, Moving Data into Hadoop, Introduction to MapReduce Programming, Big Data Fundamentals, and the Cloud U Certification. Mr. Jain is an active member of several prestigious professional bodies, such as the Indian Society for Technical Education (ISTE), Computer Science Teachers Association (CSTA), Academy & Industry Research Collaboration Center (AIRCC), International Association of Computer Science and Information Technology (IACSIT), and the International Association of Engineers (IAENG). His research interests span a wide range of emerging technologies, including cloud computing, machine learning, deep learning, fuzzy logic, Internet of Things, genetic algorithms, big data, and software engineering.

DOI: 10.6977/IJoSI.202508_9(4).0006
V. N. Gaikwad, R. K. Budania/Int. J. Systematic Innovation, 9(4), 71-83 (2025)

Decoding Marathi emotions: Enhanced speech emotion recognition through deep belief network-support vector machine integration

Varsha Nilesh Gaikwad[1]*, Rahul Kumar Budania[2]

[1]Department of Electronics and Telecommunication Engineering, School of Engineering, RMD Sinhgad Technical Institute, Pune, Maharashtra, India

[2]Department of Electronics and Communication Engineering, Institute of Engineering, Shri JJT University, Jhunjhunu, Rajasthan, India

*Corresponding author E-mail: nvarsha29619@gmail.com

(Received 29 October 2024; Final version received 9 December 2024; Accepted 12 December 2024)

Abstract

Speech emotion recognition in Marathi presents considerable hurdles due to the language's distinct grammatical and emotional characteristics. This paper presents a robust methodology for classifying emotions in Marathi speech utilizing advanced signal processing, feature extraction, and machine learning techniques. The method entails collecting diverse Marathi speech samples and using pre-processing steps such as pre-emphasis and voice activity detection to improve signal quality. Speech signals are segmented using the Hamming window to reduce discontinuities, and features such as Mel-frequency cepstral coefficients, pitch, intensity, and spectral properties are retrieved. For classification, an attentive deep belief network is paired with a support vector machine, which uses attention techniques and batch normalization to improve performance and reduce overfitting. The suggested approach surpasses existing models, with 98% accuracy, 98% F1-score, 99% specificity, 99% sensitivity, 98% precision, and 98% recall.

Keywords: Speech Emotion Recognition, Voice Activity Detection, Mel-Frequency Cepstral Coefficient, Deep Belief Network, Support Vector Machine

1. Introduction

User interfaces are growing more complicated, with voice processing technology allowing users to communicate without physically using a keyboard (Chaudhari et al., 2023). Speech is an important type of human-to-human communication that provides emotional and psychological information. Speech processing provides sound qualities and characteristics that can be used to extract meaningful information (Papala et al., 2023). Speech emotion classification is not only at the core of human life and action, as most scientific and psychiatric endeavors have shown, but it can also be studied using the computing tools of today's modernist conception of science. One unanswered topic, nevertheless, is how the application of machine learning techniques to the study of the typical, human, and empirically observed dynamics of emotion classification has changed, evolved, or expanded in scope (Akinpelu & Viriri, 2024). However, identifying emotions from speech remains challenging due to the range of expressions, even for the same feeling (Lieskovská et al., 2021). Joy, fury, fear, and sorrow have similar acoustic features, such as voice volume, pitch, and the number of times their speech meets the zero axis (Madanian et al., 2023). This issue stems from the recognition of these two sets of emotions, which we extract directly from speech signals or text, and the feature set used for emotion detection (Hammed & George, 2023). Acoustic elements of speech, such as pitch, intensity, and volume, can also be deceptive when considered alone (Kaur & Singh, 2023). People employ speech signal features and speech semantics to communicate their emotions in everyday situations, emphasizing the significance of extracting emotions from both acoustic and semantic variables before concluding the underlying emotions in a speech signal (Zaidi et al., 2023).

Artificial intelligence (AI) advancements have improved the comfort and convenience of human-computer contact (Yang et al., 2024). The next wave of AI development will focus on enhancing speech emotion recognition (SER), which has both theoretical and practical ramifications (Harhare & Shah, 2021). Feature extraction is critical in speech signal processing, and hand-designed features have been used for SER (Bachate et al., 2022). The spectral feature, which considers both the frequency and time axes, has gained prominence in recent years. There are several challenges with the traditional method of identifying emotion from speech utterances. Many of the current methods, including the support vector machine (SVM), hidden Markov model, and Gaussian mixture model, rely on automated speech recognition, which is highly dependent on dataset manipulation. Any changes may necessitate reconstructing the entire model. It is impossible to categorize emotion lightly because it contains important information that has the power to either make or break a person's personality. To avoid some of these issues with the conventional method (Akçay & Oğuz, 2020). However, these manual abilities are limited and cannot adequately portray emotions in speech. To solve this, neural networks have proposed a solution that incorporates deep learning into the model-building process (Alam Monisha & Sultana, 2022).

Deep learning features extract specialized feature representations from big learning problems, which reduce the incompleteness caused by artificially created features (Padman & Magare, 2022). The standardized pre-trained (Chai et al., 2021) model is typically used to address the issue of the inadequate training dataset in transfer learning, a fundamental area of deep learning that has demonstrated effectiveness in a variety of computer vision-related applications, including emotion identification (Li et al., 2021). It is a deep convolutional neural network (DCNN) subdivision. Due to its inherent capacity to extract speech features from speech signals distinctively and efficiently, the DCNN application to emotion categorization gained prominence (Oh & Kim, 2022). Researchers are constantly on the lookout for new DCNN techniques (Byun & Lee, 2021) that can produce more noticeable results, but the current findings have exposed the long-standing issues of inadequate label datasets for the classification of speech emotion and a high level of parameterization of the field. As a result, there is a need to develop a dependable method for automatic speech detection.

2. Literature Survey

Section 2 presents an overview of the current research in SER, summarizing and discussing key literature. SER is critical for understanding human emotional behavior and relies on identifying distinguishing traits. Alluhaidan et al. (2023) enhanced SER system performance by combining Mel-frequency cepstral coefficients (MFCCs) and time-domain features. To create the SER model, a convolutional neural network (CNN) was fed the suggested hybrid features. The current work limits the acquisition of high-level acoustic information crucial for accuracy because it does not compare SER approaches across datasets and does not include recurrent neural networks. Kawade & Jagtap (2024) employed a DCNN and multiple acoustic features and proposed a cross-corpus SER (CCSER) for an Indian corpus. For feature selection, Fire Hawk-based optimization reduces computational complexity and enhances feature distinctiveness. Better correlation, greater feature representation, and a more accurate description of the speech signal's timbre, intonation, and pitch variation are all provided by the DCNN method. However, its capacity to generalize across a variety of emotional circumstances is diminished by its lack of domain adaptability and inadequate global and local acoustic properties. Bhangale & Kothandaraman (2023) displayed the acoustic feature set using the following methods to increase the feature distinctiveness: MFCC, linear prediction cepstral coefficients (LPCCs), wavelet packet transform (WPT), zero crossing rate (ZCR), spectrum centroid, spectral roll-off, spectral kurtosis, root mean square (RMS), pitch, jitter, and shimmer. In addition, a lightweight, compact one-dimensional DCNN is employed to capture the spoken emotion signal's long-term relationships and reduce computational complexity. The system is not resilient under CCSER under different noise settings and suffers from class imbalance as a result of unequal dataset training. Farooq et al. (2020) conducted a project to improve SER by employing a DCNN to classify emotions during human-machine interaction accurately. The DCNN extracts features from complex speech-emotional datasets using a correlation-based feature selection method. The approach obtains 95.10% accuracy in speaker-dependent SER tests with four publicly available datasets: Emo-DB, SAVEE, IEMOCAP, and RAVDESS. This is especially essential in audio conferencing, where traditional machine learning methods are less trustworthy due to noise sensitivity and accent fluctuations. Sonawane & Kulkarni (2020) used a deep learning strategy for emotion speech detection that employs a multilayer CNN and a simple K-nearest neighbor classifier, which has been shown to outperform the current MFCC method in real-time testing on the YouTube database. This technology is critical for SER in real-time applications such as human behavior assessment, human–robot interaction, virtual reality, and emergency rooms. Sajjad & Kwon (2020) created a new framework for SER that selects sequence segments based on cluster-level similarity measures. The short-time Fourier transform technique transforms the sequence into a spectrogram, which is then input into a CNN model for feature extraction. The CNN features are normalized for accurate recognition and input into bidirectional long short-term memory for emotion recognition. The system is evaluated using typical datasets to enhance recognition accuracy and processing time.

The contributions of this work are as follows:

- The study focuses on creating a diverse dataset of Marathi voice samples to capture the unique grammatical and emotional characteristics of the Marathi language.
- It employs advanced feature extraction techniques, including pre-emphasis; voice activity detection (VAD); and acoustic, prosodic, and spectral features, to capture nuanced emotional traits in Marathi speech.
- The study also develops a hybrid classification model combining a deep belief network (DBN) and an SVM, enhancing feature representation and emotion recognition accuracy.

3. Proposed Methodology

SER in Marathi is a significant difficulty due to the language's unique grammatical and emotional characteristics. To address these challenges, a complex methodology was developed for improving emotion classification accuracy in Marathi speech samples. The dataset compilation included specifics about the Marathi voice samples, such as their source, size, emotional categories, and demographic diversity, alongside the procedures for collection and annotation. For pre-processing, the exact parameters for pre-emphasis filters improved the speech signal by removing low-frequency noise. VAD was then utilized to determine the beginning and conclusion of speech by combining temporal and frequency domain approaches. The speech signal was divided into smaller frames using a Hamming window to reduce disruptions at frame boundaries, which must be clearly outlined. Feature extraction processes detailed the computation of MFCCs, including the number of coefficients and window settings, as well as the calculation methods for prosodic features such as pitch, intensity, and duration, and spectral features such as spectral centroid and zero-crossing rate, spectral bandwidth, and spectral roll-off, provide additional information about the frequency distribution and periodicity of the speech signal. The model design specified the architecture of the attentive DBN, including layer configurations, hidden units, activation functions, and the attention mechanism, as well as the kernel type and hyperparameters for the SVM. Regularization techniques, such as batch normalization, included parameter settings, and the training process were described in terms of optimizers, learning rate, epochs, and stopping criteria. Evaluation methods outlined the performance metrics, validation strategies, and comparison benchmarks. This methodology aims to overcome the limits of existing SER methods by increasing the feature representation of Marathi speech and employing modern machine learning algorithms for more accurate emotion recognition (Fig. 1).

3.1. Pre-emphasis

A pre-emphasis filter was used to boost the high-frequency components of an audio signal. This step was essential for capturing the characteristics of the input speech samples. In our method, we started by eliminating noise from the input samples, which helped with feature extraction.

3.2. VAD

The VAD is a technique for determining whether speech is present in each frame of a noisy signal. It consists of two processing stages: gathering features from noisy signals to discriminate between speech and noise and applying a detection approach to these data. This article examines the extraction of features and the performance of VAD algorithms (Fig. 2). Speech detection has poor temporal resolution since it is usually divided into shorter frames rather than deciding for each sample: Equations (1-3).

$$xl = [x(lL - N + 1), \ldots, x\,(lL - 1), x\,(lL)\,]^2 \tag{1}$$

$$H_1: x(l) = b(l) + s(l) \tag{2}$$

$$H_0: x(l) = b(l) \tag{3}$$

The noisy frame can be assumed to be a combination of speech components (s [L] and noise (b [L]) or simply noise. The decision for one hypothesis, Equations (4 and 5).

$$VAD_{ftr}\left(n,l\right) = \begin{cases} 1, \textit{When } H_1 \textit{ is accepted} \\ 0 \textit{ When } H_0 \textit{ is accepted} \end{cases} \tag{4}$$

$$VAD_{ftr}\left(n,l\right) = \begin{cases} 1, \textit{where } ftr(x(l) > n \\ 0, \textit{where } ftr(x(l) \le n \end{cases} \tag{5}$$

3.3. Acoustic Feature

Acoustic aspects of a voice signal describe its physical characteristics in terms of frequency,

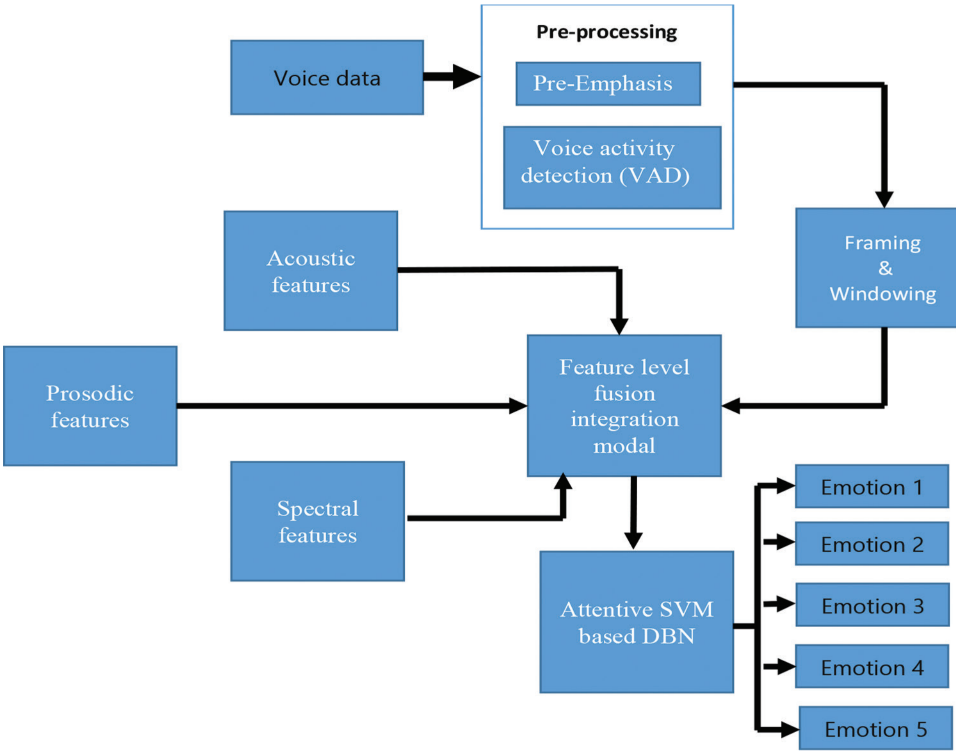


Fig. 1. Block diagram of the proposed methodology
Abbreviations: DBN: Deep belief network; SVM: Support vector machine

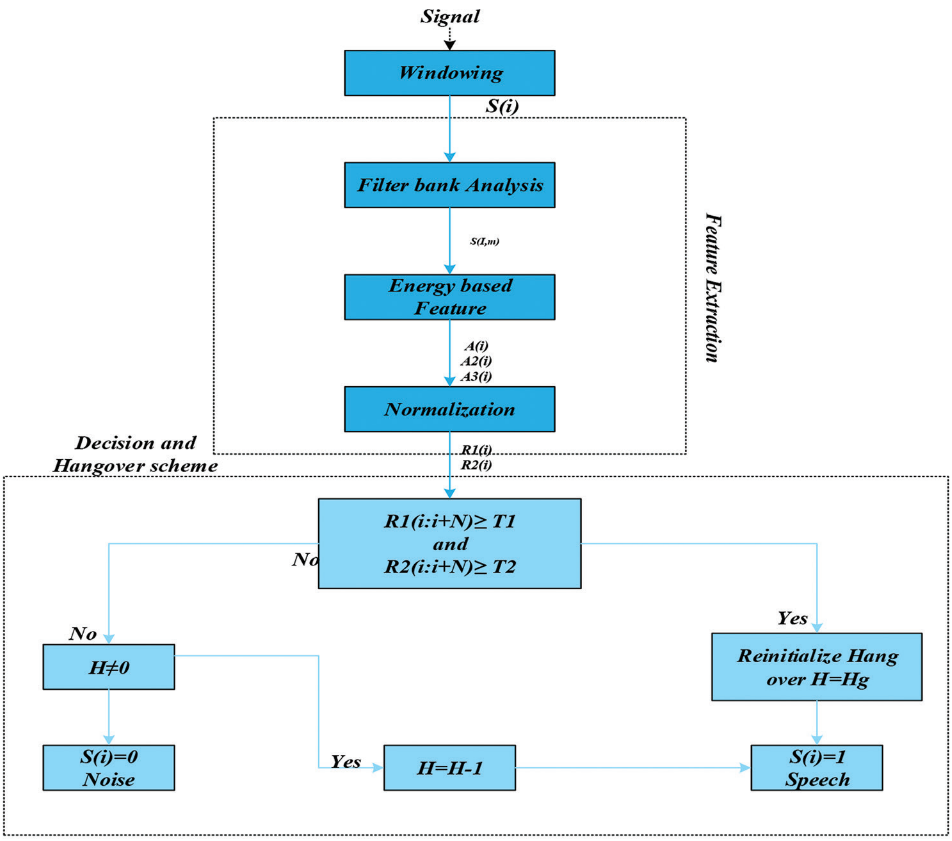


Fig. 2. Block diagram of voice activity detection

amplitude, and volume. The proposed acoustic feature set consisted of various spectral features, time-domain features, and voice quality factors that describe speech emotion. The acoustics features extracted are MFCC, LPCC, WPT, ZCR, RMS, SK, jitter, shimmer, pitch frequency, formants, and their mean and standard deviation. To eliminate noise and disturbances, the speech stream was sent through a moving average filter before being processed into various properties.

3.3.1. MFCC feature extraction

The MFCC is a technique for extracting spectral data regarding speech and human hearing perception (Abdel-Hamid et al., 2020; Shah et al., 2021). It entails normalizing the emphasis, removing noise and disturbances in raw emotional speech, and dividing the signal into 40-ms frames with a 50% frameshift (Fig. 3). For four-second voice signals, 199 frames were generated, each with a 40-ms frame width and 50% overlapping. Equation (6) demonstrates that the nearest frequency components are combined with a single Hamming window and a sample duration of 30 ms.

$$H(n) = (1-\alpha) - \alpha \times cos\frac{2\pi n}{(N-1)}, 0 \le n \le N-1 \tag{6}$$

The discrete Fourier transform converts time-domain emotion speech data into frequency-domain counterparts, revealing vocal tract characteristics. The signal was processed using Mel-frequency triangular filter banks, which provided perceptual information for speech hearing. Equations (10 and 11), which convert linear to Mel frequency and vice versa, ensure appropriate interpretation of speech-hearing perceptual information.

$$X(K) = \sum_{n=0}^{N-1} x(n) \times H(n) \times e^{-\frac{j2\pi nk}{N}}, 0 \le n, k \le N-1 \tag{7}$$

$$X_k = \frac{1}{N} | X(K) |^2 \tag{8}$$

$$ET_m = \sum_{k=0}^{k=1} \nabla_m(k) \times X_k; m = 1, 2, \dots M \tag{9}$$

$$Mel = 2595 \log 1 + \frac{f}{700} \tag{10}$$

$$f = 7010^{\frac{Mel}{2595}} - 1 \tag{11}$$

The discrete cosine transform of the log-filter bank energy signal yields an L number of cepstral coefficients, as shown in Equation (12).

$$MFCC_i = \sum_{m=1}^{M} log_{10}(ET_m) \times cosj(m+0.5)\frac{\pi}{m} \quad \text{For } j=1, 2 \dots L \tag{12}$$

According to earlier research, the MFCC contains 39 variables, including the speech signal's energy, 12 coefficients, and 26 derivatives, all of which are critical for distinguishing emotional speech shifts (Er, 2020; Kishor & Mohanaprasad, 2022).

3.4. Prosodic Feature

Prosodic features are components of speech that extend beyond phonetic segments, including intonation, stress, rhythm, and tempo. These qualities are essential for conveying meaning, emotion, and intent in spoken language. Understanding and analyzing prosodic features can help improve speech recognition systems, natural language processing, and communication interfaces. Prosodic features such as pitch, intensity, and duration are crucial for analyzing and interpreting vocal characteristics, especially in the context of emotional expression.

To calculate the pitch period (T_0), the following formula in Equation (13) was used,

$$T_o = \frac{1}{F_0} \tag{13}$$

Intensity assesses the loudness or vigor of the voice. Variations in intensity can reflect emotional states, with higher intensity frequently associated with anger or enthusiasm and lower intensity with melancholy or peacefulness (Equation [14]).

$$RMS = \sqrt{\frac{1}{N}\sum_{N=1}^{N} x[i^2]} \tag{14}$$

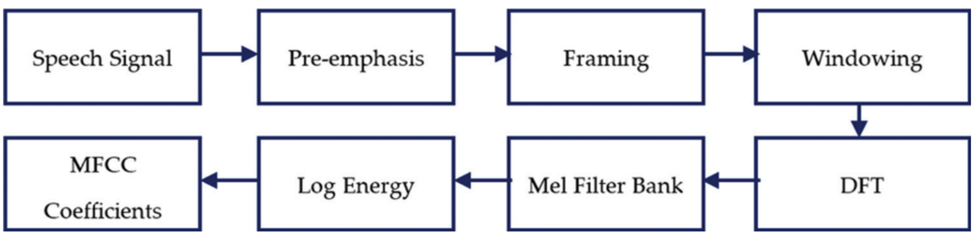


Fig. 3. Process flow of Mel-frequency cepstral coefficient feature extraction
Abbreviation: DFT: Discrete Fourier transform

Duration refers to the length of time a sound is held. The duration of spoken words and the gaps between them can reveal information about a speaker's emotional state. For example, lengthier durations and pauses may imply reluctance or reflection, whereas shorter durations may show hurry or excitement (Equation [15]).

$$Duration = N \times T_s \tag{15}$$

Where $T_N = \dfrac{1}{f_s}$ and windowing f_s is the sampling frequency.

3.5. Spectral Features

They are important in audio signal processing because they provide a complete picture of the signal's properties. Each of these characteristics contains important information about the signal's frequency distribution and periodicity, resulting in a more complete and detailed feature set.

A higher spectral centroid usually indicates a brighter sound, which is common in speech and music analysis. For example, in voice processing, the spectral centroid can aid in discriminating between distinct phonemes and emotional tones.

Spectral bandwidth measures the width of the spectrum and offers information about the range of frequencies in the signal. This function is especially important for determining the timbral properties of sound. A broader bandwidth typically suggests a richer and more complex sound, which can be critical in music genre classification and speaker recognition.

The spectral roll-off is useful for distinguishing between percussive and harmonic sounds. In music, it can aid in instrument detection, while in speech processing, it can help distinguish between voiced and unvoiced speech parts.

The ZCR is the rate at which a signal changes sign from positive to negative, or vice versa. It is a simple but powerful function for detecting the noise and roughness of an audio source. High ZCR readings frequently indicate the presence of high-frequency components, which are found in fricative sounds in speech and some musical instruments such as cymbals.

Incorporating these spectral properties into audio analytic frameworks can improve the accuracy and resilience of a wide range of applications, including speech recognition, music information retrieval, and audio categorization. These elements enabled a more comprehensive comprehension of the audio content by collecting specific information about the signal's frequency distribution and periodicity.

3.6. DBN

The DBNs are classified and feature extracted using restricted Boltzmann machines with two visible and hidden layers (Li et al., 2022). These layers are connected by weights, but nodes within the same layer are not. Backpropagation was used to train DBNs, as it is for all multilayer neural networks. In the first step, input data were used to forward-propagate restricted Boltzmann machines in each layer, and high-level abstractions were constructed by translating feature vectors to different feature spaces. V0 acted as both the first visible and input layer, whereas parameter W0 was learned from training data to rebuild the hidden and second visible layers (Arul, 2021) (Fig. 4).

A DBN is useful for estimating the posterior probability of a given feature vector. The parameters of DBN are W, b, and c. The probability of input vector v and output vector h is given below in Equation (16):

$$p(v,h) = \frac{e^{-E(v,h)}}{Z} \tag{16}$$

Where E (v, h) is the energy function (Equation [17]):

$$E(v, h) = -b^T v - c^T h - h^T Wv \tag{17}$$

The normalizing factor, Z, was calculated by adding the numerator of (16) to all conceivable h and v statuses (Equation [18]).

$$Z = \sum_{v,h} e^{-E(v,h)} \tag{18}$$

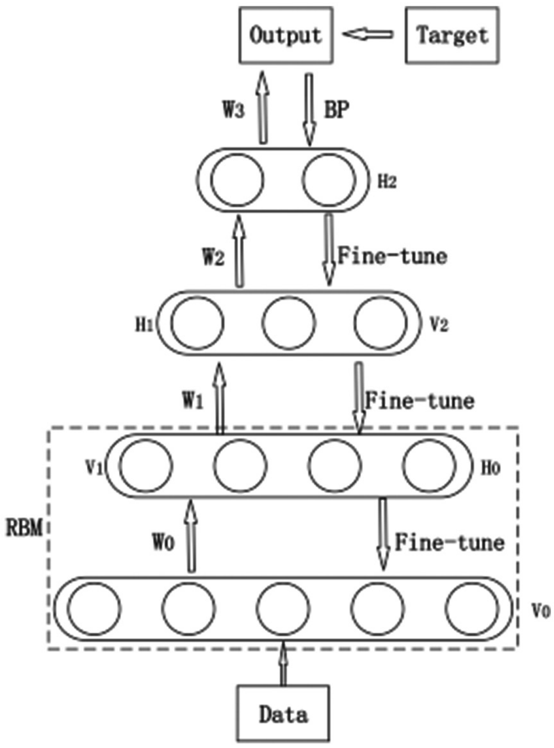


Fig. 4. Structure of a deep belief network
Abbreviation: RBM: Restricted Boltzmann machine

The DBN produced discriminative features that resemble non-linear correlations in voice samples, hence reducing the need for sophisticated feature extraction and selection. Hidden layers learned audio features from the input layer, which are subsequently used by the output layer to classify the input sample.

3.7. SVM for Emotional Classification

SVM is a nonlinear classifier that employs a kernel mapping function to convert input feature vectors into a higher-dimensional feature space (Fig. 5). To maximize discrimination, the separation plane should be advantageously positioned between the borders of two classes (Kok et al., 2024). Support vectors span the plane and reduce the number of references. The global goal is to find the equation of a hyperplane that divides p (Equation [19]).

$$y_i\,[(w.x_i)+b]\geq 1\ \forall i=1,2,\ldots,N \tag{19}$$

The pair (w, b) defines a hyperplane (Equation [20]):

$$(w.x_i)+b=0 \tag{20}$$

This plane is known as the separating hyperplane. To identify the best-separating hyperplane, the following optimal problem in Equations (21 and 22) was considered:

$$minimize\, w(\propto)=\sum_{i=1}^{N}\propto_i-\frac{1}{2}\sum_{i,j,=1}^{N}\propto_i\propto_j\, y_i y_j K(x_i,x_j) \tag{21}$$

$$\text{Subject to } \sum_{i=1}^{N}\propto_i y_i=0, \propto_i\geq 0, \forall i=1,2,\ldots,N \tag{22}$$

Proper non-linear kernels can transform non-linear classifiers into linear ones in the feature space. Some common kernel functions (Tiwari et al., 2022) are listed below in Equations (23-25):

Linear kernel:

$$K(x_i,x_j)=x_i.x_j \tag{23}$$

Polynomial kernel:

$$K(x_i,x_j)=x_i.x_j+\beta)^d \tag{24}$$

Radial basis function kernel:

$$k\left(x_i,x_j\right)=exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right) \tag{25}$$

For classification, a novel attentive DBN was paired with an SVM. This hybrid technique used the strengths of both models to improve classification performance. Attention methods were built into the DBN to dynamically weigh the contributions of various hidden units at its levels. This enabled the model to focus on salient traits that are most essential to the classification goal while ignoring irrelevant ones, resulting in increased accuracy. Regularization approaches, such as batch normalization, were also used to prevent overfitting in the model. These strategies serve to stabilize the learning process and increase the model's generalization capacity, ensuring that it performs well on both training and unseen data. This strategy was especially effective for complex classification tasks since it combines attentive processes with robust regularization methods.

4. Results and Discussion

The implementation was conducted using the Python programming language on a Windows 7 (64-bit) operating system with an Intel Pentium CPU and 8 GB of memory.

4.1. Data Description

The trials were carried out with the EMODB speech emotion database, which contains 535 utterances of seven emotions from ten German professional actors (Abdusalomov et al., 2023), and the RAVDESS emotional speech dataset, which contains 1440 samples from 24 actors. The RAVDESS dataset contains five emotions: anger, none, happiness, calmness, and fear (Abdusalomov et al., 2023). The original EMODB database samples are down-sampled to 16 kHz, yielding four-second samples.

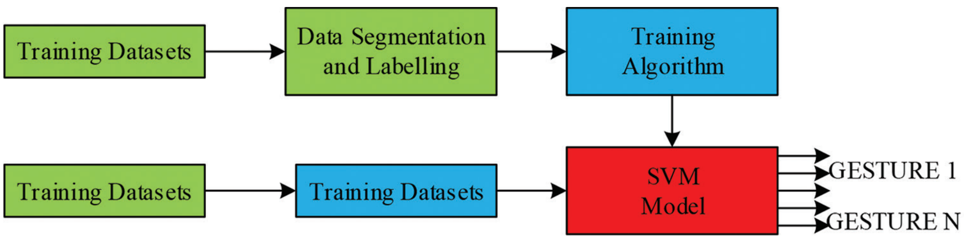


Fig. 5. Block diagram of the support vector machine system

4.2. Killer Whale Optimization Algorithm

Fig. 6 depicts the fitness graph for the Killer Whale Optimization algorithm; the fitness value at iteration 1 is 0.0920, showing that the algorithm is in the early stages of investigating possible solutions. By iteration 2, the fitness value has improved to 0.0751, indicating that the algorithm is effectively refining its search and discovering superior solutions. From iterations 3 to 5, the fitness value remains constant at 0.0751, indicating that the algorithm has hit a local optimum and is performing consistently without major improvements. Between iterations 6 and 10, the fitness value drops significantly to 0.0750, indicating that the algorithm is still engaged in exploration and exploitation, but the benefits are minor at this time.

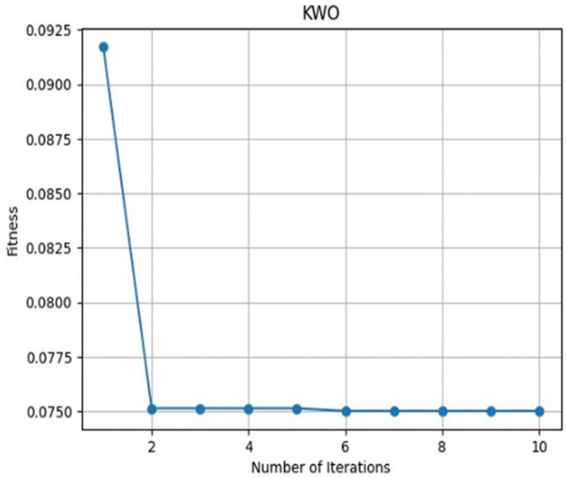


Fig. 6. Killer Whale Optimization (KWO) algorithm

4.3. Confusion Matrix

Fig. 7 compares a model's predictions to real emotions in a dataset, indicating both strengths and opportunities for development. The model guessed "none" 210 times, "angry" 35 times, "happy" 146 times, "calm" 66 times, and "fearful" 234 times. Diagonal values indicate good predictions, and higher values imply better classification accuracy. For example, the model accurately predicted "fearful" 234 times, indicating excellent performance. Off-diagonal values show misclassifications and provide areas for improvement. Overall, the model's performance is evaluated to identify its strengths and opportunities for development.

Fig. 7. Graph of the confusion matrix

4.4. Receiver Operating Characteristic Curve Under the Area Under the Curve

The true positive rate (TPR), also known as sensitivity or recall, quantifies the proportion of actual positives properly detected by the model. It is calculated using the following formula in Equation (26):

$$TPR = \frac{True\,Positive\ (TP)}{True\,Positive\left(TP\right) + False\,Negative\ (FN)} \quad (26)$$

The false-positive rate (FPR) is the percentage of actual negatives that are wrongly recognized as positives by the model. It is calculated using the following formula in Equation (27):

$$FPR = \frac{False\,Positive\ (TP)}{False\,Positive\left(TP\right) + True\,Negative\ (FN)} \quad (27)$$

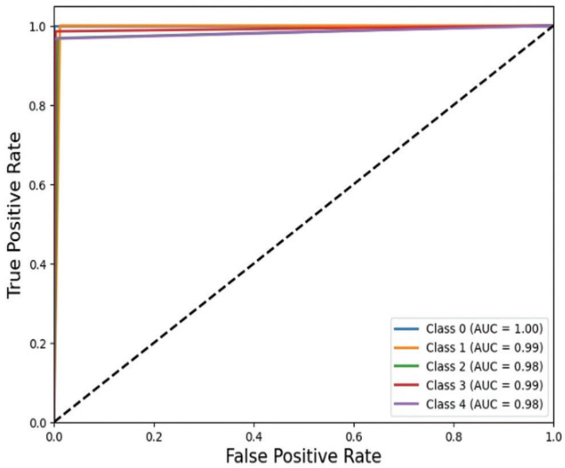


Fig. 8. Receiver operating characteristic curve under the area under the curve

Fig. 8 depicts the relationship between the TPR and the FPR, with each axis ranging from 0.0 to 1.0. The area under the curve (AUC) for each class is shown below: Class Zero has an AUC of 100%, Class One has an AUC of 99%, Class Two has an AUC of 98%, Class Three has an AUC of 99%, and Class Four has an AUC of 98%. These AUC values indicate the classifier's performance, with higher values suggesting a stronger ability to differentiate between classes.

4.5. Performance Metrics

The performance metrics include a variety of critical indicators for assessing a model's success.

Fig. 9 displays the performance metrics, with the x-axis representing the score ranging from 0.0 to 1.0. The results achieved are as follows: accuracy 98%, precision 98%, recall 98%, F1-score 98%, sensitivity 99%, and specificity 99%.

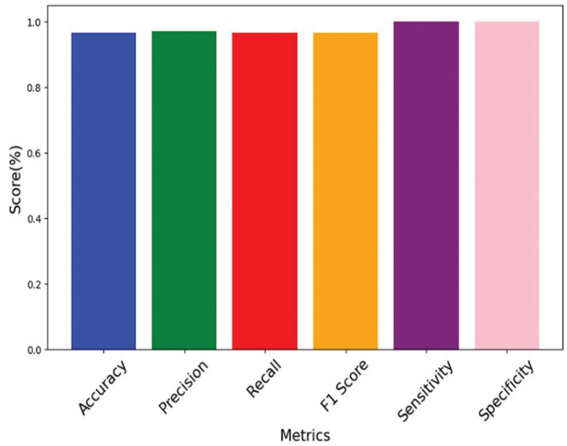


Fig. 9. Performance metrics

4.6. Precision for Each Class

The precision formula calculates the accuracy of a model's positive predictions. The definition goes as follows in Equation (28):

$$Precision = \frac{TP}{TP + FP} \tag{28}$$

Fig. 10 shows the precision reached for each emotion class, with x-axis values ranging from 0.0 to 1.0. The emotions are divided into five categories: angry, none, happy, calm, and fearful. The precision for each class is as follows: angry 99%, none 82%, happy 98%, calm 98%, and fearful 99%.

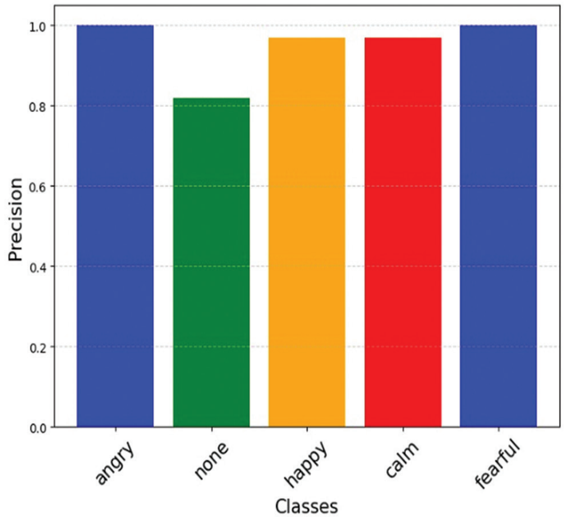


Fig. 10. Precision for each class

4.7. F1-Score for Each Class

The F1-score combines precision and recall into a single metric. It is very useful when trying to achieve a balance between precision and recall, particularly if your class distribution is asymmetrical (Equation [29]).

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{29}$$

Fig. 11 shows the F1-score reached for each emotion class, with x-axis values ranging from 0.0 to 1.0. The emotions are divided into five categories: angry, none, happy, calm, and fearful. The F1-score for each class is as follows: angry 99%, none 90%, happy 97%, calm 98%, and fearful 98%.

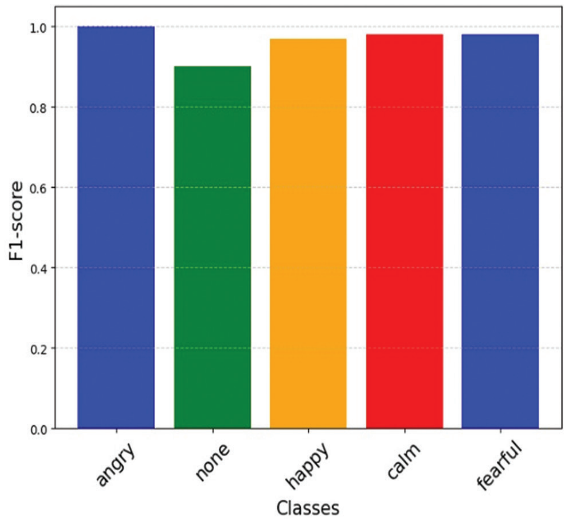


Fig. 11. F1-score for each class

4.8. Recall for Each Class

Recall, also known as sensitivity or true positive rate, is a metric that measures a model's ability to accurately identify all relevant instances in a dataset. It is highly useful for reducing false negatives (Equation [30]).

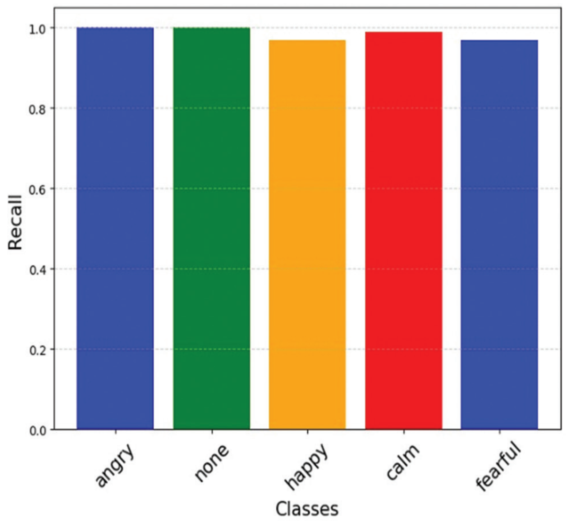


Fig. 12. Recall for each class

$$Recall = \frac{TP}{TP + FN} \quad (30)$$

It is vital in situations when missing a positive instance is more expensive or significant than mistakenly identifying a non-positive example.

Fig. 12 depicts the recall scores for each emotion class, with x-axis values ranging from 0.0 to 1.0. The emotions are classified into five types: angry, none, happy, calm, and fearful. The recall for each class is as follows: angry 99%, none 99%, happy 97%, calm 98%, and fearful 96%.

4.9. Comparative Analysis

This section demonstrates that the suggested methodology outperforms alternative models with fewer parameters, such as logistic regression

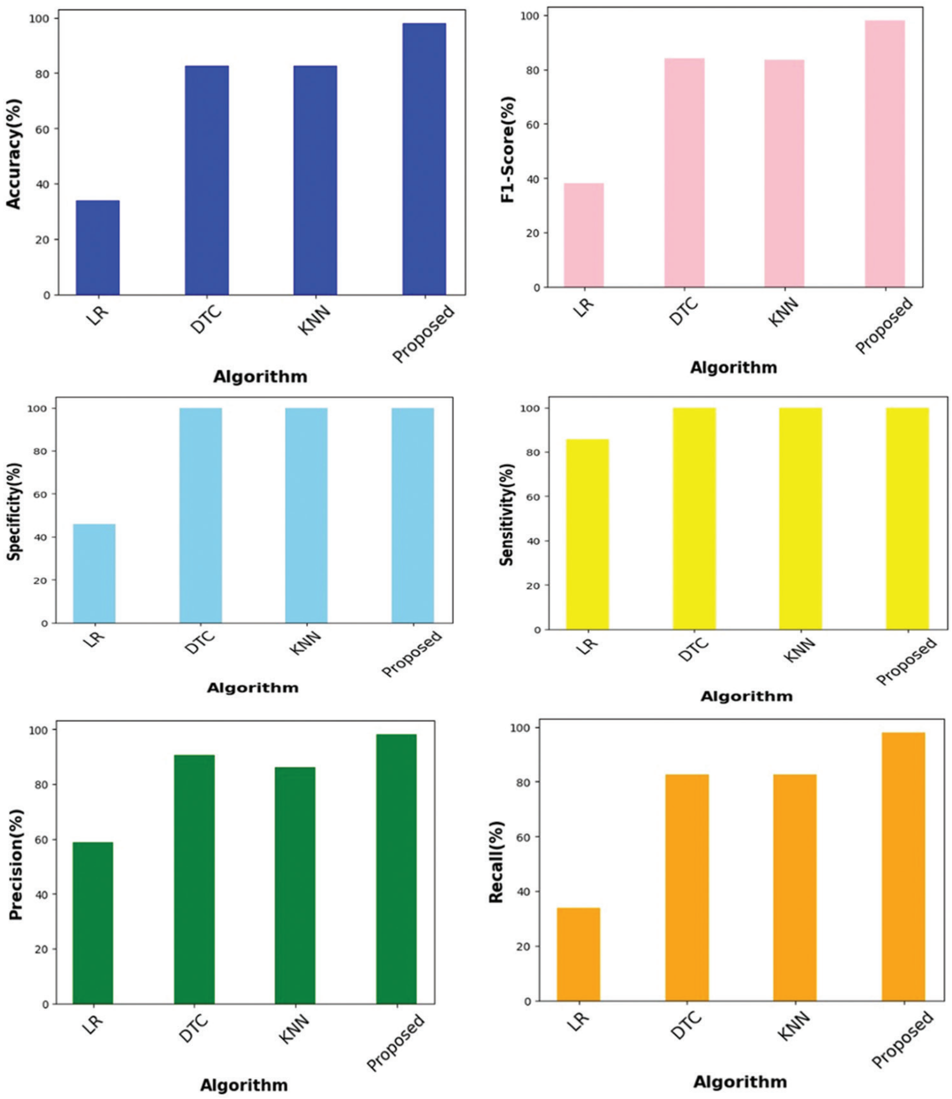


Fig. 13. Comparative analysis
Abbreviations: DTC: Decision tree classifier; KNN: K-nearest neighbor; LR: Logistic regression

Table 1. Comparative analysis

Methods	Accuracy%	F1-score%	Specificity%	Sensitivity%	Precision%	Recall%
Logistic regression	35	38	45	85	60	35
Decision tree classifier	82	84	99	99	90	83
K-nearest neighbors	80	82	99	99	85	83
Proposed	98	98	99	99	98	98

(Luna-Jiménez et al., 2021), decision tree classifier (Amartya & Kumar, 2022), and K-nearest neighbors (Subbarao et al., 2021).

Table 1 and Fig. 13 compare the performance of Marathi SER algorithms in terms of accuracy, F1-score, specificity, sensitivity, precision, and recall, demonstrating their usefulness. The logistic regression model achieved 35% accuracy, 38% F1-score, 45% specificity, 85% sensitivity, 60% precision, and 35% recall. The decision tree classifier model performed better, with an accuracy of 82%, an F1-score of 84%, a specificity of 99%, a sensitivity of 99%, a precision of 90%, and an 83% recall. The K-nearest neighbors approach achieved 80% accuracy, 82% F1-score, 99% specificity, 99% sensitivity, 85% precision, and 83% recall. In contrast, the proposed technique surpassed all other models, with 98% accuracy, 98% F1-score, 99% specificity, 99% sensitivity, 98% precision, and 98% recall.

5. Conclusion

The study's objectives are effectively addressed by the suggested methodology for SER in Marathi, which overcomes the shortcomings of current approaches and captures the distinctive grammatical and emotional subtleties of Marathi speech. The model attains remarkable performance metrics, including 98% accuracy, 98% F1-score, 99% specificity, 99% sensitivity, 98% precision, and 98% recall, using sophisticated signal processing, thorough feature extraction, and a novel classification strategy that combines an attentive DBN with an SVM. The study's shortcomings offer the potential for further research, even if the methodology addresses important issues and establishes a new standard for SER in Marathi. These include adding more languages, improving real-time processing for mobile apps, strengthening resilience to various noise and acoustic conditions, incorporating multimodal data such as physiological signals and facial expressions, and utilizing innovative architectures like transformers to further improve performance.

Acknowledgments

The authors would like to thank the Deanship of RMD Sinhgad School of Engineering for supporting this work.

References

Abdel-Hamid, L., Shaker, N.H., Emara, I. (2020). Analysis of linguistic and prosodic features of bilingual Arabic-English speakers for speech emotion recognition. *IEEE Access*, 8, 72957–72970.

Abdusalomov, A., Kutlimuratov, A., Nasimov, R., & Whangbo, T.K. (2023). Improved speech emotion recognition focusing on high-level data representations and swift feature extraction calculation. *Computers, Materials and Continua*, 77(3), 2915–2933.

Akçay, M.B., & Oğuz, K. (2020). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication*, *116*, 56–76.

Akinpelu, S., & Viriri, S. (2024). Deep learning framework for speech emotion classification: A survey of the state-of-the-art. *IEEE Access*, 12, 152152.

Alam Monisha, S.T., & Sultana, S. (2022). A review of the advancement in speech emotion recognition for indo-aryan and dravidian languages. In: *Advances in Human-Computer Interaction*. Wiley, Hoboken.

Alluhaidan, A.S., Saidani, O., Jahangir, R., Nauman, M.A., & Neffati, O.S. (2023). Speech emotion recognition through hybrid features and convolutional neural network. *Applied Sciences*, 13(8), 4750.

Amartya, J.G.M., & Kumar, S.M. (2022). Speech emotion recognition in machine learning to improve accuracy using novel support vector machine and compared with decision tree algorithm. *Journal of Pharmaceutical Negative Results,* 185–192.

Arul, V.H. (2021). Deep learning methods for data classification. In: *Artificial Intelligence in Data Mining*. Academic Press, p87–108

Bachate, R.P., Sharma, A., Singh, A., Aly, A.A., Alghtani, A.H., & Le, D.N. (2022). Enhanced marathi speech recognition facilitated by grasshopper optimisation-based recurrent neural network. *Computer Systems Science and Engineering,* 43(2), 439–454.

Bhangale, K., & Kothandaraman, M. (2023). Speech emotion recognition based on multiple acoustic features and deep convolutional neural network. *Electronics*, 12(4), 839.

Byun, S.W., & Lee, S.P. (2021). A study on a speech emotion recognition system with effective acoustic features using deep learning algorithms. *Applied Sciences*, 11(4), 1890.

Chai, J., Zeng, H., Li, A., & Ngai, E.W. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, 6, 100134.

Chaudhari, P., Nandeshwar, P., Bansal, S., & Kumar, N. (2023). MahaEmoSen: Towards Emotion-aware Multimodal Marathi Sentiment Analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(9), 1–24.

Er, M.B. (2020). A novel approach for classification of speech emotions based on deep and acoustic features. *IEEE Access*, 8, 221640–221653.

Farooq, M., Hussain, F., Baloch, N.K., Raja, F.R., Yu, H., & Zikria, Y.B. (2020). Impact of feature selection algorithm on speech emotion recognition using deep convolutional neural network. *Sensors*, 20, 6008.

Hammed, F.A., & George, L. (2023). Using speech signal for emotion recognition using hybrid features with SVM classifier. *Wasit Journal of Computer and Mathematics Science*, 2(1), 27–38.

Harhare, T., & Shah, M. (2021). Linear mixed effect modelling for analyzing prosodic parameters for marathi language emotions. *International Journal of Advanced Computer Science and Applications*, 12(12).

Kaur, K., & Singh, P. (2023). Comparison of various feature selection algorithms in speech emotion recognition. *AIUB Journal of Science and Engineering (AJSE)*, 22(2), 125–131.

Kawade, R., & Jagtap, S. (2024). Indian cross corpus speech emotion recognition using multiple spectral-temporal-voice quality acoustic features and deep convolution neural network. *Revue d'Intelligence Artificielle*, 38(3), 913–927.

Kishor, B., Mohanaprasad, K. (2022). Speech emotion recognition using mel frequency log spectrogram and deep convolutional neural network. In: *Futuristic Communication and Network Technologies*. Springer, Singapore, p241–250.

Kok, C.L., Ho, C.K., Tan, F.K., & Koh, Y.Y. (2024). Machine learning-based feature extraction and classification of emg signals for intuitive prosthetic control. *Applied Sciences*, 14(13), 5784.

Li, R., Zhao, J., & Jin, Q. (2021). Speech Emotion Recognition Via Multi-Level Cross-Modal Distillation. In: *Proceedings of Interspeech*, p4488–4492.

Li, Z., Huang, H., Zhang, Z., & Shi, G. (2022). Manifold-based multi-deep belief network for feature extraction of hyperspectral image. *Remote Sensing*, 14(6), 1484.

Lieskovská, E., Jakubec, M., Jarina, R., & Chmulík, M. (2021). A review on speech emotion recognition using deep learning and attention mechanism. *Electronics*, 10(10), 1163.

Luna-Jiménez, C., Kleinlein, R., Griol, D., Callejas, Z., Montero, J.M., & Fernández-Martínez, F. (2021). A proposal for multimodal emotion recognition using aural transformers and action units on raves dataset. *Applied Sciences*, 12(1), 327.

Madanian, S., Chen, T., Adeleye, O., Templeton, J.M., Poellabauer, C., Parry, D., & Schneider, S.L. (2023). Speech emotion recognition using machine learning-a systematic review. *Intelligent Systems with Applications*, 20, 200266.

Oh, S., & Kim, D.K. (2022). Comparative analysis of emotion classification based on facial expression and physiological signals using deep learning. *Applied Sciences*, 12(3), 1286.

Padman, S., & Magare, D. (2022). Regional language speech emotion detection using deep neural network. *ITM Web of Conferences*, 44, 03071.

Papala, G., Ransing, A., & Jain, P. (2023). Sentiment analysis and speaker diarization in hindi and marathi using finetuned whisper: Sentiment analysis in Hindi and Marathi. *Scalable Computing: Practice and Experience*, 24(4), 835–846.

Sajjad, M., & Kwon, S. (2020). Clustering-based speech emotion recognition by incorporating learned features and deep BiLSTM. *IEEE Access*, 8, 79861–79875.

Shah, F.M., Ranjan, A., Yadav, J., Deepak, A. (2021). A survey of speech emotion recognition in the natural environment. *Digital Signal Process*, 110, 102951.

Singh, Y.B., & Goel, S. (2021). An efficient algorithm for recognition of emotions from speaker and language independent speech using deep learning. *Multimedia Tools and Applications*, 80(9), 14001–14018.

Sonawane, S., & Kulkarni, N. (2020). Speech emotion recognition based on MFCC and convolutional neural network. *International Journal of Advance Scientific Research and Engineering Trends*, 5, 18–22.

Subbarao, M.V., Terlapu, S.K., Geethika, N., & Harika, K.D. (2021). Speech emotion recognition using k-nearest neighbor classifiers. *In: Recent Advances in Artificial Intelligence and Data Engineering: Select Proceedings of AIDE*.

Springer Verlag, Singapore, p123–131.
Tiwari, P., Dehdashti, S., Obeid, A.K., Marttinen, P., & Bruza, P. (2022). Kernel method based on non-linear coherent states in quantum feature space. *Journal of Physics A: Mathematical and Theoretical*, 55(35), 355301.
Yang, Z., Zhou, S., Zhang, L., & Serikawa, S. (2024). Optimizing Speech Emotion Recognition with Hilbert Curve and convolutional neural network. *Cognitive Robotics*, 4, 30–41.
Zaidi, S.A.M., Latif, S., & Qadi, J. (2023). *Cross-Language Speech Emotion Recognition Using Multimodal Dual Attention Transformers*. [arXiv Preprint].

AUTHOR BIOGRAPHIES

Varsha Gaikwad completed her Bachelor's Degree in Electronics and Telecommunication Engineering from BMIT, Solapur University in 2010 and Master's Degree in Electronics and Telecommunication Engineering with specialization in Signal Processing from SPPU Pune University in 2014. At present, she is working as an Assistant Professor in STES's RMD Sinhgad School of Engineering, Pune. She is a Life Member of ISTE, IAEG India. Her areas of interest are Speech Signal Processing and Digital Communication, Machine and Deep Learning, and Pattern Recognition. She has attended many national and international conferences. She has published six papers in the International Journal, two papers in International Conferences, and two papers in National Conferences. She has 6-year Professional Experience as Assistant Professor and she has 4-year Professional Experience as Lecturer.

DOI: 10.6977/IJoSI.202508_9(4).0007
R. Sreemol, M. B. Santosh Kumar, etc./Int. J. Systematic Innovation, 9(4), 84-105 (2025)

Fusion Net-3: Denoising-based secure biometric authentication using fingerprints

R. Sreemol[1]*, M. B. Santosh Kumar[2], A. Sreekumar[1]

[1]Department of Computer Applications, Faculty of Technology, Cochin University of Science and Technology, Kochi, Kerala, India

[2]Department of Information Technology, School of Engineering, Faculty of Engineering, Cochin University of Science and Technology, Kochi, Kerala, India

*Corresponding author E-mail: sreemolr@cusat.ac.in

(Received 3 October, 2024; Final version received 24 June, 2025; Accepted 9 July, 2025)

Abstract

Fingerprint-based authentication is a critical biometric approach for ensuring security and accuracy. Traditional methods often face challenges such as noise and suboptimal feature extraction. To address the challenges, Fusion Net-3, an extensive model, is proposed to improve the speed, precision, and security level of fingerprint-based authentication systems. Fusion Net-3 operates through two separate stages: enrollment and authentication. During the enrollment phase, advanced pre-processing of fingerprint images was performed, incorporating an enhanced bilateral filter optimized with the seagull optimization algorithm. After pre-processing, features were obtained using a two-phase method: Zernike moments for shape-based features and local binary patterns for texture-based features. This helped ensure that fingerprint features were considered comprehensive for representation. For feature selection optimization, the falcon-inspired jackal optimization algorithm was proposed, a hybrid method combining the strengths of the golden jackal optimization and falcon optimization algorithm. Then, the selected features were combined using a combination of the geometric mean and the Fisher score to facilitate classification for a balanced and novel representation. During authentication, fingerprints were processed using similar techniques for consistency. Each fingerprint was labeled as genuine or fraudulent with the aid of the Fusion Net-3 model, which leverages the combined strengths of convolutional neural networks, ResNet-50, and U-Net. The model achieved an accuracy of 98.956% and a mean squared error of 0.0234 when implemented on a Python platform. Overall, the Fusion Net-3 model demonstrated superior performance compared to existing methods, effectively enhancing authentication accuracy and security.

Keywords: Authentication, Bilateral Filtering, Enrollment, Falcon Optimization, Fusion Net-3, Golden Jackal Optimization, Seagull Optimization.

1. Introduction

Biometric authentication has become the backbone of today's security systems, using unique biological characteristics to verify identity and control access. Among the modalities, fingerprint authentication has emerged as one of the most robust, reliable, and widely adopted methods in diverse applications, ranging from unlocking mobile devices to national identification programs (Adiga & Sivaswamy, 2019; Akter et al., 2024). As a unique biometric characteristic, fingerprints offer non-intrusive, high-accuracy authentication based on the uniqueness of the minutiae patterns. To provide access, fingerprint identification systems first collect a person's fingerprints, create a customized fingerprint template, and then compare it to a database of previously approved users (Ali et al., 2020; Balsiger et al., 2020). Even though fingerprint recognition is a useful biometric authentication technique, several issues need to be resolved. The security of the system is among the important elements. Since fingerprint data are private information, it must be shielded from online dangers. Fingerprint data can be stolen and exploited by hackers for financial benefit. Therefore, designing a safe fingerprint authentication system is necessary. Efficiency is another challenge fingerprint recognition systems face. Large-scale deployments, such as those in government or corporate settings, require the system to promptly and reliably authenticate a vast number of users (Santos et al., 2024).

To address these challenges, researchers and developers have considered cutting-edge techniques and algorithms to enhance the efficacy of fingerprint recognition technologies (Zhang et al., 2019). For example, eye-tracking data, such as pupil dilation and fixation time, can be used to accurately predict cognitive load through machine learning models, including random forest (RF) and multi-layer perceptron (Dhiman & Kumar, 2019; Ding et al., 2020; Nasri et al., 2024). In addition, using an authentication system based on reconstruction is one innovative method. Reconstructing the original fingerprint picture acquired from the minute spots is how the reconstruction-based authentication system operates (Ephin & Vasanthi, 2013; Galbally et al., 2020). The distinctive qualities of a fingerprint, known as minutiae points, are the foundation of a fingerprint template. The technology can authenticate fingerprints and thwart fraudulent assaults by reconstructing the original image. However, to overcome new difficulties, traditional encryption or detection-based protection paradigms are insufficient. Traffic reshaping-based solutions, such as traffic morphing and frame quantization, provide a partial defense against specific threats but lack verifiable assurances (Abolfathi et al., 2022). Compared to conventional fingerprint recognition techniques, the reconstruction-based approach offers several benefits. First, it improves security by guarding against deceptive tactics, such as spoofing, in which a hacker fabricates a false fingerprint using synthetic materials (Gao et al., 2020; Gavaskar & Chaudhury, 2018). It is harder for fraudulent methods to deceive the system because the system authenticates fingerprints by reconstructing the original image. Second, compared to conventional fingerprint recognition methods, the reconstruction-based approach is more efficient (Gupta et al., 2020).

A reconstruction-based system can process the authentication request considerably faster because it does not need to process the entire fingerprint image. In addition, it reduces the amount of storage space needed to store fingerprint data, facilitating extensive system deployments. More research and instructional initiatives are required to enhance privacy-aware software development, while role-dependent solutions are needed to address privacy concerns in software development (Prybylo et al., 2024). The advent of hostile attacks has led to an ongoing interaction between the development of advanced attack methods and the application of strong countermeasures. This has encouraged the development of a wide range of attack techniques, each specifically designed to provide a challenge to neural networks (NNs) in different contexts. Moreover, there are critical challenges in the security and efficiency of fingerprint-based systems, including vulnerability to cyberattacks, susceptibility to spoofing, and a need for rapid and accurate performance in large-scale implementations. These challenges must be addressed to enhance the reliability of biometric systems, particularly in sensitive domains, such as financial transactions and border control (Banitaba et al., 2024; Wong & Lai, 2020). The super-learner attack, a new attack model targeting fingerprinting of HTTPS websites, has led to the development of the HTTPS obfuscation defender as a protection tactic. This defense mechanism uses adversarial example algorithms and introduces fictitious packets to interfere with categorization processes (Abolfathi et al., 2024). In parallel, reconstruction-based authentication methods have shown promise in enhancing fingerprint recognition systems' effectiveness. These approaches can stop fraudulent attacks and promptly and accurately process authentication requests (Husson et al., 2018; Xu et.al, 2019). As biometric authentication grows in popularity, designing secure and effective authentication techniques is crucial to mitigate potential cyberattacks. The main objectives of the research are as follows:

(i) Improved denoising: Enhanced bilateral filtering optimized through the seagull optimization algorithm (SOA) to preserve edge features while effectively removing noise
(ii) Robust feature extraction: Combined use of Zernike moments (shape features) and local binary pattern (LBP; texture features) for comprehensive fingerprint representation
(iii) Optimal feature selection: A novel falcon-inspired jackal optimization (FIJO) algorithm, hybridizing golden jackal optimization (GJO) and falcon optimization algorithm (FOA), was proposed to select the most discriminative features
(iv) Secure and accurate classification: Integration of convolutional NNs (CNNs), ResNet-50, and U-Net, into Fusion Net-3 to classify genuine versus fraudulent fingerprints
(v) Secure transmission: Incorporation of blockchain technology to safeguard fingerprint data integrity and confidentiality during authentication.

The remaining parts of the research include Related Works in Section 2, Proposed Model in Section 3, Results and Discussion in Section 4, and Conclusions in Section 5.

2. Related Works

The research conducted by various researchers on secure biometric authentication using fingerprints is provided in this section (Table 1).

Jia et al. (2019) proposed a highly secure biometric authentication system that can be created using a robust 3D fingerprint template to ensure the uniqueness of users' identities. This was achieved by computing minutiae triplets from the fingerprints' minutiae points, which were then used to generate the secured user template.

Kareem & Okur (2021) explored how a compressed sensing-based compression reconstruction method was developed to improve heart signal biometric recognition using portable remote bioelectric signal recognition equipment. This approach utilizes bioelectric signals to effectively enhance the limited resources of the equipment, resulting in more accurate recognition.

Khodadoust et al. (2020) proposed a mathematical model to examine the effects of transient-state excitation and k-space undersampling on magnetic resonance fingerprinting reconstructions. The model establishes a direct relationship across time-varying RF excitation, k-space sampling, and reconstruction errors, all of which are dependent on spatial variations.

Koonce & Koonce (2021) used an autoencoder network to detect presentation attacks on fingerprints. A one-class approach was used to improve detection accuracy in the study. The proposed method aims to detect fingerprint presentation attacks using only one class of data.

Lee et al. (2022) developed an FPD-M-net, which is an end-to-end CNN architecture for fingerprint image denoising and inpainting. By treating a problem as a segmentation task and incorporating a structure similarity loss function, the architecture can effectively extract fingerprints from a noisy background. The network is based on the M-net and is fully trainable.

Li et al. (2018) used an adaptive sampling strategy that utilizes an approximate volume sampling method to enhance the accuracy of radio maps for fingerprint-based indoor localization. This scheme employs a low-tubal-rank tensor to model all reference points' Wi-Fi fingerprints, aiming to reduce the expenditure required for reconstruction. The proposed approach is effective in enhancing the accuracy of indoor localization.

Li et al. (2022) introduced CRISLoc, the first localization prototype system that used channel-state information (CSI) fingerprinting based on ubiquitous smartphones. This system can passively overhear packets in real-time for its own CSI acquisition, eliminating the need for active user participation. With its innovative approach, CRISLoc demonstrates the feasibility of using smartphones for accurate and efficient localization.

Lin & Kumar (2018) used a new method for enhancing latent fingerprints, based on Finger Net, a CNN inspired by recent advancements in CNN development. The Finger Net architecture comprises a shared common convolution component and two separate deconvolution components, consisting of the enhancement and orientation branches. This

Table 1. Comparison of existing literature

Authors	Method	Advantage	Disadvantage
Jia et al. (2019)	3D fingerprint via minutiae triplets	High security	Needs 3D sensors
Kareem & Okur (2021)	Compressed sensing on heart signals	Efficient, accurate	Sensitive to signal noise
Khodadoust et al. (2020)	Magnetic resonance fingerprinting model	Explains error causes	Complex to apply
Koonce & Koonce (2021)	One-class autoencoder for presentation attack detection (PAD)	Works with one-class data	May miss unseen attacks
Lee et al. (2022)	FPD-M-net for denoising	Accurate, end-to-end	Needs a large training set
Li et al. (2018)	Adaptive sampling for radio maps	Lower cost	Sampling-dependent
Li et al. (2022)	CRISLoc (channel-state information via smartphones)	Passive, no user input	Varies by environment
Lin & Kumar (2018)	Finger Net convolutional neural network	Enhanced latent prints	Limited to poor prints
Liu et al. (2020a)	Contactless 3D with Siamese nets	No touch needed	Complex setup
Liu et al. (2021)	Cahn–Hilliard for restoration	Simple, effective	Narrow scope
Liu et al. (2022)	CFD-PAD with presentation attack-adaptation loss	Better spoof detection	High training cost
Liang & Liang (2023)	Res-WCAE for denoising	Lightweight, detailed	Limited global view
Rahman et al. (2022)	Minutiae and chaffs for security	High privacy	Complex template
Algarni (2024)	Multi-fingerprint (BioPass)	More secure login	Needs user effort

Abbreviations: CFD: Channel-wise feature denoising; Res-WCAE: Residual wavelet-conditioned convolutional autoencoder.

approach is effective in improving the quality of latent fingerprints for better identification.

Liu et al. (2020a) proposed a contactless 3D fingerprint representation learning model, utilizing a CNN. The model incorporates a fully convolutional network for fingerprint segmentation and three Siamese networks to learn a multi-view 3D fingerprint feature representation. This approach successfully produced precise 3D fingerprint representations without requiring physical contact.

Liu et al. (2021) presented a reliable and efficient fingerprint image restoration technique, utilizing a non-local Cahn–Hilliard equation designed for modeling microphase separation of di-block copolymers. The method employs a Gauss–Seidel-type iterative approach, resulting in a straightforward implementation process that enhances the quality of fingerprint images with effectiveness and efficiency.

Liu et al. (2022) proposed a new channel-wise feature denoising fingerprint presentation attack detection technique that addresses the redundant noise data that were overlooked in earlier research. The suggested approach determines discriminative and "noise" channels by evaluating the significance of each channel to learn significant fingerprint picture features. To reduce interference, "noise" channel propagation is then muted in the feature map. To make the feature distribution of spoof fingerprints more dispersed and that of live fingerprints more aggregate, a presentation attack-adaptation loss is specifically introduced to restrict the feature distribution.

Liang & Liang (2023) presented the residual wavelet-conditioned convolutional autoencoder (Res-WCAE), a lightweight and reliable deep learning architecture with the Kullback–Leibler divergence regularization that is specifically designed for fingerprint image denoising. Res-WCAE consists of one decoder and two encoders: a wavelet encoder and an image encoder. The bottleneck layer is conditioned on the compressed representation of features derived from the wavelet encoder, which processes both approximation and detail sub-images in the wavelet-transform domain. Residual connections between the image encoder and decoder are employed to preserve fine-grained spatial features.

Rahman et al. (2022) proposed a strategy based on minutiae to defend fingerprint templates against security breaches. Even though the database provides an attacker with these safe minutiae templates, it is difficult to access the actual minutiae features of a user fingerprint. Minutiae-based techniques have been applied in the study, and the fingerprint minutiae characteristics and their associated parameters have been investigated. This technique creates a secure template by altering the actual minutiae information and adding additional chaffs (fake minutiae) to safeguard the minutiae features. It is nearly impossible to obtain fingerprint features or vault information using this method because the template pattern is completely different for each new fingerprint, even if the parameters are the same for the same fingerprint.

Algarni (2024) presented the novel idea of a multi-fingerprint sequence authentication procedure for user verification. For improved convenience and security, this multifactor methodology combines the use of several fingerprints with a sequence pattern, as opposed to the conventional, single-fingerprint methods. In addition, as an alternative to biometric usernames and text passwords, this study offers a thorough assessment of BioPass, a novel authentication mechanism that uses a multi-fingerprint sequence pattern.

New biometric systems utilizing optimal-effort fingerprint templates demonstrate improved verification security, accuracy, and efficiency. However, these systems face limitations, such as model and hardware complications, data training requirements, and potential issues with compromised signals and noise. A unification of systems integrating multimodal has not been synthesized, and computational value overstated cases. This literature review highlights the potential of emerging approaches that maintain optimal-effort security, optimize user experience, minimize hardware dependency, and mitigate potential exploitation risks, all without compromising user choice or agency. The proposed Fusion Net-3 model, optimized using SOA, incorporates enhanced bilateral filtering, hybrid feature extraction, a novel FIJO-based feature selection method, and an integrated CNN–ResNet-50–U-Net model for classification.

3. Fusion Net-3

The proposed model, Fusion Net-3, comprises two stages, enrollment and authentication. Using enhanced bilateral filtering, noises are removed from the images. Filter parameters are optimized using an SOA, and the images are enhanced using a contrast enhancement technique. The pre-processed output is then used to extract features based on shapes and textures. Ultimately, a unique FIJO, a combination of the GJO and FOA, is used for feature selection. The features are combined using geometric mean and Fisher score. The fingerprint images are given as input into the second phase, where pre-processing, feature extraction, feature selection, and feature fusion are conducted using similar approaches. Finally, the efficient (correct or incorrect) fingerprints are detected using the Fusion Net-3 model, which combines CNN, ResNet-50, and U-Net models. Fig. 1 illustrates the proposed model.

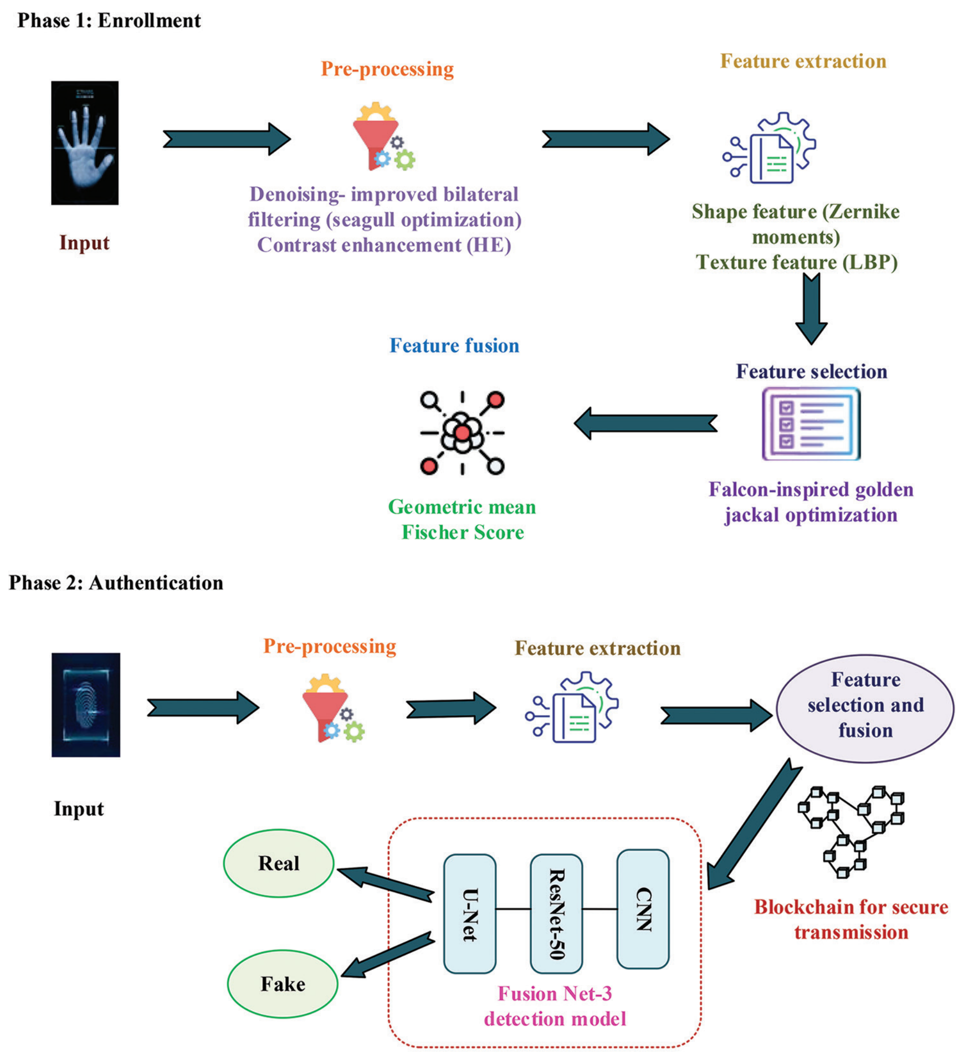


Fig. 1. Architecture for the proposed authentication system

3.1. Enrollment

Enrollment is the first phase of the model, which collects the information of scanned hands. The following steps provide a detailed explanation of this phase.

3.1.1. Pre-processing

Pre-processing is a necessary step that transforms raw datasets into a desired format, ensuring the accuracy and applicability of the data. This study employed two methodologies, improved bilateral filtering (Liu et al., 2020b; Mahum et al., 2023) and histogram equalization (HE) (Narodytska & Kasiviswanathan, 2017; Paris et al., 2009).

(a) Noise Reduction Using Improved Bilateral Filtering

A non-linear filter–the bilateral filter–preserves edges while removing noise. It considers the geometric proximity of adjacent pixels and the similarity of their gray levels (Afshari et al., 2017). The filter computes the local neighborhood's weighted sum of pixels. These pixels are replaced with the weighted average of their neighbors (Gavaskar and Chaudhury, 2018). Both the intensity difference and the spatial distance of the pixel relative to its neighborhood can be used to determine the weights. It is formulated as:

$$O(p,q) = M\left(\frac{1}{A(p,q)}\sum_{(i,j)\in N(p,q)} f_b\begin{pmatrix} I(x,y) \\ -I(i,j)\end{pmatrix} .f_c\big((p,q)-(i,j)\big)I(i,j)\right) \quad (1)$$

Where $O(p, q)$ is the filtered output image at pixel (p, q), M is the input data, $A(p, q)$ is the normalization factor, $N(p, q)$ is the spatial neighborhood of $I(p, q)$, and f_b is the intensity kernel, which is the difference in intensities between the center pixel (*p, q*) and the surrounding pixels (*i, j*), f_c is the spatial kernel, which is the spatial proximity between (*p, q*) and (i, j). f_b and f_c can be defined as:

$$f_b\left(d\right)=e^{-\frac{d^2}{2\sigma_b^2}} \tag{2}$$

Where d is the intensity difference between pixels, σ_b controls the intensity kernel's standard deviation;

$$f_c\left(d\right)=e^{-\frac{d^2}{2\sigma_c^2}} \tag{3}$$

Where d is the Euclidean distance between pixels, σ_c controls the spatial kernel's standard deviation.

$$O\left(x,y\right)=\frac{1}{A(p,q)}\sum\nolimits_{(i,j)\in N(p,q)} e^{\frac{-\left|I(i,j)-I(p,q)\right|^2}{2\sigma_b^2}}.e^{\frac{-(p,q)-(i,j)}{2\sigma_c^2}}, I\left(i,j\right)=O \tag{4}$$

As a result, edge preservation and noise reduction can be accomplished using bilateral filter (BF). The improved bilateral filtering strategy was utilized in previous studies, involving the application of an SOA (Praseetha et al., 2019) to tune the filter's parameters, including the spatial kernel (σ_s) and intensity kernel (σ_r), to increase denoising efficiency. The main purpose of the SOA is to mimic the migratory and predatory behaviors of seagulls in their native environment, achieving optimal denoising performance. For example, the pursuit of food is a defining characteristic of gull migratory behavior. The manner in which seagulls hunt migratory birds at sea is referred to as "attack behaviors." The system replicates the gulls' migratory patterns from one area to another. To locate the search agent, *L* (mobile behavior) is considered to avoid collisions among seagulls:

$$\overrightarrow{N_q}=L\times\overrightarrow{S_q}\left(p\right) \tag{5}$$

Where $\overrightarrow{N_q}$ and $\overrightarrow{S_q}$ are the search agent's position and present positionrespectively.

$$L=f_c-\left(p\times\left(\frac{f_c}{\max iter}\right)\right); p=0,1,\ldots,\max iter \tag{6}$$

Where f_c is the parameter's frequency. When the search agent avoids collisions with seagulls, it moves in the optimal nearby direction:

$$\overrightarrow{M_q}=B\times\left(\overrightarrow{S_{bq}}\left(p\right)-\overrightarrow{S_q}\left(p\right)\right) \tag{7}$$

Where $\overrightarrow{M_q}$ indicates the search agent's location, $\overrightarrow{S_q}$ is the direction of the optimum search agent $\overrightarrow{S_{bq}}$, and *B* balances between exploration and exploitation behavior, which is stochastic and is defined as:

$$B = 2 \times L^2 \times rdm \tag{8}$$

Where *rdm* is a random number within [0,1].

Furthermore, the search agent may adjust its ranking concerning the most prominent search agent:

$$\overrightarrow{D_q}=\left|\overrightarrow{N_q}+\overrightarrow{M_q}\right| \tag{9}$$

Where $\overrightarrow{D_q}$ shows a distinction between the search agent and the most suitable search agent (i.e., the optimal seagull with a lower fitness value).

Seagulls exhibit a spiral movement while attacking their prey. The behavior in the *x*, *y*, and *z* planes can be stated as follows:

$$x' = r \times cos(l) \tag{10}$$

$$y' = r \times sin(l) \tag{11}$$

$$z' = r \times l \tag{12}$$

$$r = g \times e^{lh} \tag{13}$$

The updated position of the search agent is given as:

$$\overrightarrow{S_q}\left(p\right)=\left(\overrightarrow{D_q}\times x'\times y'\times z'\right)+\overrightarrow{S_{bq}}\left(p\right) \tag{14}$$

Configurations for reproducibility include a 512 × 512 pixel input image size, a 5 × 5 filter window, and optimized parameter values of $\sigma_s = 1.5$ and $\sigma_r = 0.8$.

(b) Contrast Enhancement using Histogram Equalization

Dynamic range, or the ratio of the brightest to darkest pixel intensities, determines image contrast. There are various applications for contrast enhancement techniques to improve low-contrast images. HE is a commonly employed technique. The probability distribution of input gray levels is used to map the gray levels. The histogram of the image is stretched and flattened to increase contrast. The probability density function $S(I_y)$ for image *I* is provided as follows:

$$S\left(I_y\right)=\frac{n_y}{n} \tag{15}$$

Where $y = 0, 1, \ldots, L-1$, *y* is the series of time of the level I_y in input images, and *n* indicates samples.

The cumulative density function is defined as:

$$c\left(i\right)=\sum\nolimits_{i=0}^{y}S\left(I_i\right) \tag{16}$$

Where $I_y = i$, $y = 0,1\ldots,L-1$, and $c(I_{L-1}) = 1$ (default).

The transform function $f(i)$, based on the above equation, is given as:

$$f(i) = I_0 + (I_{L-1} - I_0)\, c(i) \tag{17}$$

The HE results, V, which is the function of $\{S(q, r)\}$, is given as:

$$V = f(I) = \{f(I(k,i)) | \forall I(k,i) \in I\} \tag{18}$$

As a result, the contrast of the images is enhanced.

3.1.2. Feature extraction

Feature extraction is the process of extracting a set of features from the pre-processed data. The shape and texture features are extracted using the Zernike moments (Shadab et al., 2022) and LBP (Shehu et al., 2018; Vogel, 2022) techniques.

(a) Shape Feature Extraction Using Zernike Moments

The low-level feature that Zernike moments yield is significant. Zernike moments provide rotationally invariant descriptors based on the order n and repetition m of Zernike polynomials, computed from radial polynomials which capture the finer details of shape. Since it is rotationally invariant, recognition is based on the magnitude of these moments. The following are the Zernike radial polynomials:

$$R_{nm}(p,q) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \times \frac{\left(p^2+q^2\right)^{\frac{n-2s}{2}} (n-s)!}{s!\left(\left(n+\frac{|m|}{2}-s\right)!\left(\left(n-\frac{|m|}{2}-s\right)!\right)\right)} \tag{19}$$

Where n is a non-negative integer, m is a non-zero integer, n-$|m|$ is even, and $|m| \leq n$.

The (n, m) order of the Zernike bias function is given as:

$$V_{nm}(p,q) = R_{nm}(p,q)\, e^{jm\theta} \tag{20}$$

Where j is $\sqrt{-1}$, and θ is $tan^{-1}\left(\frac{y}{x}\right)$.

The Zernike moments of order n and repetition m of a function $f(p, q)$ are defined as:

$$Z_{nm} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(p,q) V_{nm}^*(p,q)\, dp dq \tag{21}$$

Where V_{nm}^* is a complex conjugate of V_{nm}.

(b) Texture Feature Extraction Using Local Binary Pattern

Color features use individual pixels, whereas texture features use groups of pixels. In the feature maps, an LBP is computed for every pixel. After comparing the data, the results are binary encoded. A collection of binary characteristics is produced, capturing certain local texture patterns. It derives texture information from the surface features, patterns, and edges. The (p_c, q_c) gray value of a center pixel is compared to the pixels of its eight neighbors to create an ordered binary set, or LBP. As a result, the LBP code is expressed as a decimal octet value.

$$z = LBP(p_c, q_c) = M\left(\sum_{n=0}^{7} S(i_n - i_c) 2^n\right) \tag{22}$$

Where i_c is the gray value of the center pixel (p_c, q) and i_n is the gray value of the pixels of its eight neighbors. After transformation, the result obtained is given as:

$$S(i_n - i_c) = \begin{cases} 1 \; ; (i_n - i_c) \geq 0 \\ 0 \; ; (i_n - i_c) < 0 \end{cases} \tag{23}$$

The LBP is a texture representation method where the intensity difference between a central pixel and its neighbors is encoded. LBP involves parameters, such as radius (R) and neighbors (P), for generating binary patterns. Configurations include a radius of 1–3 pixels, 8–16 neighbors, and uniform pattern mapping. This two-phase approach will thus ensure full representation of features since it captures both the geometric structure and surface details of the fingerprint images.

3.1.3. Feature fusion

By utilizing the geometric mean and Fisher score, the fusion of selected features is achieved. This methodology ensures that every feature carries equal weight, while simultaneously maximizing the distinction between genuine and imposter fingerprints.

(a) Geometric Mean

The geometric mean is a widely used mathematical concept in finance, science, and engineering, providing a measure of central tendency for a set of numbers. Unlike the arithmetic mean, the geometric mean multiplies the values and takes the n^{th} root of the product, making it more sensitive to changes in smaller values. It is particularly useful in datasets with extreme values or outliers. The geometric mean is commonly used to calculate rates of change, such as growth or inflation rates. It is given as:

$$Gm = \sqrt[m]{a_1 \times a_2 \times .... a_m} \tag{24}$$

Where $a_1, a_2 ..., a_m$ are the observations.

(b) Fisher Score

Fisher score is an effective method for reducing data feature dimension. Its major goal is to discover a feature subset that maximizes the selected features in a data space. Techniques such as clustering and dimensionality reduction are utilized to decrease distances between data points within a class and increase distances between data points in different classes. The Fisher score of the j^{th} feature is calculated as:

$$FS1\left(f_j\right) = \frac{L_y(f_j)}{\sum_{K=1}^{E} L_T^{(K)}(f_j)} \tag{25}$$

Where $L_y(f_j)$ is the between-class scatter, which measures the spread of data points between different classes in a selected feature space, and $L_T^{(K)}\left(f_j\right)$ is the within-class scatter, which measures the spread of data points within each class for a particular feature. The between-class scatter of the j^{th} feature is calculated by taking the sum of the squared differences across the mean μ_j of the j^{th} feature in each class and the overall mean $\mu_j^{(K)}$ of the j^{th} feature, multiplied by the number of samples m_K in each class:

$$L_y\left(f_j\right) = \sum_{K=1}^{E} m_K (\mu_j^{(K)} - \mu_j)^2 \tag{26}$$

The within-class scatter matrix of the j^{th} feature calculates the variance of that feature in the K^{th} class by summing the squared differences between each data point and the mean of j^{th} feature in the same class:

$$L_T^{(K)}\left(f_j\right) = \sum_{i=1}^{m_K} (a_{ji}^{(K)} - \mu_j^{(K)})^2 \tag{27}$$

Finally, the features are combined and the results are stored.

3.2. Authentication

Authentication is the second phase of the model, which uses the finger images as the input. In this phase, the input finger images are pre-processed, and features are extracted and selected using approaches similar to those employed during the enrollment phase. Once the feature selection is completed, blockchain (Akanfe et al., 2024) technology is used to securely transfer the generated data. With the use of a fixed distributed database and a hash chain of blocks that each include time-stamped transactions, this system organizes data. Every block in the chain has a distinct code, or hash, that identifies it and establishes a sequential relationship with the blocks that came before it. The hash function employed in blockchain depends on several important factors for its success.

On secure transmission of data through the blockchain, finally, the Fusion Net-3 architecture is used to detect real and fake fingerprints (Fig. 2). CNNs, ResNet-50, and U-Net are combined to provide an efficient method for extracting and processing image features. Raw picture data are first fed into the network to start the process. The CNN-ResNet-50 (Wang et al., 2020) branch uses convolutional layers to extract local image features, batch normalization to normalize these features, and activation functions to introduce non-linearity. To obtain robust features and reduce computational costs, feature maps are down-sampled through pooling layers. Deeper architectures are made possible by ResNet-50 blocks, which solve the vanishing gradient issue in deep NNs. The encoder-decoder structure used by the U-Net (Wang & Yang, 2024) branch allows it to extract both high-level and low-level image features simultaneously. While pooling layers down-sample features in the encoder path, convolutional layers extract local features at various scales. In the decoder path, up-sampling layers retrieve spatial information and merge features from corresponding encoder and decoder layers. The complementary information is then combined by merging the feature maps from the two branches. To map features to class probabilities, a fully connected layer receives this flattened representation of the combined features. The ultimate class probabilities for classification are output by a SoftMax layer.

3.2.1. Convolutional neural network–Resnet-50

Convolutional layers are usually the most important components in CNNs. At each convolutional layer, the input is convolved with an array of learnable filters, yielding a variety of feature maps. Let s_i represent i^{th} feature map of S. We can utilize the weight W_k and bias b_k to characterize the c^{th} filter. The k^{th} output is given as:

$$y_k = \sum_{i=1}^{d} f\left(s_i \times W_k + b_k\right);\ \mathrm{k} = 1,\ 2,\ \mathrm{n} \tag{28}$$

Where f (•) is an activation function, while W_k and b_k are weights and bias, respectively. The extraction of local features from input fingerprint images is largely dependent on the CNN component. It identifies images' patterns and textures, both of which are essential for fingerprint recognition, by applying convolutional filters to images.

A 50-layer network, called ResNet-50, has demonstrated efficacy in pre-trained image classification. Deeper NN trainings are difficult due to disappearing gradient problems. Such problems are attempted to be addressed by residual learning. A layer that learns low- or high-level features is taught specifically for that task. Deep NN training can be stabilized and sped up with the use of batch normalization.

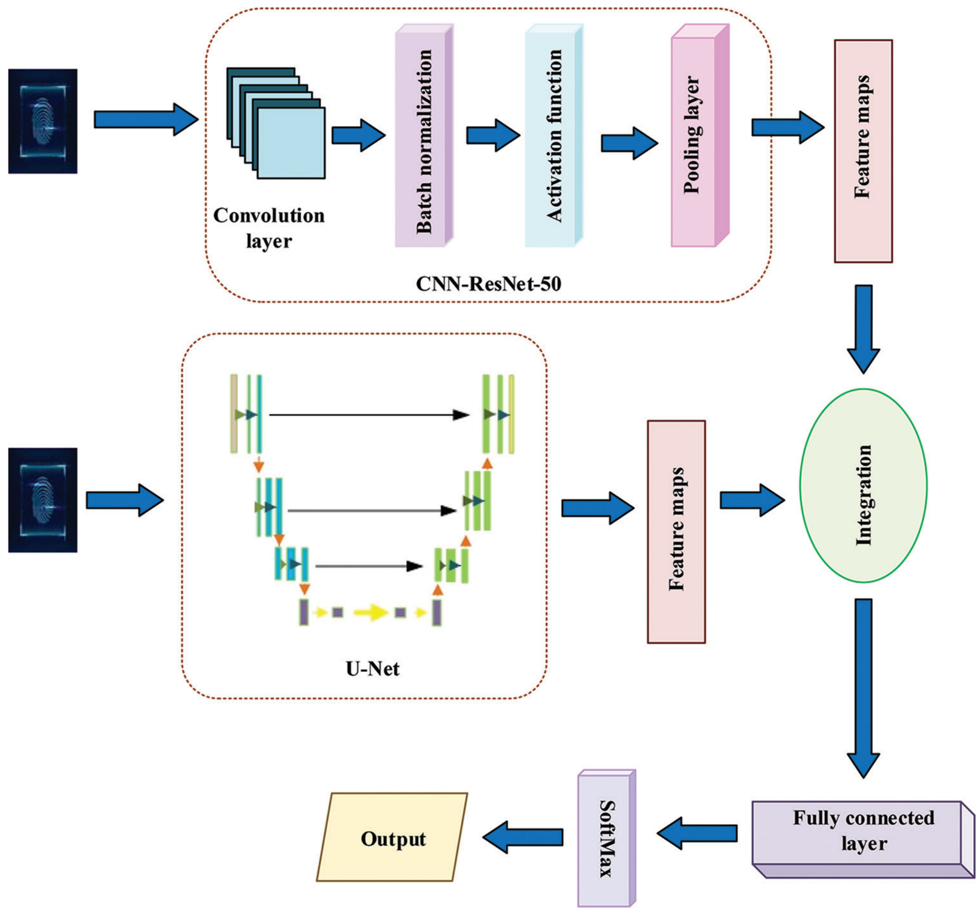


Fig. 2. Fusion Net-3 architecture

Fusion Net-3 integrates ResNet-50 into its architecture. The residual connections in the model enable the network to learn residual functions rather than directly approximating the underlying mapping. ResNet-50 allows Fusion Net-3 to train deeper NNs without sacrificing efficiency. In addition, batch normalization is applied to accelerate convergence and stabilize the training process.

Batch normalization works by modifying and scaling each layer's activations to normalize them. It is usually used before the activation function in ResNet-50, following the convolutional layers. Each feature map essentially generates a new mean and standard deviation for every pixel. The procedure involves normalizing the *z*-score, then multiplying the results by an arbitrary scale parameter (α) and adding another arbitrary offset value (β). These are the specifications for the batch normalization.

$$s` = \left(\frac{(s-\mu)}{\sigma} \times \alpha \right) + \beta \tag{29}$$

Where s, s` are feature maps and batch-normalized value, element, and μ is the mean.

The activation function is used after batch normalization, improving non-linearity. The rectified linear unit (ReLU) is used, and it is given as:

$$\sigma(s) = max(0, s) \tag{30}$$

3.2.2. Pooling layer

At times, there is redundant information existing in signals. The pooling process reduces the spatial size of the feature maps steadily, reducing the computation and parameter count of the network. With an $n \times n$ window-size neighbor denoted as P, the standard pooling technique is represented as:

$$Z = \frac{1}{F} \sum_{i,j \epsilon P} s_{i,j} \tag{31}$$

Where F is the number of elements in P, and $s_{i,j}$ is an activation value of position i, j.

3.2.3. U-Net

The U-Net architecture is used to extract contextual information and fine-grained details from images, providing pertinent data for categorization. In the Fusion Net-3 model, the U-Net component is crucial for extracting both high-level and low-level features from the input fingerprint. The autoencoder architecture, in which the left path (encoder) is referred to as the contracting or compressive path and is built on a standard CNN deep network, is the most similar to the Fusion Net-3 model's basic structure, which consists of two main paths. The network's second path, known as the decoder or expanding path (also called the up-sampling or synthesis path in some references), is made up of both convolutional and deconvolutional layers. The expanding path uses optimized techniques, such as concatenating skip connections, to recover the input image resolution and spatial structure, which are both compromised during down-sampling. The network generates dense predictions at higher resolutions in the expanding path, helping it to learn spatial classification information. Furthermore, it increases the output's resolution, which is then transferred to the final convolutional layer to produce a segmented image with the same shape as the input image.

The contracting path is a typical CNN network, consisting of two successive 3 × 3 convolutions, followed by non-linear activations (e.g., ReLU), and a max pooling layer. This structure is repeated numerous times until the bottleneck is reached. The strided convolutions and pooling layers in the contracting path decrease dimensions while increasing the channel number and receptive field. This expanding path, which involves up-sampling feature maps from the bottleneck using 2 × 2 up-convolutions to recover the input image's dimensions, is where the novelty of the U-Net originates. There are normal 3 × 3 convolutions and ReLU activations along with a 2 × 2 up-convolution in every step of the expanding path. This path's up-sampling ability reduces the number of channels by half, while the image's width and height are increased through the up-convolution. After cropping, a concatenation from the same level layer in the feature map's contracting path is added to expand the image's dimensions, while the spatial features are preserved following each 2 × 2 up-convolution.

3.2.4. Fully connected layer

The combined output features of CNN–ResNet-50 and U-Net are input into the fully connected layer to improve detection accuracy. The output is input into the SoftMax function to predict the class of an input image.

$$\sigma(z) = \frac{\exp(z)}{\sum_{i=1}^{r} \exp(z_i)} \tag{32}$$

Where z and r are input and classes, respectively. These architectures are combined in Fusion Net-3, which is used in this study, to produce a strong feature representation that improves its ability to distinguish between real and fake fingerprints. ResNet-50 is used as a pre-trained backbone, while CNN and U-Net convolutional layers are configured based on empirical results. Key hyperparameters, including learning rate, batch size, number of epochs, dropout rate, and Adam optimizer, are selected based on grid search experiments from the training dataset. The FIJO algorithm is used to optimize feature selection, whereas hyperparameters and network topology are manually tuned to balance system performance and training time. When compared to using individual models alone, this integration improves fingerprint authentication's robustness and accuracy.

4. Results and Discussion

Performance indicators were used to assess the outcomes of the dataset obtained. The computed results were compared across the proposed model and existing models (CNN, ResNet-50, and GoogLeNet) through a Python platform. The results were computed by considering a learning rate of 70% for training and 30% for testing. The equations are shown as follows:

- Accuracy: It is the proportion of correctly predicted values to all observations, including total positive (TP), total negative (TN), false positive (FP), and false negative (FN). It is expressed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{33}$$

- Precision: It is the percentage of a model's true positive predictions. Precision is key in determining whether an observation reflects a real phenomenon. It is stated as:

$$Precision = \frac{TP}{TP + FP} \tag{34}$$

- Recall: It is the percentage of true positives that are correctly identified. It is stated as:

$$Recall = \frac{TP}{TP + FN} \tag{35}$$

- F-measure: It is a statistic from the combination of precision and recall, serving as a general performance evaluation score. It is stated as:

$$F\text{-}measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{36}$$

- Matthews correlation coefficient (MCC): It is the degree of correlation between the predicted and actual outcomes. It is expressed as:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{37}$$

- Peak signal-to-noise ratio (PSNR): It is the ratio between the examined image and the recovered image. *Max* represents an image's maximum value, where the maximum value is 255 if 8 bits/sample are utilized to represent the pixel. Higher PSNR indicates better quality. It is given as:

$$PSNR = 10log\left(\frac{Max^2}{MSE}\right) \tag{38}$$

- Mean squared error (MSE): It is a regression model's confidence measure, indicating the error between predicted class probabilities and true labels. A lower MSE value indicates stronger prediction confidence, especially when using soft probability outputs from the SoftMax layer. The average squared variance of the actual and predicted values is MSE. It is expressed as:

$$MSE = \frac{1}{N}\sum\nolimits_{x=1}^{N}\left[Z_x - \widehat{Z_x}\right]^2 \tag{39}$$

Where Z_c and $\widehat{Z_c}$ are the actual and projected values, respectively, and V is the total number of data points.

- False acceptance rate (FAR): It is calculated by dividing the number of false-positive recognitions by the total number of identified attempts. It is expressed as:

$$FAR = \frac{FP}{FP + TN} \times 100 \tag{40}$$

- False rejection rate (FRR): It is the number of false rejections divided by the total number of transactions. A lower FRR indicates that fewer cases are being rejected by the biometric system. It is expressed as:

$$FRR = \frac{FN}{FN + TP} \times 100 \tag{41}$$

4.1. Parameter Evaluation

Performance measures, including accuracy, precision, recall, F1-score, MCC, MSE, FAR, and FRR, were used to evaluate the proposed Fusion Net-3 model and conventional strategies (CNN, ResNet-50, and GoogLeNet) for three datasets (LUMID, LivDet, and Biometrika; refer to Appendix A1). The models' numerical results for Datasets 1, 2, and 3 are shown in Tables 2-4, respectively.

Fig. 3 shows the graphical analysis of accuracy and precision of all tested models across databases, with the Fusion Net-3 model achieving the highest accuracy of 98.956%. In Dataset 2, the accuracy of the proposed model (95.654%) outperformed other models, followed by GoogLeNet (93.876%), ResNet-50 (92.654%), and CNN (91.765%). Similarly, in Dataset 3, Fusion Net-3 recorded an accuracy of 95.432%, markedly higher than those of CNN (92.876%), ResNet-50 (93.654%), and GoogLeNet (94.876%). The proposed model's precision was 98.548%, indicating a better performance compared to CNN, ResNet-50, and GoogLeNet. In Datasets 2 and

Table 2. Numerical results for Dataset 1 by models

Parameter	Fusion Net-3	CNN	ResNet-50	GoogLeNet
Accuracy (%)	98.956	94.562	95.632	96.327
Precision (%)	98.548	94.685	95.975	96.852
Recall (%)	98.967	94.536	95.524	96.384
F-measure (%)	98.675	94.687	95.862	96.247
MCC (%)	98.635	94.368	95.427	96.784
FAR (%)	15.573	45.453	34.543	52.112
FRR (%)	19.534	46.642	36.8765	48.765
PSNR (dB)	18.524	15.527	14.753	12.864
Time complexity	5.324	9.325	8.357	7.368
MSE	0.0234	0.0612	0.0560	0.0474

Abbreviations: CNN: Convolutional neural network; MCC: Matthews correlation coefficient; MSE: Mean squared error; PSNR: Peak signal-to-noise ratio.

Table 3. Numerical results for Dataset 2 by models

Parameter	Fusion Net-3	CNN	ResNet-50	GoogLeNet
Accuracy (%)	95.654	91.765	92.654	93.876
Precision (%)	94.876	91.543	92.876	92.543
Sensitivity	94.123	91.876	92.543	92.432
Specificity	93.543	91.321	92.654	92.765
Recall (%)	94.432	91.765	92.876	92.654
F-measure (%)	93.765	91.654	92.765	92.876
MCC (%)	93.432	90.876	91.876	92.123
FAR (%)	12.987	22.876	21.543	17.654
FRR (%)	10.543	23.654	20.876	18.543
PSNR (dB)	32.432	29.876	16.432	15.876
Time complexity	5.987	17.543	16.432	15.654
MSE	0.0201	0.1123	0.0987	0.0912

Abbreviations: CNN: Convolutional neural network; FAR: False acceptance rate; FRR: False rejection rate; MCC: Matthews correlation coefficient; MSE: Mean squared error; PSNR: Peak signal-to-noise ratio.

3, the proposed model achieved the highest precision of 94.876% and 95.654%, respectively, indicating accurate fingerprint classification. This precision value is markedly higher than those in previous works by Yin et al. (2019).

Fig. 4 compares the recall and F-measure results of the tested models. The Fusion Net-3 model achieved the highest recall with a value of 98.967%, outperforming CNN, ResNet-50, and GoogLeNet. The proposed model's recall values were 94.432% for Dataset 2 and 95.543% for Dataset 3, also outperforming other models. Similarly, the F-measure of the proposed model was 98.675% for Dataset 1, markedly higher than CNN, ResNet-50, and GoogLeNet. Its F-measure values were 93.765% for Dataset 2 and 95.432% for Dataset 3. These findings indicate the proposed model's superiority over other models in previous research by Zheng et al. (2019). They also highlight the importance of considering both recall and F-measure in the development of effective machine learning models.

Table 4. Numerical results for Dataset 3 by models

Parameter	Fusion Net-3	CNN	ResNet-50	GoogLeNet
Accuracy (%)	95.432	92.876	93.654	94.876
Precision (%)	95.654	92.543	93.432	94.543
Sensitivity	95.321	92.654	93.876	94.321
Specificity	94.876	92.876	93.543	94.876
Recall (%)	95.543	92.765	93.432	94.654
F-Measure (%)	95.432	92.876	93.654	94.876
MCC (%)	94.654	91.543	92.654	94.123
FAR (%)	12.543	20.654	18.543	15.654
FRR (%)	10.876	21.876	19.876	16.432
PSNR (dB)	32.654	30.876	31.765	30.876
Time complexity	9.123	16.654	15.876	10.543
MSE	0.0434	0.1034	0.0976	0.0897

Abbreviations: CNN: Convolutional neural network; FAR: False acceptance rate; FRR: False rejection rate; MCC: Matthews correlation coefficient; MSE: Mean squared error; PSNR: Peak signal-to-noise ratio.

The graphical analyses of MCC and MSE are presented in Fig. 5. The results showed that the MCC of the Fusion Net-3 model has a higher MCC value of 98.635% in comparison to CNN, ResNet-50, and GoogLeNet, which recorded 94.368%, 95.427%, and 96.784%, respectively, in Dataset 1. In Datasets 2 and 3, the MCC of the proposed model also demonstrated the highest values of 93.432% and 94.654%, respectively. These results indicate that Fusion Net-3 is superior to the models in previous research Akanfe et al. (2024). Meanwhile, the comparison results across the tested models in terms of MSE showed that, in Dataset 1, the Fusion Net-3 model achieved a lower rate, with a MSE value of 0.0234, compared to CNN (0.0612), ResNet-50 (0.0560), and GoogLeNet (0.0474). In Datasets 2 and 3, the proposed model's MSEs were 0.0201 and 0.0434, respectively. Although the error rate was comparatively low, some research Stolk & Sbrizzi (2019) reported even lower rates, highlighting the need for further research in this area.

4.1.1. Time complexity

The graphical analysis of time complexity is presented in Fig. 6. Time complexity is a measure of computing difficulty that characterizes how long a model takes to execute. The results revealed that the proposed model exhibited a lower time complexity of 5.324, compared to CNN (9.325), ResNet-50 (8.357), and GoogLeNet (7.368). In Datasets 2 and 3, the time complexity values of the proposed model were 5.987 and 9.123, respectively, indicating a higher efficiency than the other models. With its lower time complexity, along with superior precision and accuracy, the proposed model outperformed all other models in the

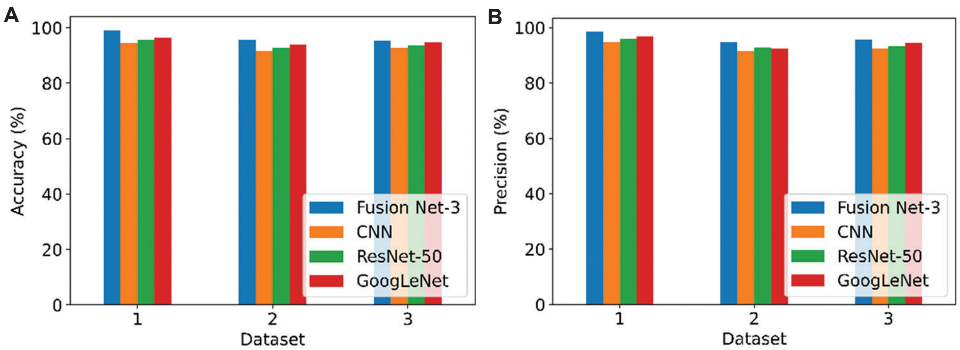


Fig. 3. Graphical representation of the (A) accuracy and (B) precision results
Abbreviation: CNN: Convolutional neural network

DOI: 10.6977/IJoSI.202508_9(4).0007
R. Sreemol, M. B. Santosh Kumar, etc./Int. J. Systematic Innovation, 9(4), 84-105 (2025)

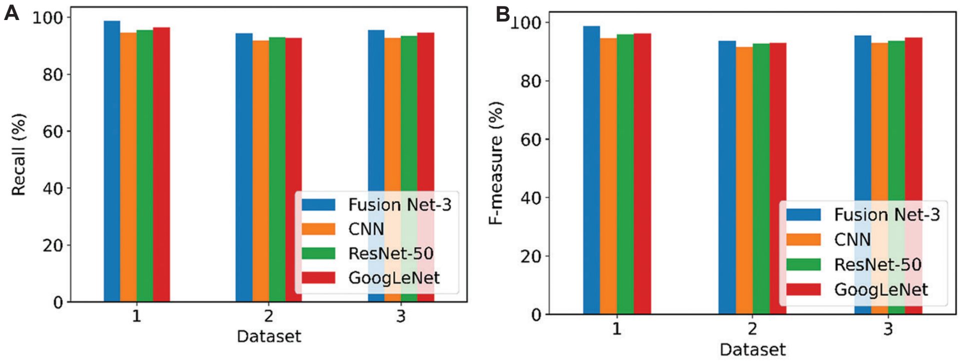


Fig. 4. Graphical representation of the (A) recall and (B) F-measure analysis
Abbreviation: CNN: Convolutional neural network

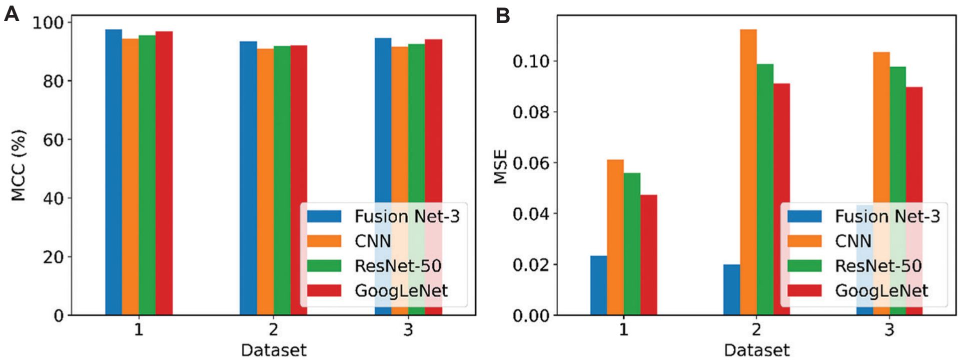


Fig. 5. Graphical representation of the (A) Matthews correlation coefficient (MCC) and (B) mean squared error (MSE) analysis
Abbreviation: CNN: Convolutional neural network

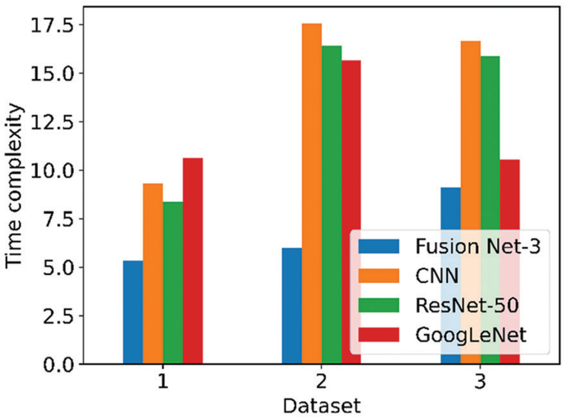


Fig. 6. Time complexity analysis
Abbreviation: CNN: Convolutional neural network

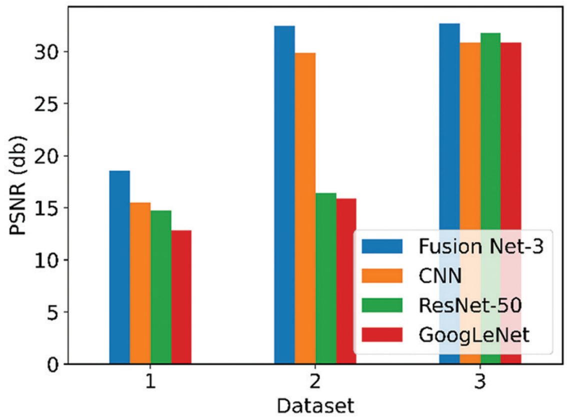


Fig. 7. Peak signal-to-noise ratio analysis
Abbreviation: CNN: Convolutional neural network

present study as well as in earlier research Arini et al. (2022) and Chopra & Ansari (2022).

The PSNR graphical analysis is displayed in Fig. 7. The results revealed that, in Dataset 1, the PSNR of the Fusion Net-3 model attained a higher value of 18.524 dB, compared to CNN (15.527 dB), ResNet-50 (14.753 dB), and GoogLeNet (12.864 dB). In Datasets 2 and 3, the PSNR values of the proposed model were 32.432 dB and 32.654 dB, respectively, suggesting better image clarity. Notably, previous research has reported lower PSNR values than the proposed model.

Fig. 8 shows the FAR and FRR of the tested models. The results showed that, in Dataset 1, the FAR of the Fusion Net-3 model exhibited a superior performance with an FAR of 0.5%, compared to CNN (1.2%), ResNet-50 (1.6%), and GoogLeNet (1.9%). In Datasets 2 and 3, the FARs of Fusion Net-3 were relatively lower at 12.987% and 12.543%, respectively.

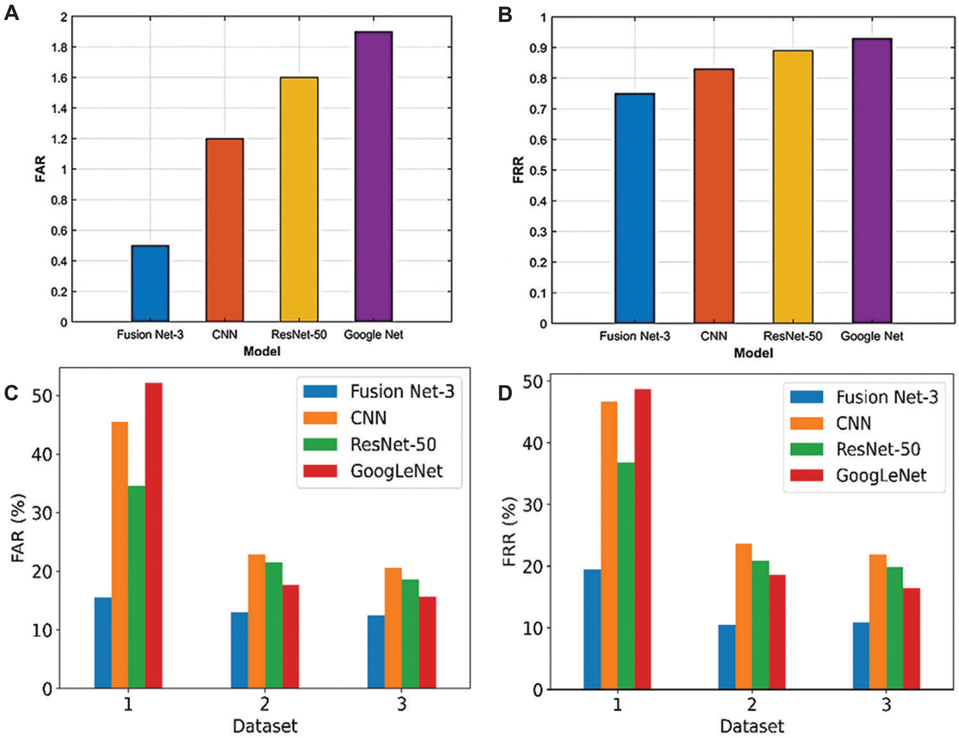


Fig. 8. Graphical representation of the (A) false acceptance rate and (B) false rejection rate by models
Abbreviation: CNN: Convolutional neural network

These comparatively low FAR results indicate a strong reliability of the proposed model. Besides, in Dataset 1, the FRR of the Fusion Net-3 model was 0.75%, lower than those of CNN (0.83%), ResNet-50 (0.89%), and GoogLeNet (0.93%). In Datasets 2 and 3, the FRRs of Fusion Net-3 were 10.543% and 10.876%, respectively. These findings suggest the reliability of the proposed model when compared with the other models.

4.1.2. Computational complexity

This section presents the performance of computational complexity in runtime measurements. As shown in Table 5, the Fusion Net-3 model demonstrates superior runtime performance compared to the other models. The higher flop performance indicates the superiority of the proposed model over the other models.

Table 5. Performance of flops across models

Model	Flops
CNN	634×10^{9}
ResNet-50	1.57×10^{9}
GoogLeNet	3.80×10^{9}
Fusion Net-3	4.20×10^{9}

Abbreviation: CNN: Convolutional neural network.

4.2. Statistical Analysis

The statistical analysis evaluates the performance and data balance of the Fusion Net-3 model using various statistical tests, such as the Mann–Whitney *U*-test, Kruskal–Wallis test, and chi-squared test. These tests were applied to a balanced dataset to assess whether the models exhibit statistically significant differences. Fig. 9A-C illustrates the results of the chi-squared test, Kruskal–Wallis test, and Mann–Whitney *U*-test, respectively.

Significant differences were observed between expected and observed image counts across categories, especially in the "Altered-easy," "Altered-medium," and "Real" categories. To ensure the proposed approach was trained and assessed on a balanced dataset, statistical tests were used to detect dataset imbalances. Table 6 depicts a comparison of the three statistical tests in terms of mean, error, and *p*-value.

With an error of 36,680.745 and p=0.0300, the results displayed a high Chi-squared statistic of 4,497,877.2. This result illustrates notable variations and potential biases in the procedures used to define categories or collect data. When two independent groups were compared using the Mann–Whitney *U*-test, the results showed a mean of 4,630,356.2, an error of 45,933.445, and p=0.0146. These results suggest a

DOI: 10.6977/IJoSI.202508_9(4).0007
R. Sreemol, M. B. Santosh Kumar, etc./Int. J. Systematic Innovation, 9(4), 84-105 (2025)

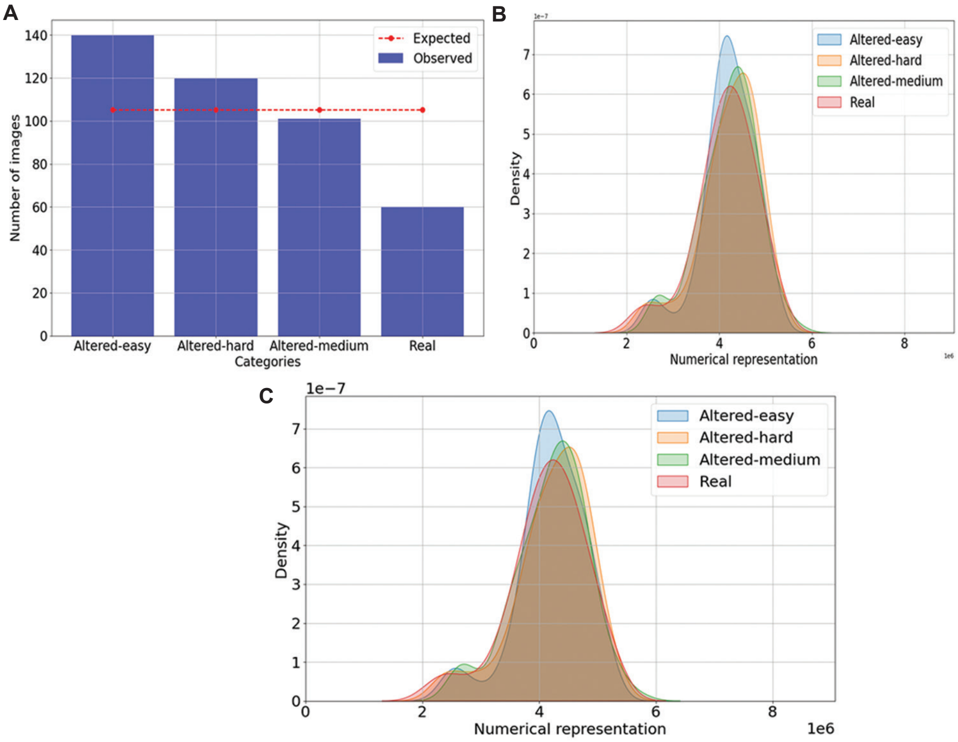


Fig. 9. Statistical analyses of the Fusion Net-3 model. (A) Chi-squared test. (B) Kruskal–Wallis test. (C) Mann–Whitney *U*-test

Table 6. Comparison of the statistical tests of Fusion Net-3

Statistical test	Mean	Error	*p*-value
Chi-square test	4,497,877.2	36,680.745	0.0300
Mann–Whitney *U*-test	4,630,356.2	45,933.445	0.0146
Kruskal–Wallis test	4,564,060.8	38,432.850	4.04×10^{-49}

significant difference between the groups compared, adding to the evidence supporting the effectiveness of the proposed Fusion Net-3 model in various scenarios. Notably, when comparing more than two groups, the Kruskal–Wallis test yielded a mean of 4,564,060.8, an error of 38,432.850, and an incredibly low p=4.04×10^{-49}. This remarkably low p-value indicates significant variations across the groups, underscoring the resilience and efficacy of the proposed model. The combined outcomes of these experiments indicate that the Fusion Net-3 model performs noticeably better than CNN, ResNet-50, and GoogLeNet, while effectively managing dataset imbalances.

4.3. Different Adversarial Attack Comparison

Tables 7-9 compare the Fusion Net-3 model with existing fingerprint authentication methods, focusing on accuracy as a common metric due to the lack of publicly available metrics, such as precision, recall, and MCC, in many baseline studies. Table 7 presents the robustness of fingerprint authentication models against adversarial attacks using the fast gradient sign method (FGSM). Fusion Net-3 exhibited the highest accuracy of 98.956%, surpassing CNN (94.562%), ResNet-50 (95.632%), and GoogLeNet (96.327%). The table also presents the adversarial samples' impact, where Fusion Net-3 demonstrated a minimal impact with an adversarial sample value of 0.37678, compared to CNN (0.32263), ResNet-50 (0.36442), and GoogLeNet (0.36010). These findings imply that Fusion Net-3 is more resistant to adversarial attacks, even in the presence of potential perturbations in fingerprint image input.

Table 8 details the accuracy obtained by fingerprint-based authentication models when subjected to adversarial attacks generated using the projected gradient descent (PGD) method. The proposed model reported an accuracy of 98.245%, outperforming CNN (92.654%), ResNet50 (93.432%), and GoogleNet (92.543%). The minimum difference was noted with Fusion Net-3 models at 0.38234, compared to larger differences recorded for CNN (0.32212), ResNet-50 (0.33345), and GoogLeNet (0.34123). These results indicate that Fusion Net-3 offers superior resilience to PGD attacks, retaining high accuracy with minimal degradation in the presence of adversarial perturbations.

Table 7. Accuracy and adversarial samples of models against black-box attack using fast gradient sign method

Dataset	Metric	Fusion Net-3	CNN	ResNet-50	GoogLeNet
1	Accuracy (%)	98.956	94.562	95.632	96.327
	Adversarial samples	0.37678	0.32263	0.36442	0.360100
2	Accuracy (%)	98.234	92.543	92.123	96.327
	Adversarial samples	0.36749	0.31122	0.35895	0.34921
3	Accuracy (%)	97.932	93.765	93.432	92.327
	Adversarial samples	0.39234	0.32865	0.36675	0.35472

Abbreviation: CNN: Convolutional neural network.

Table 8. Accuracy and adversarial samples of models against black-box attack using projected gradient descent

Dataset	Metric	Fusion Net-3	CNN	ResNet-50	GoogLeNet
1	Accuracy (%)	98.245	92.654	93.432	92.543
	Adversarial samples	0.38234	0.32212	0.33345	0.34123
2	Accuracy (%)	98.543	93.234	93.765	92.543
	Adversarial samples	0.37122	0.30643	0.31434	0.32344
3	Accuracy (%)	97.976	94.432	93.432	92.543
	Adversarial samples	0.36533	0.34245	0.33123	0.32875

Abbreviation: CNN: Convolutional neural network.

Table 9. Accuracy and adversarial samples of models against black-box attacks using a replay attack

Dataset	Metric	Fusion Net-3	CNN	ResNet-50	GoogLeNet
1	Accuracy (%)	97.542	92.643	93.234	92.123
	Adversarial samples	0.37455	0.30865	0.31235	0.32133
2	Accuracy (%)	97.654	93.765	92.876	92.123
	Adversarial samples	0.37546	0.33434	0.32675	0.32764
3	Accuracy (%)	98.123	92.876	93.76	92.123
	Adversarial samples	0.38123	0.34342	0.32450	0.31457

Abbreviation: CNN: Convolutional neural network.

Table 9 summarizes the replay attack resilience of the models. Replay attacks try to cheat the system using replayed biometric data captured previously. Fusion Net-3 demonstrated the highest performance with an accuracy of 97.542%, compared to CNN (92.643%), ResNet-50 (93.234%), and GoogLeNet (92.123%). The value of the adversarial sample of Fusion Net-3 was 0.37455, lower than CNN (0.30865), ResNet-50 (0.31235), and GoogLeNet (0.32133). The results indicate that Fusion Net-3 is effective in countering threats from adversarial replay attacks, making it a secure solution for fingerprint authentication.

4.4. Results in the Pre-Processing and Feature Selection Phases of the Models

Denoising performance was evaluated using the signal-to-noise ratio (SNR). The ratio measures the proportion of the image's valuable information relative to undesired artifacts or disruptions. Denoising the input images during the pre-processing phase revealed that the proposed improved bilateral filtering method achieved a higher SNR of 20.8 dB. In comparison, the median and Gaussian filters produced lower SNRs of 17.3 dB and 18.6 dB, respectively.

Feature selection using optimization, which is used for selecting the accurate features, demonstrated that the proposed Fusion Net-3 model outperformed existing algorithms by achieving a 98.12% performance. In comparison, the FOA and GJO achieved lower performances of 97.35% and 97.65%, respectively.

The accuracy of 98.956% achieved by the proposed model surpasses that of previous research Srinivasan et al. (2023), Trivedi et al. (2020), indicating its superiority over the other models used in this paper and those from previous works by other researchers. The enhanced bilateral filtering approach for denoising and the combination of CNN–ResNet-50 and U-Net features are some of the architectural enhancements that contribute to the superior performance of the proposed Fusion Net-3 model. Feature extraction and noise reduction are enhanced by these factors, resulting in increased accuracy and robustness. The model's capacity to preserve excellent image quality and prediction accuracy is further supported by the lower MSE and higher PSNR. In addition, the computational efficiency of the proposed method is also enhanced, with a time complexity of 5.324, demonstrating its usefulness in real-world applications.

The graphical analysis of receiver operating characteristic for the proposed and existing methods is shown in Fig. 10. Fusion Net-3 exhibited the highest area under the curve score (0.97), indicating its superior ability to balance reducing false positives (high specificity) with accurately identifying positive cases (high sensitivity).

The study's findings show that the proposed Fusion Net-3 model consistently outperforms other models, CNN, ResNet-50, and GoogLeNet, in a number of performance metrics. Specifically, Fusion Net-3 achieves an accuracy of 98.956%, surpassing CNN (94.562%), ResNet-50 (95.632%), and GoogLeNet (96.327%). It also exhibits higher precision (98.548%), recall (98.967%), and F-measure (98.675%) than the other models. In addition, Fusion Net-3 performs better than other models according to the MCC, which stands at 98.635%. This suggests a strong correlation between expected and actual outcomes. Furthermore, the proposed model outperforms the other models in its resilience to black-box attacks using FGSM, attaining an accuracy of 98.956% against adversarial samples.

The proposed model of Fusion Net-3 demonstrates an accuracy level of 98.956%, alongside enhanced security features that address crucial challenges in fingerprint-based authentication,

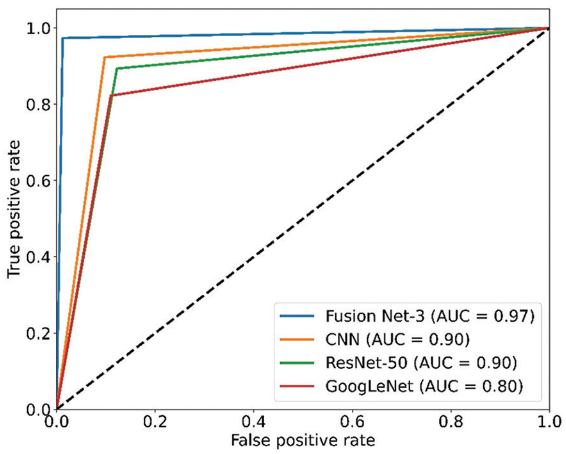


Fig. 10. Receiver operating characteristic curve analysis
Abbreviations: AUC: Area under the curve; CNN: Convolutional neural network

including noise reduction, feature extraction, and resistance to cyberattacks. In industries such as finance, healthcare, and national security, where the integrity of data and authenticity are important, this model would improve access control mechanisms, reduce fraud cases, and enhance operational efficiency. Moreover, the integration of blockchain technology in the model ensures secure data transmission, making it a viable solution for large-scale deployments in government and corporate settings. Furthermore, improved processing speed and low error rates also make it suitable for real-time applications, such as mobile authentication and border control systems, fostering wider acceptance of biometric security solutions. Such advances not only raise the confidence level of users in biometric systems but also pave the way for integrating them with newer technologies, such as Internet of Things-enabled smart environments and automated customer service kiosks.

The most significant limitation of this study is the variability in fingerprint image quality due to variations in acquisition devices, environment, and user-related inconsistencies. This may introduce noise and degrade the robustness of the model. In addition, adversarial attacks still pose a threat since an intelligent spoofing technique can potentially bypass the system with advanced security functionalities. To address these limitations, future work should expand the dataset diversity, improve computational efficiency, and further strengthen resilience against emerging attack vectors.

5. Conclusion

A fingerprint-based authentication system is a biometric authentication method that uses multiple techniques to enhance overall security and authentication accuracy. This study proposes the denoising-based Fusion Net-3 model to address the shortcomings of the currently existing methods. There are two stages to it: enrollment and authentication. The scanned information of hands is collected and pre-processed in the first phase using enhanced bilateral filtering, where the filter parameters are optimized using an SOA and the images are enhanced using a contrast enhancement technique. The pre-processed output is then used to extract features based on shapes and textures. Ultimately, a unique FIJO algorithm, a combination of the FOA and GJO algorithms, is used for feature selection to extract the optimal features. The selected features are combined using geometric mean and Fisher score. In the second phase, the fingerprint images are input and undergo pre-processing, feature extraction, and feature selection using similar methods utilized during the enrolment phase. The efficient (correct or incorrect) fingerprints are detected using the Fusion Net-3 model, which combines CNN, ResNet-50, and U-Net models. By implementing the proposed model on the Python platform, it achieved an accuracy of 98.956%, a precision of 98.548%, a recall of 98.967%, an F-measure of 98.675%, an MCC of 98.635%, a time complexity of 5.324, a PSNR of 18.524 dB, and an MSE of 0.0234. The Fusion Net-3 model outperforms the currently existing models based on these results. The proposed model is a classifier with high performance, but it has limitations, such as being influenced by poor-quality fingerprints and computational costs. Further work includes multimodal biometric systems and lightweight NNs in mobile or embedded systems, as well as investigating adversarial robustness and real-time spoof detection.

6. Acknowledgments

The authors would like to thank the Deanship of Cochin University of Science and Technology for their institutional support of this work.

References

Abolfathi, M., Inturi, S., Banaei-Kashani, F., & Jafarian, J.H. (2024). Toward enhancing web privacy on HTTPS traffic: A novel SuperLearner attack model and an efficient defense approach with adversarial examples. *Computers and Security*, 139, 103673. https://doi.org/10.1016/j.cose.2023.103673

Abolfathi, M., Shomorony, I., Vahid, A., & Jafarian, J.H. (2022). A game-Theoretically Optimal Defense Paradigm Against Traffic Analysis Attacks using Multipath Routing and Deception. In: *Proceedings of the 27th ACM Symposium on Access Control Models and Technologies*, p67–78. https://doi.org/10.1145/3532105.3535015

Adiga, V.S., & Sivaswamy, J. (2019). Fpd-m-net: Fingerprint image denoising and inpainting using m-net based convolutional neural networks. In: *Inpainting and Denoising Challenges*. Cham: Springer International Publishing, p51–61. https://doi.org/10.1007/978-3-030-25614-2_4

Afshari, H.H., Gadsden, S.A., & Habibi, S. (2017). Gaussian filters for parameter and state estimation: A general review of theory and recent trends. *Signal Processing*, 135, 218–238. https://doi.org/10.1016/j.sigpro.2017.01.001

Akanfe, O., Lawong, D., & Rao, H.R. (2024). Blockchain technology and privacy regulation: Reviewing frictions and synthesizing opportunities. *International Journal of Information Management*, 76, 102753. https://doi.org/10.1016/j.ijinfomgt.2024.102753

Akter, A., Nosheen, N., Ahmed, S., Hossain, M., Yousuf, M.A., Almoyad, M.A.A., et al. (2024). Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor. *Expert Systems with Applications*, 238, 122347. https://doi.org/10.1016/j.eswa.2023.122347

Algarni, M.H. (2024), Fingerprint sequencing: An authentication mechanism that integrates fingerprints and a knowledge-based methodology to promote security and usability. *Engineering, Technology and Applied Science Research*, 14(3), 14233–14239. https://doi.org/10.48084/etasr.7250

Ali, S.S., Baghel, V.S., Ganapathi, I.I., & Prakash, S. (2020). Robust biometric authentication system with a secure user template. *Image and Vision Computing*, 104, 104004. https://doi.org/10.1016/j.imavis.2020.104004

Arini, F.Y., Sunat, K., & Soomlek, C. (2022). Golden jackal optimization with joint opposite selection: An enhanced nature-inspired optimization algorithm for solving optimization problems. *IEEE Access*, 10, 28800–128823. Available from: https://www.kaggle.com/datasets/ruizgara/socofing

Balsiger, F., Jungo, A., Scheidegger, O., Carlier, P.G., Reyes, M., & Marty, B. (2020). Spatially regularized parametric map reconstruction for fast magnetic resonance fingerprinting. *Medical Image Analysis*, 64, 101741. https://doi.org/10.1016/j.media.2020.101741

Banitaba, F.S., Aygun, S., & Najafi, M.H. (2024). *Late Breaking Results: Fortifying Neural Networks: Safeguarding Against Adversarial Attacks with Stochastic Computing*. [arXiv Preprint].

Chopra, N., & Ansari, M.M. (2022). Golden jackal optimization: A novel nature-inspired optimizer for engineering applications. *Expert Systems with Applications*, 198, 116924.

Dhiman, G., & Kumar, V. (2019). Seagull optimization algorithm: Theory and its applications for large-scale industrial engineering problems. *Knowledge-Based Systems*, 165, 169–196. https://doi.org/10.1016/j.knosys.2018.11.024.

Ding, B., Wang, H., Chen, P., Zhang, Y., Guo, Z., Feng, J., et al. (2020). Surface and internal fingerprint reconstruction from optical coherence tomography through convolutional neural network. *IEEE Transactions on Information Forensics and Security*, 16, 685–700. https://doi.org/10.1109/TIFS.2020.3016829

Ephin, M., & Vasanthi, N.A. (2013). A highly secure integrated biometrics authentication using finger-palmprint fusion. International Journal of Scientific and Engineering Research, 4(1).

Galbally, J., Beslay, L., & Böstrom, G. (2020).

3D-FLARE: A touchless full-3D fingerprint recognition system based on laser sensing. *IEEE Access*, 8, 145513–145534. https://doi.org/10.1109/ACCESS.2020.3014796

Gao, Z., Gao, Y., Wang, S., Li, D., Xu, Y. (2020). CRISLoc: Reconstructable CSI fingerprinting for indoor smartphone localization. *IEEE Internet of Things Journal*, 8(5), 3422–3437. https://doi.org/10.1109/JIOT.2020.3022573

Gavaskar, R.G., & Chaudhury, K.N. (2018). Fast adaptive bilateral filtering. *IEEE Transactions on Image Processing*, 28(2), 779–790. https://doi.org/10.1109/TIP.2018.2871597

Gupta, R., Khari, M., Gupta, D., & Crespo, R.G. (2020). Fingerprint image enhancement and reconstruction using the orientation and phase reconstruction. *Information Sciences*, 530, 201–218. https://doi.org/10.1016/j.ins.2020.01.031

Husson, L., Bodin, T., Spada, G., Choblet, G., & Kreemer, C. (2018). Bayesian surface reconstruction of geodetic uplift rates: Mapping the global fingerprint of Glacial Isostatic Adjustment. *Journal of Geodynamics*, 122, 25–40. https://doi.org/10.1016/j.jog.2018.10.002

Jia, H., Xing, Z., & Song, W. (2019). A new hybrid seagull optimization algorithm for feature selection. *IEEE Access*, 7, 49614–49631. https://doi.org/10.1109/ACCESS.2019.2909945

Kareem, S.W., & Okur, M.C. (2021). Falcon optimization algorithm for bayesian network structure learning. *Computer Science*, 22, 553–569. https://doi.org/10.7494/csci.2021.22.4.3773

Khodadoust, J., Khodadoust, A.M., Mirkamali, S.S., & Ayat, S. (2020). Fingerprint indexing for wrinkled fingertips immersed in liquids. *Expert Systems with Applications*, 146, 113153. https://doi.org/10.1016/j.eswa.2019.113153

Koonce, B., & Koonce, B.E. (2021). *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*. Apress, New York, USA. https://doi.org/10.1007/978-1-4842-6168-2

Lee, S.H., Kim, W.Y., & Seo, D.H. (2022). Automatic self-reconstruction model for radio map in Wi-Fi fingerprinting. *Expert Systems with Applications*, 192, 116455. https://doi.org/10.1016/j.eswa.2021.116455

Li, J., Feng, J., & Kuo, C.C.J. (2018). Deep convolutional neural network for latent fingerprint enhancement. *Signal Processing: Image Communication*, 60, 52–63. https://doi.org/10.1016/j.image.2017.08.010

Li, Y., Xia, Q., Lee, C., Kim, S., & Kim, J. (2022). A robust and efficient fingerprint image restoration method based on a phase-field model. *Pattern Recognition*, 123, 108405. https://doi.org/10.1016/j.patcog.2021.108405

Liang, Y., & Liang, W. (2023). *ResWCAE: Biometric Pattern Image Denoising Using Residual Wavelet-Conditioned Autoencoder*. [Preprint]. https://doi.org/10.1007/978-981-96-6603-4_17

Lin, C., & Kumar, A. (2018). Contactless and partial 3D fingerprint recognition using multi-view deep representation. *Pattern Recognition*, 83, 314–327. https://doi.org/10.1016/j.patcog.2018.05.004

Liu, F., Kong, Z., Liu, H., Zhang, W., & Shen, L. (2022). Fingerprint presentation attack detection by channel-wise feature denoising. *IEEE Transactions on Information Forensics and Security*, 17, 2963–2976. https://doi.org/10.1109/TIFS.2022.3197058

Liu, F., Liu, G., Zhao, Q., & Shen, L. (2020a). Robust and high-security fingerprint recognition system using optical coherence tomography. *Neurocomputing*, 402, 14-28. https://doi.org/10.1016/j.neucom.2020.03.102

Liu, F., Liu, H., Zhang, W., Liu, G., & Shen, L. (2021). One-class fingerprint presentation attack detection using auto-encoder network. *IEEE Transactions on Image Processing*, 30, 2394–2407. https://doi.org/10.1109/TIP.2021.3052341.

Liu, F., Shen, C., Liu, H., Liu, G., Liu, Y., Guo, Z., et al. (2020b). A flexible touch-based fingerprint acquisition device and a benchmark database using optical coherence tomography. *IEEE Transactions on Instrumentation and Measurement*, 69(9), 6518–6529. https://doi.org/10.1109/TIM.2020.2967513

Mahum, R., Irtaza, A., Nawaz, M., Nazir, T., Masood, M., Shaikh, S., et al. (2023). A robust framework to generate surveillance video summaries using combination of zernike moments and r-transform and a deep neural network. *Multimedia Tools and Applications*, 82(9), 13811–13835. https://doi.org/10.1007/s11042-022-13773-4

Narodytska, N., & Kasiviswanathan, S.P. (2017). Simple Black-Box Adversarial Attacks on Deep Neural Networks. In: *CVPR Workshops*. Vol. 2. Available from: https://openaccess.thecvf.com/content_cvpr_2017_workshops/w16/papers/kasiviswanathan_simple_black-box_adversarial_cvpr_2017_paper.pdf

Nasri, M., Kosa, M., Chukoskie, L., Moghaddam, M., & Harteveld, C. (2024). Exploring Eye Tracking to Detect Cognitive Load in Complex Virtual Reality 1 Training. In: *2024 IEEE International Symposium on Mixed and Augmented Reality*

Adjunct (ISMAR-Adjunct). IEEE, p51–54. https://doi.org/10.1109/ISMAR-Adjunct64951.2024.00022

Paris, S., Kornprobst, P., Tumblin, J., & Durand, F. (2009). Bilateral filtering: Theory and applications. *Foundations and Trends® in Computer Graphics and Vision*, 4(1), 1–73. https://doi.org/10.1561/0600000020

Praseetha, V.M., Bayezeed, S., & Vadivel, S. (2019). Secure fingerprint authentication using deep learning and minutiae verification. *Journal of Intelligent Systems*, 29(1), 1379–1387. https://doi.org/10.1515/jisys-2018-0289

Prybylo, M., Haghighi, S., Peddinti, S. T., & Ghanavati, S. (2024b). *Evaluating privacy perceptions, experience, and behavior of software development teams. USENIX*. Available from: https://www.usenix.org/conference/soups2024/presentation/prybylo

Rahman, M.M., Mishu, T.I., & Bhuiyan, M.A.A. (2022). Performance analysis of a parameterized minutiae-based approach for securing fingerprint templates in biometric authentication systems. *Journal of Information Security and Applications*, 67, 103209. https://doi.org/10.1016/j.jisa.2022.103209

Santos, S., Breaux, T., Norton, T., Haghighi, S., Ghanavati, S. (2024). *Requirements Satisfiability with in-Context Learning*. [arXiv Preprint]. https://doi.org/10.1109/RE59067.2024.00025

Shadab, S.A., Ansari, M.A., Singh, N., Verma, A., Tripathi, P., & Mehrotra, R. (2022). Detection of Cancer from Histopathology Medical Image Data Using ML with CNN ResNet-50 Architecture. In: *Computational Intelligence in Healthcare Applications*. Academic Press, United States, p237–254 https://doi.org/10.1016/B978-0-323-99031-8.00007-7

Shehu, Y.I., Ruiz-Garcia, A., Palade, V., & James, A. (2018). *Sokoto Coventry Fingerprint Dataset.* [arXiv Preprint]. https://doi.org/10.48550/arXiv.1807.10609

Srinivasan, D.S., Ravichandran, S., Indrani, T.S., & Karpagam, G.R. (2023). Local Binary Pattern-Based Criminal Identification System. In: *Sustainable Digital Technologies for Smart Cities*. United States: CRC Press. p45–56. Available from: https://www.taylorfrancis.com/chapters/edit/10.1201/9781003307716-4/local-binary-pattern-based-criminal-identification-system-dhana-srinithi-srinivasan-soundarya-ravichandran-thamizhi-shanmugam-indrani-karpagam

Stolk, C.C., & Sbrizzi, A. (2019). Understanding the combined effect of ${k} $-space undersampling and transient states excitation in MR fingerprinting reconstructions. *IEEE Transactions on Medical Imaging*, 38(10), 2445–2455. https://doi.org/10.1109/TMI.2019.2900585

Trivedi, A.K., Thounaojam, D.M., & Pal, S. (2020). Non-invertible cancellable fingerprint template for fingerprint biometrics. *Computers and Security*, 90, 101690. https://doi.org/10.1016/j.cose.2019.101690

Vogel, R.M. (2022). The geometric mean? *Communications in Statistics-Theory and Methods*, 51(1), 82–94. https://doi.org/10.1080/03610926.2020.1743313

Wang, D., Ostenson, J., & Smith, D.S. (2020). snapMRF: GPU-accelerated magnetic resonance fingerprinting dictionary generation and matching using extended phase graphs. *Magnetic Resonance Imaging*, 66, 248–256. https://doi.org/10.1016/j.mri.2019.11.015

Wang, W., & Yang, Y. (2024). A histogram equalization model for color image contrast enhancement. *Signal, Image and Video Processing*, 18(2), 1725–1732. https://doi.org/10.1007/s11760-023-02881-9

Wong, W.J., & Lai, S.H. (2020). Multi-task CNN for restoring corrupted fingerprint images. Pattern Recognition, 101, 107203. https://doi.org/10.1016/j.patcog.2020.107203

Xu, Z., Ye, H., Lyu, M., He, H., Zhong, J., Mei, Y. (2019). Rigid motion correction for magnetic resonance fingerprinting with sliding-window reconstruction and image registration. *Magnetic Resonance Imaging*, 57, 303–312. https://doi.org/10.1016/j.mri.2018.11.001

Yin, X., Zhu, Y., & Hu, J. (2019). 3D fingerprint recognition based on ridge-valley-guided 3D reconstruction and 3D topology polymer feature extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(3), 1085–1091. https://doi.org/10.1109/TPAMI.2019.2949299

Zhang, L., Tan, T., Gong, Y., & Yang, W. (2019). Fingerprint database reconstruction based on robust PCA for indoor localization. *Sensors (Basel)*, 19(11), 2537.

Zheng, H., Gao, M., Chen, Z., Liu, X.Y., & Feng, X. (2019). An adaptive sampling scheme via approximate volume sampling for fingerprint-based indoor localization. *IEEE Internet of Things Journal*, 6(2), 2338–2353. https://doi.org/10.1109/jiot.2019.2906489

AUTHOR BIOGRAPHIES

R. Sreemol received her Bachelor's degree in Computer Science and Engineering in 2014 and her Master's degree in Computer Science and Engineering in 2017. She earned her Ph.D. from the Department of Computer Applications, Cochin University of Science and Technology, South Kalamassery, Kochi, Kerala, India. Her research interests include pattern recognition and image processing, with applications in biometrics and security systems

M. B. Santosh Kumar received his Ph.D. in 2018 from Cochin University of Science and Technology, South Kalamassery, Kochi, Kerala, India. His dissertation was titled Development of Innovative Procedures for Information Technology Articulated Agriculture. He is a Professor in the Division of Information Technology, School of Engineering, CUSAT. He was granted a patent for a Portable Agriculture Network System in 2017. His research interests focus on knowledge-based systems in agriculture and artificial intelligence applications in agriculture

A. Sreekumar received his Ph.D. in Cryptography from Cochin University of Science and Technology, South Kalamassery, Kochi, Kerala, India, in 2010. He is a retired Professor from the Department of Computer Applications, CUSAT, with 27 years of teaching experience. His research interests include cryptography, number theory, and secret sharing.

Appendix

Appendix A1. Dataset details

Dataset 1: Six thousand fingerprints from 600 African participants make up the Sokoto Coventry Fingerprint Dataset (SOCOFing; https://www.kaggle.com/datasets/ruizgara/socofing; Xiao et al., 2019). The experimental settings and datasets used were consistent across all models. Each participant, all of whom were at least 18 years old, submitted 10 fingerprints. Gender designations and hand and finger names are included in SOCOFing. The STRANGE toolbox was utilized to generate synthetic modifications of these fingerprints, including three distinct levels of obliteration, central rotation, and *z*-cut change. The SDU03PTM sensor (SecuGen, USA) and Hamster Plus sensor (HSDU03PTM, SecuGen, USA) scanners were used to acquire the original images.

Dataset 2: FVC2002 fingerprints (https://www.kaggle.com/datasets/nageshsingh/fvc2002-fingerprints). These datasets were chosen due to their inclusion of benchmarked and standard fingerprints collected from various sensors, displaying various characteristics.

Dataset 3: Fingerprint Dataset for FVC2000_DB4_B (https://www.kaggle.com/datasets/peace1019/fingerprint-dataset-for-fvc2000-db4-b). A collection of fingerprint photos utilized for fingerprint recognition studies. This fingerprint dataset can also be used for data augmentation activities. It comprises 800 excellent fingerprint photographs, each measuring 160 × 160 pixels and having a resolution of 500 DPI.

Reference

Xiao, J., Hu, F., Shao, Q., & Li, S. (2019). A low-complexity compressed sensing reconstruction method for heart signal biometric recognition. *Sensors (Basel)*, 19(23), 5330. https://doi.org/10.3390/s19235330

DOI: 10.6977/IJoSI.202508_9(4).0008
N. U. Khan, R. Riaz, etc./Int. J. Systematic Innovation, 9(4), 106-122 (2025)

Enhanced group recommendation system: A hybrid context-aware approach with collaborative filtering for location-based social networks

Naimat Ullah Khan[1]*, Rubina Riaz[2], Jutao Hao[1], Xianzhi Wang[1], Huan Huo[1]

[1]School of Computer Science, Faculty of Engineering and IT, University of Technology Sydney, Ultimo, Australia
[2] School of Software Engineering, Dalian University of Technology, Dalian, China

*Corresponding author E-mail: naimatullah.khan@uts.edu.au

(Received 14 April 2025; Final version received 16 June 2025; Accepted 23 June 2025)

Abstract

In recent years, location-based social networks (LBSNs) have gained significant popularity, enabling users to interact with points of interest (POIs) using modern technologies. As more people rely on LBSNs for finding interesting venues, contextually aware and relevant recommendation systems have become very beneficial with practical applications. In this research, we propose an enhanced hybrid recommendation system, designed for LBSNs to improve the accuracy of suggestions by integrating collaborative filtering methods with singular value decomposition to handle sparse data, along with context-aware modeling to tailor recommendations based on user interests, and group recommendation to accommodate multi-user scenarios. In addition, we incorporate contextual aspects, such as spatial proximity and temporal behavior, into the model to ensure recommendations align closely with the user's present surroundings and preferences. The proposed method extends further to group recommendations by considering individual inclinations into cohesive suggestions for groups interested in visiting POIs together. The proposed method is assessed using precision, recall, and F1 score, ensuring a thorough evaluation of its performance. To further highlight context-aware recommendations, we use clustering based on user preference, temporal behavior, and category-wise interaction to identify patterns across various venue types. The proposed method shows improved recommendations, specifically based on data from LBSNs, and develops an efficient solution for balanced user preferences with contextual influences.

Keywords: Collaborative Filtering, Context-Aware Recommendations, Data Sparsity, Group Recommendation, Hybrid Recommendation System, Location-Based Social Networks, Singular Value Decomposition

1. Introduction

Location-based social networks (LBSNs) have become an integral part of daily life, largely due to the widespread use of smartphones and global positioning technology. Platforms such as Weibo, Gowalla, Foursquare, Yelp, and Google Maps allow users to share their locations, check in at various places, and connect with friends. These networks not only track a user's movements but also create a dynamic system in which individuals interact with their physical environment and social networks in meaningful ways (Barai & Bhaumik, 2015). An opportunity and the big challenge are developing recommendations that are more relevant by suggesting points of interest (POIs) – places the user might like based on where they have been, who they know, and what activities the users enjoy frequently – with the help of methods such as content-based filtering and collaborative filtering (CF) traditional recommendation systems, such as those used by Netflix or Amazon, are effective in domains including movies and product recommendations due to well-established preference patterns. However, such models are not directly suitable for LBSNs, which involve more nuanced spatial, social, and temporal dynamics (Ezin, 2024). However, these methods cannot be implemented directly when it comes to location-based recommendations, which involve more complicated factors such as location, social, or temporal aspects related to the users' check-ins. When recommending a venue, systems now consider much more than just general past preferences; factors such as physical proximity and typical visitation times for similar venues are taken into consideration. It is a much more nuanced approach to helping people discover interesting places nearby. These approaches represent a significant shift in how technology understands and anticipates our daily experiences, transforming simple check-in data into sophisticated, personalized recommendation models.

1.1. Challenges in LBSN Recommendations

One of the significant challenges in LBSN recommendation systems is the critical influence of user proximity to other users in the physical world (Dutta et al., 2025). Unlike previous recommendation systems that focused on item features such as type, price, or characteristics, LBSN users tend to frequent POIs located near their current location (Sánchez & Bellogín, 2022). Location plays a pivotal role in generating meaningful recommendations. Previous research has demonstrated that individuals typically travel from their current location before heading to the recommended venue, underscoring the importance of spatial data in recommendation algorithms. A recommendation system might suggest appealing places, but its utility diminishes if it fails to consider geographical distance (Dietz et al., 2025). Beyond geographical considerations, temporal parameters are equally crucial in understanding user decision-making. Individuals' preferences vary throughout the day and week. For example, the users might seek cafés in the morning, restaurants in the evening, or parks on weekends. Therefore, LBSN recommendations must incorporate time-based factors that reflect users' behavioral patterns during specific periods (Zhang et al., 2019). These temporal elements are not merely supplementary but fundamental to recommendation accuracy. Ignoring time-related context can result in inappropriate suggestions that feel disconnected from users' actual behavioral patterns, fundamentally compromising the system's functionality and relevance (Redondo et al., 2020; Zheng & Zhou, 2024). The major challenge lies in developing recommendation systems that can integrate spatial and temporal dimensions, providing suggestions that are not just potentially interesting but practically accessible within a user's context.

However, user preferences and social factors present additional complexities in LBSN recommendation systems. Empirical research demonstrates that individual historical behaviors, such as previous check-ins, repeated visits, and consistent location choices, significantly influence preferred visiting venues (Teoman, 2022; Wachyuni & Kusumaningrum, 2020). Past preference patterns dramatically shape venue selections. Users with a history of social interactions tend to gravitate toward vibrant, interactive venues, such as bars and festivals. On the other hand, those with more solitary past behaviors prefer more contemplative spaces, such as libraries and parks. Location-based data provides significant opportunities to enhance recommendation accuracy by integrating contextual factors, including user location, temporal variables, and local events. Individual preferences can vary substantially between travel and home environments, with local conditions and events further modulating decision-making patterns. The use of this contextual information becomes crucial in creating more accurate recommendations. By combining these factors, recommendation systems can generate suggestions that are aligned with users' circumstances and past preferences. The fundamental challenge lies in developing recommendation systems that can analyze and model the relationship between users' past behaviors, spatial context, and temporal dynamics (Wang et al., 2024).

1.2. Hybrid Recommendation Systems For LBSNs

The complex nature of user preferences in LBSNs makes it crucial to explore hybrid recommendation systems to get more relevant suggestions. In rapidly changing environments like LBSNs, hybrid systems can effectively address the limitations of individual methods by combining the strengths of various recommendation approaches, ultimately providing more accurate recommendations. Most conventional recommendation approaches, such as CF, are typically designed to predict a specific user's preference based on their past activities. However, these traditional methods often fall short in capturing the complex nature of user interactions within the location-based systems. Hybrid recommendation systems offer a more comprehensive solution by integrating multiple recommendation techniques, allowing for a more nuanced and adaptable approach to understanding user preferences. By integrating different methodological strategies, these systems can provide more robust and context-aware recommendations that better reflect the multilayered user behavior in LBSNs (Eliyas & Ranjana, 2022). There are two main types of memory-based CF approaches (Teoman, 2022): User-based CF, which recommends POIs based on preferences of similar users, and item-based CF, which suggests POIs similar to those the user has previously visited. Despite their effectiveness, individual methods often struggle with data sparsity and lack contextual adaptability, both of which are critical for LBSNs. Given the complex nature of user preferences in LBSNs, hybrid recommendation systems have emerged as a promising solution, integrating CF with singular value decomposition (SVD) and context-aware techniques to offer more accurate and personalized recommendations. In this study, we propose a contextual hybrid recommendation strategy that integrates user preferences, geographical location, and temporal behavior factors with CF and SVD matrix factorization to address the unique challenges of LBSNs. Using standard performance metrics, such as precision, recall, and F1 score, we assess the quality of the proposed system's recommendations. This serves as a powerful dimensionality reduction technique that enables the identification of hidden patterns within user-item interaction data that might otherwise remain imperceptible (Wachyuni & Kusumaningrum, 2020). While SVD improves recommendation quality and addresses data sparsity issues, it does not consider contexts, such as spatial or temporal factors, that are essential in LBSN recommendation systems.

Key contributions to this research:

(i) This research introduces a novel hybrid recommendation framework that integrates user-based, item-based CF with SVD, and context-aware user modeling. This combination is specifically designed to address and mitigate the limitations found in traditional recommendation approaches within LBSNs, enabling more accurate and context-sensitive recommendation strategies.

(ii) By incorporating factors such as spatial proximity and temporal behaviors, our framework enhances the personalization of recommendations. This integration allows for recommendations that adapt to varying user contexts, aligning suggestions with location-based and temporal dimensions to closely reflect diverse user needs and preferences, accumulating to group recommendations.

(iii) Through extensive experimentation with standard evaluation metrics (precision, recall, F1 score, and accuracy), we demonstrate that the proposed hybrid model outperforms traditional methods. This thorough assessment highlights the framework's potential to maintain high recommendation quality and accuracy in data-sparse and dynamic environments, marking a significant improvement over the conventional method in LBSN settings.

The next section of this paper presents a review of the relevant literature, followed by Section 3, which describes the material and methods, including the detailed steps in the architecture of the proposed system. Section 4 provides the results and comparisons, followed by the conclusion and potential future directions.

2. Related Work

The research on efficient recommendation engines is becoming increasingly important as LBSNs emerge as a primary platform for real-world user interactions. LBSNs present unique recommendation challenges by integrating contextual factors such as user location, check-in timing, and personal preferences, extending beyond traditional user-item interaction models. The CF approaches are widely used in most recommender systems, excel at generating similar-based recommendations tailored to user behaviors (Papadakis et al., 2022). However, the contextual aspects inherent in LBSNs, when the user preferences dynamically shift with location and time, present significant limitations for traditional CF techniques. The complexity of LBSNs has motivated research into integrating contextual information to enhance recommendation accuracy and relevance (Mahajan & Kaur, 2023). While CF techniques have been successfully applied across domains, including media streaming and e-commerce (Hu et al., 2019), their direct application to location-based networks remains challenging. Matrix factorization methods, including SVD, have improved CF effectiveness by reducing user-item interaction matrix dimensionality and uncovering hidden factors. Nevertheless, the data sparsity problem in LBSNs, where users visit few POIs, creates substantial gaps in interaction matrices (Tourinho & Rios, 2021). These constraints highlight the need to develop innovative recommendation methodologies that can effectively capture real-world contextual features and manage sparse data environments.

To address the limitations of conventional CF for large-scale social networks, context-aware recommendation systems have been proposed in a previous study (Ezin, 2024). These systems integrate external contextual factors, such as user time and location, into recommendation mechanisms. LBSN users typically prefer locations spatially closer to their current position, making geo-influence a critical system component. As demonstrated by Yuan & Chen (2017), users visit different places at varying times, establishing temporal factors as essential in LBSN recommendations. Some studies have expanded contextual considerations beyond spatial and temporal parameters. For example, individual personality differences significantly impact location preferences: Extroverted individuals might seek active environments, while introverted users prefer calmer settings (Deldjoo et al., 2020). For instance, people with different personality traits, such as extroverted individuals, typically prefer vibrant environments, while introverted people seek serene settings. Previous studies by Hossein (2018) and Arabi (2018) discuss that integrating personality traits can enhance recommendation targeting, potentially improving overall user satisfaction in LBSNs. While exploring recommendations for LBSN, considering factors such as proximity might enhance user satisfaction.

LBSN research demonstrates that social connections are a valuable source of information for improving recommendation quality. Social network analysis, as discussed by Wang et al. (2019), can be integrated into LBSNs to leverage user relationships, particularly in scenarios with limited user-item interaction data. Wang et al. (2019) suggested that users' preferences are significantly influenced by their social circles, friends, coworkers, and contacts, which can inform and shape location choices. Social trust-based models, introduced by Kanfade et al. (2018) and Bhaumik (2016), extend traditional CF by incorporating trust scores between users, enabling more nuanced preference generalization. Wang et al. (2013) proposed a circle-based recommendation system that generates suggestions by grouping friends into distinct circles, focusing on preferences within these social networks. Recent developments in recommender systems increasingly recognize that multiple factors beyond traditional user-item interactions shape user preferences in LBSNs. Recently, decision-making in these systems incorporates multi-criteria considerations, such as cost, ambiance, accessibility, and user reviews. Multi-criteria rating systems enable more complex, detailed recommendations compared to traditional CF techniques (Dadoun et al., 2019). Unlike the conventional CF approach by Nian (2021), multi-criteria CF (Davtalab & Alesheikh, 2023) extracts multiple dimensions of user preferences to capture the trade-offs users consider when selecting locations. In LBSNs, this approach allows users to prioritize criteria such as distance, reflecting the nature of location-based decision-making.

Traditional CF methods do not consider handling contextual features, leading to the development of hybrid recommendation approaches that combine multiple strategies to enhance overall performance (Zheng, 2022). These hybrid systems integrate advantages from CF, context-aware modules, and content-based strategies to mitigate limitations in individual methods. CF and context-aware approaches, including SVD, are frequently used to capture complex interactions between users and items. This approach is especially vital in LBSNs where user-item interactions are typically sparse and latent factors can mask user-POI relationships (Sun et al., 2022). The increasing trend of collaborative destination planning has shifted focus from individual to group recommendation systems (Zhao et al., 2023). Group recommendation systems face the complex challenge of satisfying multiple users' preferences, often complicated by conflicting individual desires. Various grouping methods have emerged to address these challenges, including social preference analysis and voting processes. These aggregation techniques aim to compute group-level recommendations while carefully balancing individual preferences (Zhou et al., 2024). By incorporating contextual and social aspects, group recommendation algorithms become more sophisticated in suggesting relevant choices for a group along with various venue categories.

The advancement in hybrid systems that integrate CF and social trust modeling has enhanced personalization that is hindered by the limitations of traditional CF in handling sparse data and contextual factors, such as time and location in LBSNs. However, hybrid systems still face challenges in combining contextual and social dimensions, especially for group recommendations. In line with these findings, it is necessary to propose a hybrid recommendation framework that not only leverages CF and SVD for handling data sparsity but also integrates contextual factors, including time and location. In addition, the inclusion of group recommendation methods ensures that our system can balance individual preferences in multi-user scenarios, offering a scalable and contextually adaptive solution for LBSNs.

3. Research Methodology

This section presents the methodology and framework for designing the LBSN group recommendation system. The framework in Fig. 1 represents the proposed hybrid approach designed to generate highly personalized and contextually relevant recommendations in LBSNs. To demonstrate the effectiveness of the model, we used the user check-in data from an LBSN named Gowalla (Zhou et al., 2024). The user-influence modeling component built a detailed profile for each user by analyzing key factors that contribute to a long-term understanding of user preferences. Parallelly, CF was employed, encompassing user-based CF, item-based CF, and SVD. This combination allowed the framework to uncover hidden relationships within user-item interactions, making it particularly effective in sparse data environments. In addition, the context-aware module dynamically adjusts recommendations based on the user's immediate context, including their current location and time, ensuring relevance.

Once the system had gathered insights from user influence modeling, CF, and context-aware modules, it combined these in the recommendation

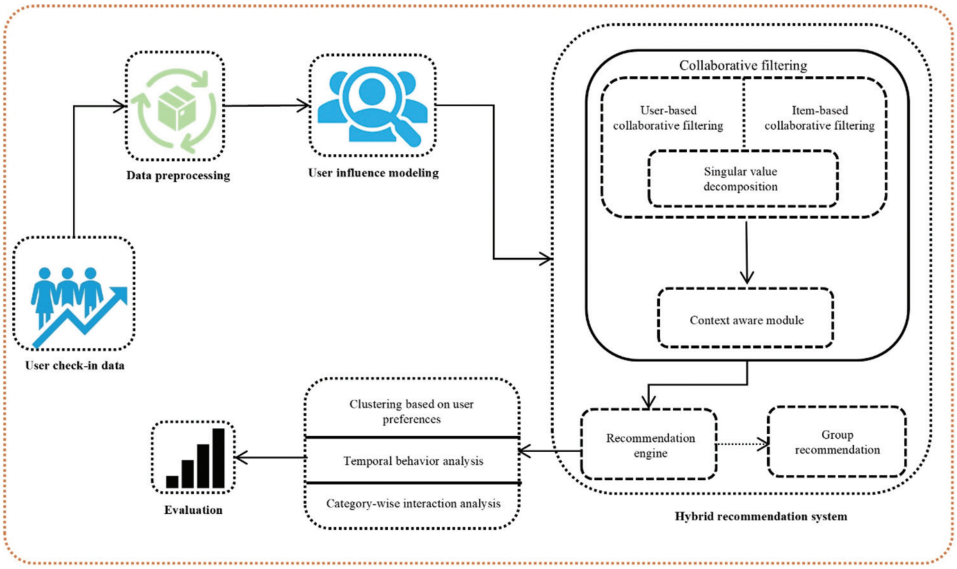


Fig. 1. Proposed framework of LBSN group recommendation system

engine to generate a final set of recommendations tailored to the individual. For group scenarios, the group recommendation module compiled individual preference scores to provide unified recommendations that met the collective needs of multiple users, balancing diverse tastes within the group. Finally, the evaluation component measured the system's performance using metrics, such as precision and recall, along with analyses based on user preferences, temporal behaviors, and category-specific interactions. Together, these components formed a robust, scalable recommendation system that effectively balanced long-term personalization with adaptive, recommendation contextual adjustments, resulting in recommendations that were both accurate and highly relevant to the user's preferences.

3.1. Hybrid Recommendation System For LBSN

The hybrid recommendation system integrates multiple recommendation techniques to provide highly personalized and contextually relevant recommendations for LBSNs. The dataflow diagram in Fig. 2 shows the steps to achieve this goal. The first step involved preprocessing the user-item interaction matrix and extracting contextual information of the user, such as geographical location and temporal behavior. These contextual factors were key to understanding the user's preferences more comprehensively. CF methods were applied to predict scores for POIs, as shown in Eq. (1). We computed the final recommendation score for each user-item pair by integrating user-based CF, item-based CF, and SVD using a weighted sum, as follows:

$$F_{final} = w_1 \cdot F_{up} + w_2 \cdot F_{ip} + w_3 \cdot F_{svd} \tag{1}$$

where F_{final}: final predicted score, F_{up}: user-based CF score, F_{ip}: item-based CF score, F_{svd}: score from SVD, and w_1, w_1, w_1: weights for each method.

In our hybrid approach, the weights assigned to each component method (user-based CF, item-based CF, and SVD) were initially set based on empirical tuning, with equal or proportional to the assigned values derived from preliminary experiments. This initialization provided a balanced integration of the three methods, ensuring that the system did not overly depend on any single approach in the beginning. However, to adapt to our datasets, the weights were fine-tuned during training, considering characteristics such as data sparsity and interaction density. This weighting approach demonstrated the potential to improve recommendation accuracy by enabling the system to respond to various user behaviors and contexts.

User-based CF searched for similar users to the target user and utilized their preferences to generate recommendations, while the item-based CF searched for items that the user had interacted with. The SVD was used to uncover latent relationships between users and items, helping in cases where the data were sparse. The Fig. 2 incorporates contextual information to adjust the recommendations, making them more relevant based on the user's preferences. The final scores were calculated by combining the scores from user-based CF, item-based CF, SVD, and context-aware modeling, using weighted averages. If the recommendation was intended for a group, the preferences of all group members were aggregated, and the final scores were adjusted accordingly. This ensured that the recommendations satisfied the entire group's preferences. Finally, the POIs were ranked based on the adjusted scores, and the top recommendations were

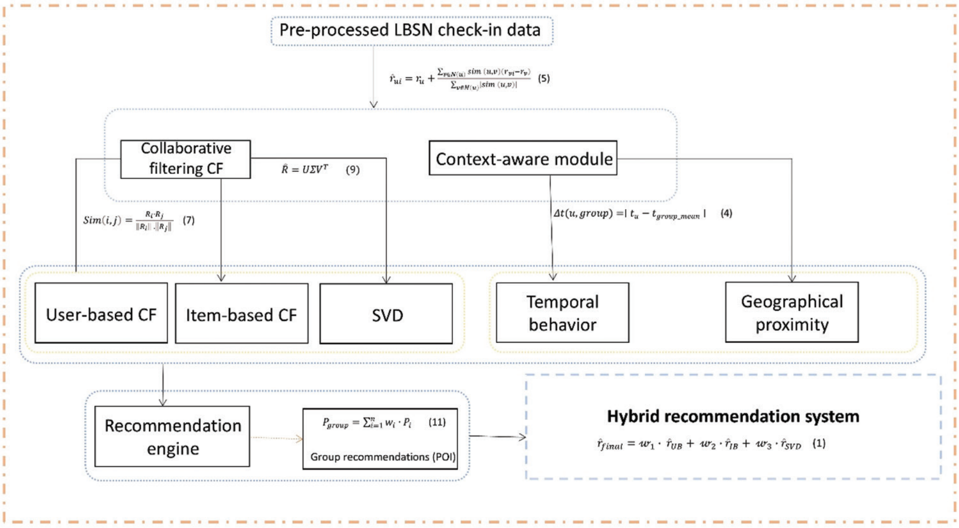


Fig. 2. Dataflow diagram of the proposed method to achieve the hybrid recommendation system
Abbreviations: CF: Collaborative filtering; LBSN: Location-based social networks; POI: Point of interest; SVD: Singular value decomposition

computed. Fig. 2 illustrates the process of generating personalized and group recommendations in LBSN. The process integrated user check-in data, CF (user-based and item-based), context-aware module SVD, and context-aware modeling (spatial proximity and temporal behavior) to produce highly accurate and tailored recommendations for both individuals and groups.

3.2. Overview of the Dataset

To show the effectiveness of the proposed method, we utilized the Gowalla dataset (Cho et al., 2011), which contains detailed check-in information of the users of the LBSN. Gowalla is a location-based service where users share their activities with friends by checking in at various POIs in the real world. The dataset includes check-in records from users across the globe, enabling a rich context for studying user mobility, spatial influences, and location recommendations, as shown in Table 1.

The dataset used in this study included 196,591 users, each identified with a unique user ID. Users in the dataset interacted with different locations by checking in at various POIs. User check-ins were timestamped, which allowed for the analysis of temporal behavior in conjunction with spatial movement patterns. There were 6,442,892 check-ins at 1,280,969 different POIs in the dataset. Each POI was associated with geographical coordinates (latitude and longitude) and represented a real-world location such as a restaurant, park, or other venue. The dataset consisted of over 6.4 million check-ins. This social network structure enables the modeling of user behavior based on social influence and peer interactions, which is valuable for group and social-aware recommendation systems.

Table 1. Overview of the dataset

Attribute	Description
Number of users	A total of 196,591 users with unique user IDs. Each user has a record of check-ins at various POIs.
Number of POIs	A total of 1,280,969 POIs with geographical coordinates (latitude and longitude).
Total check-ins	A total of 6,442,892 check-ins where users interacted with different POIs.
Geographical data	Latitude and longitude for each POI, which allow calculation of spatial proximity for the recommendation model.
Temporal information	Timestamps for each check-in that allow analysis of user behavior over time.
Categories of POIs	Food, entertainment, professional, shopping, etc.

Abbreviation: POI: Point of interest.

3.3. Data Preprocessing

Before conducting the analysis, several preprocessing steps were undertaken to ensure the quality and consistency of the data. First, the spatial data, including the latitude and longitude information for POIs, was normalized to standardize distance calculations and enable accurate spatial proximity analysis. Temporal information, specifically the timestamps for each check-in, was converted into standard date-time formats to facilitate the study of temporal patterns, such as user check-ins at different times of the day or week. To maintain data integrity and ensure the focus remained on active users and their mobility patterns, users who exhibited very low levels of activity were excluded from the dataset. The dataset used for this study included 196,591 users and 1,280,969 unique POIs, with a total of 6,442,892 check-ins. Although the dataset did not explicitly categorize POIs, they were grouped into various categories based on their nature and purpose. These categories included food (e.g., restaurants, cafés, and fast-food outlets), travel (e.g., parks and tourist spots), entertainment (e.g., concert halls, theatres, and cinema), professional (e.g., banks and offices), shopping and services (e.g., malls and retail stores), educational (e.g., schools, universities), hotels, residential (e.g., apartments), and sports. These groupings provided a clear understanding of user preferences and behaviors in different locations. Each attribute in the LBSN dataset indicated the interaction frequency between a specific user and a POI. After preprocessing, the user-POI interaction matrix, R, was constructed with Eq. (2), where each entry $R_{u,i}$ represents an interaction between a user (u) and a POI (i) with binary values indicating whether an interaction exists (1) or not (0), while m and n represent the number of users and the number of items (POIs), respectively.

$$R = \begin{bmatrix} R_{1,1} & R_{1,1} \cdots & R_{1,n} \\ R_{2,1} & R_{2,2} \cdots & R_{2,n} \\ \vdots & \vdots & \vdots \\ R_{m,1} & R_{m,2} & R_{m,n} \end{bmatrix} \quad (2)$$

The system was initialized by acquiring user check-in data from LBSNs. This data captured user interactions with various POIs, including restaurants, parks, and cafés. Each check-in record typically contained location coordinates, timestamp, and supplementary user demographic or behavioral data. These historical interactions formed the cornerstone of our recommendation process, revealing patterns in user preferences, frequently visited locations, and peak activity periods. Through systematic analysis of this data, the system uncovered underlying user preferences and their affinities for specific POIs.

3.4. User Influence Modeling

The next stage was user influence modeling, where our recommendation system improved user influence modeling by incorporating contextual factors that enhanced recommendation relevance. While traditional systems primarily analyzed user-item interactions, LBSNs required consideration of external influences. The framework evaluated spatial proximity, acknowledging that users typically preferred venues near their current location. It also accounted for temporal patterns, recognizing that preferences shifted throughout the day, such as favoring coffee shops in morning hours and restaurants in the evening. By integrating these contextual factors, we personalized recommendations based on two key influences: spatial proximity and temporal behavior patterns.

In LBSNs, physical proximity significantly influenced user preferences, with users typically favoring recommendations for nearby POIs. We evaluated the distance between users and POIs as a key factor in our recommendation system, recognizing that users closer to specific locations were more likely to interact with them and provided relevant recommendations. To quantify spatial relationships, we calculated each user's distance from the group's centroid location using latitude and longitude coordinates. This spatial proximity measurement employed the Haversine formula (Wirastuti et al., 2023), as expressed in Eq. (3).

$$d(u,group) = Haversine(lat_u|, lon_u|, lat_{mean}|, lon_{mean}|) \quad (3)$$

where: lat_u and lon_u represent the latitude and longitude of the user u, while lat_{mean} and lon_{mean} centroid the coordinates of the group.

Temporal patterns significantly influenced human behavior, with distinct location preferences emerging during work hours versus leisure time. Our recommendation system analyzed these temporal patterns by tracking when users typically interacted with various POIs. The system assigned higher priority to locations where a user's preferred check-in times aligned with peak activity periods. To quantify this temporal relationship, we evaluated the time differential between individual check-in patterns and group user behavior using Eq. (4).

$$\Delta t(u,group) = |t_u| - tgroup\_mean_{|} \quad (4)$$

We calculated user influence scores by analyzing interactions across location, time, and behavioral patterns. This scoring enabled us to prioritize recommendations that aligned with each user's preferences and behavioral patterns. Our recommendation accuracy relied on three key dimensions: geographic proximity, temporal activity patterns, and individual behavioral characteristics.

3.5. CF

CF techniques integrated user-item interaction data with sophisticated user influence factors. The basic idea of user-based CF centered on identifying users with statistically significant similarities in their spatial interaction patterns, particularly those demonstrating consistent check-in behaviors across comparable venues. The user-based CF recommended POIs that were frequently visited by users with similar traits, but remained unexplored by the target user. On the other hand, the item-based approach suggested POIs that exhibited substantial similarity to locations previously visited by the user, effectively extending the user's existing interaction profile. For example, when a user consistently visited specific types of restaurants, the model identified and recommended similar venues across diverse spatial contexts. This approach enabled the recommendation system to determine behavioral patterns, generating contextually refined and relevant recommendations as formalized in Eq. (5).

$$\hat{r}_{ui} = r_u + \frac{\sum_{v \in N(u)} sim(u,v)(r_{vi} - r_v)}{\sum_{v \in N(u)} |sim(u,v)|} \tag{5}$$

where: $\hat{r}_{ui}$ is the predicted rating for the user u on item i, $sim\ (u,v)$ is the similarity between users u and v, r_u and r_v are their average ratings, and $N(u)$ is the set of similar users.

The user-based CF approach leveraged behavioral similarities, predicting future preferences based on the users' historical interaction patterns. The identified user partners were characterized by significant behavioral similarities, particularly in spatio-temporal POI engagement. By analyzing the patterns of user check-in behaviors, the recommendation system identified the relational pattern that captured the interactions between users and venues. It involved identifying users with highly correlated interactions and subsequently utilizing their POI preferences to generate targeted recommendations for the user. This approach effectively transformed the collective user experience into a predictive recommendation mechanism. The user-based CF calculated user similarity by analyzing their interaction patterns with items (POIs), as shown in Eq. (6).

$$\hat{R}_{u.i} = \sum_{v \in similar\_users(u)} Sim(u,v) \cdot R_{v,i} \tag{6}$$

where: $\hat{R}_{u.i}$ represents the predicted score of interactions between user u and item i.

The item-based CF approach varied from traditional user-based methods by focusing on inter-item (venues) similarity relationships. This recommendation strategy, which operated mainly on spatial proximity and contextual similarity between venues, could effectively predict user preferences. By analyzing the characteristics and interaction patterns associated with specific locations, item-based CF could generate recommendations based on the inherent similarities between POIs. Consider a scenario where a user showed a distinct preference for coffee shops. The item-based CF systematically identified and recommended alternative venues that exhibited significant similarities. The similarity of two venues i and j, as well as the correlation between users who had interacted with both items, were measured using cosine, as expressed in Eq. (7). The predicted score for venue i is determined as in Eq. (8).

$$Sim(i,j) = \frac{R_i \mid \cdot R_j}{\| R_i \| . \| R_j \|} \tag{7}$$

$$\hat{R}_{u,i} = \sum_{j \in similar\_items(i)} Sim(i,j) \cdot R_{u,j} \tag{8}$$

The SVD served as a dimensionality reduction technique that effectively uncovered latent patterns and intricate relationships between users and items, particularly in large-scale datasets characterized by sparse user-item interaction matrices. This approach enabled the identification of underlying semantic factors that analyzed the check-in data, revealing features that significantly influenced user preferences. In the context of LBSNs, these latent factors involve attributes such as venue type (e.g., restaurant or park) and contextual environment (e.g., casual or formal). By extracting these hidden patterns, the recommendation system could generate highly contextualized suggestions even in scenarios with limited direct user-item interaction data. The SVD demonstrated outstanding efficacy in enhancing recommendation accuracy by discerning sensitive, non-obvious patterns that remained hidden through conventional analysis. This was achieved by decomposing the user-item interaction matrix R, thereby revealing latent structures. The reconstruction of the user-item matrix, as shown in Eq. (9), facilitated precise predictive recommendations by synthesizing these extracted characteristics.

$$\hat{R} = U \Sigma V^T \tag{9}$$

The top-N recommendations were based on the predicted score with the highest values in $\hat{R}$ calculated as in Eq. (10).

$$\hat{R}^{Hybrid} = \lambda_1 . \hat{R}^{UB} . \mid \lambda_2 . \hat{R}^{IB} . \mid \lambda_3 \hat{R}^{SVD} \tag{10}$$

where $\hat{R}^{UB}$ is the user-based CF's predicted scores, $\hat{R}^{IB}$ represents the item-based CF's predicted scores, and $\hat{\ }^{SVD}$ from SVD. The combined score is a weighted sum for the hybrid approach, where λ_1, λ_2, and λ_3 represents the weights of each method. This score was then combined with the contextual feature to get the final recommendation score.

3.6. Group Recommendation

In a hybrid recommendation engine, the system employs group recommendation methodology to integrate individual preferences into a cohesive collective recommendation. This process combined individual user preferences to identify a set of POIs that suitably satisfied the collective group's preferences. The group recommendation mechanism leveraged well-known techniques from social choice theory and the collaborative voting method to join and integrate different user preferences. By employing this approach, the model could strategically recommend venues that demonstrated the highest probability of collective appeal based on a comprehensive analysis of individual user profiles, historical patterns, and collective preferences. For each user in the group, the system computed preferences using CF, SVD, and context-aware modeling (spatial and temporal). Then, individual preferences were gathered into a group recommendation using voting techniques. A method for gathering preferences could be represented in Eq. (11).

$$P_{group} = \sum_{i=1}^{n} w_i \ P_i \tag{11}$$

where: P_{group} is the final group recommendation score, P_i is the recommendation score for individual user i based on their preferences, w_i is the weight assigned to each user's preferences, and n is the number of users in the group. The system then recommended items (POIs) that maximize P_{group}, considering each member's preferences and grouping them into a unified group decision.

3.7. User Behavior Analysis

In developing a comprehensive recommendation system, understanding user interaction patterns played a critical role. The system implemented three analytical components: clustering based on user preferences, temporal behavior analysis, and category-wise interaction analysis, each contributing to improved recommendation accuracy and personalization.

3.7.1. User Preferences

We employ K-means clustering due to its scalability and efficiency for partitioning users based on preference features. The number of clusters (k = 3) was selected empirically to balance between underfitting and over-segmentation. This step influenced clustering techniques, that is, K-means, to segment users based on their preferences and interaction behaviors. By grouping users with similar location preferences, the recommendation system can generate tailored suggestions that are more aligned with each user's interests. The clustering process often incorporates distance measures such as Euclidean or Cosine similarity to assess the proximity between users in terms of their preferences.

3.7.2. Temporal Behavior Analysis

This component analyzed the influence of time on user behavior, examining when users were most likely to interact with certain categories of locations (e.g., restaurants during lunch hours or entertainment venues in the evening). Temporal patterns helped the system to adjust recommendations based on the time of day or week. In LBSN research, the day was often divided into four distinct time segments: morning (5:00 A.M. – 11:59 A.M.), afternoon (12:00 P.M. – 5:59 P.M.), evening (6:00 P.M. – 8:59 P.M.), and night (9:00 P.M. – 4:59 A.M.), which reflected natural human activity patterns and aligned with previous studies on temporal behavior in location-based services (Choe et al., 2023). The analysis often involved temporal data partitioning, which was then visualized and measured to capture significant behavior patterns, typically using normalized interaction rates over defined periods.

This analysis focused on the categories of locations that users interacted with most frequently, providing insights into which types of venues held the most appeal for different user clusters. By computing metrics such as mean, median, and standard deviation of interactions within each category, the system could discern variations in user engagement across different categories, allowing more significant recommendations. To model these behaviors, equations were integrated to represent similar measures and interaction probabilities. For instance, the relationship between users i and j could be calculated as in Eq. (12).

$$Similarity_{ij|} = \frac{\sum\left(u_i \bullet u_j\right)}{\sqrt{\sum u_i^2} \times \sqrt{\sum u_j^2}} \tag{12}$$

where u_i and u_j are the interaction vectors for users i and j, respectively.

In addition, temporal interaction rates could be computed to normalize engagement across different time intervals. Each of these components contributed to a robust recommendation framework by capturing diverse dimensions of user behavior, thereby enhancing the relevance and personalization of the recommendations.

3.8. Evaluation

The evaluation stage of the research evaluated the system's efficiency and effectiveness through comprehensive performance metrics. These overlapping performance indicators, including precision, recall, F1 score, and accuracy, provided a multidimensional assessment of the recommendation system's ability to align recommended POIs with user preferences. The precision quantified the performance of recommended POIs calculated as the ratio of true positives (TP) to the total number of recommendations (TP + false positives [FP]), addressing the critical research question: "Of all system recommendations, what percentage accurately matches user preferences?" A high precision score indicated the system's capability to generate contextually appropriate suggestions. Recall evaluated the model by determining the ratio of TP to the total number of relevant items (TP + FP), answering: "Of all potentially relevant items, what proportion did the system successfully identify?" The F1 score, which was estimated as the mean of recall and precision, provided the balanced composite metric. This measurement was particularly valuable for assessing system performance in scenarios with irregular precision and recall, in terms of recommendation effectiveness. Accuracy represented the overall system performance by calculating the proportion of correctly identified relevant and irrelevant items. Mathematically derived as the ratio of all correct predictions (TP + TN) to the total number of predictions (TP + TN + FP + FN). This metric offered a comprehensive view of the recommendation system's capabilities.

4. Results and Discussion

This section presents a comprehensive empirical analysis of the hybrid recommendation approach for LBSNs, utilizing the Gowalla dataset. The investigation provides a thorough evaluation of the proposed methodology, systematically examining multiple critical dimensions of recommendation generation. The research explored the interactions between CF techniques, spatial proximity, and temporal behaviors. The study employed key performance metrics, including precision, recall, F1 score, and accuracy, to evaluate the performance of the model and recommendation efficiency.

4.1. User Influence Modeling

To highlight the influence of spatial and temporal features on the recommendation quality, we applied context-aware approaches, as shown in Fig. 3. This analysis optimized the matrices, streamlining their manipulation in subsequent computational phases.

The integration of spatial proximity demonstrated significant improvements in the performance of the recommendation system. Empirical findings uncovered

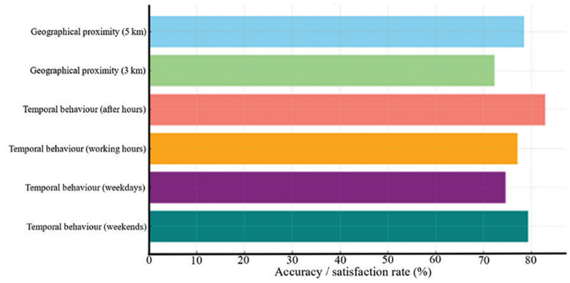


Fig. 3. User influence modeling. This chart shows how spatial and temporal patterns (e.g., proximity and working hours) impact user satisfaction with POI recommendations

that 78.5% of the generated suggestions corresponded to POIs within a 5-km radius of the user's current location, aligning closely with established behavioral patterns in LBSNs. The analysis substantiated that users commonly interact with proximate POIs, with a notable 72.3% interaction rate for locations within a 3-km radius. These results also emphasize the critical significance of spatial proximity in determining recommendation relevance. The temporal features also showed significant implications for recommendation accuracy. The model demonstrated remarkable capability in predicting user preferences across distinct time segments by analyzing check-in patterns. During non-working hours, the recommendation mechanism successfully suggested entertainment and food venues with an accuracy of approximately 82.9%. Similarly, during working hours, it strategically demonstrated work-related location recommendations, achieving a precision of 77.1%. Consistent behavioral patterns were observed across weekday and weekend contexts, confirming the key role of temporal features in enhancing the relevance of recommendations and contextual alignments with users' daily activities.

4.2. Temporal Distribution

The analysis of user activities within different venue categories during various periods of the day can be observed with the help of the temporal distribution of the check-ins. Fig. 4 demonstrates user interactions spread out over time in various categories, including educational, food, shopping and services, hotel, entertainment, travel, sports, professional, and residential venues.

Each cell represents the percentage of interactions within a specific category during the corresponding period, with darker shades indicating higher levels of interaction. Educational venues showed peak activity of 51.61% in the morning, aligning with typical school and university hours, while entertainment venues dominated in the evening and night, with

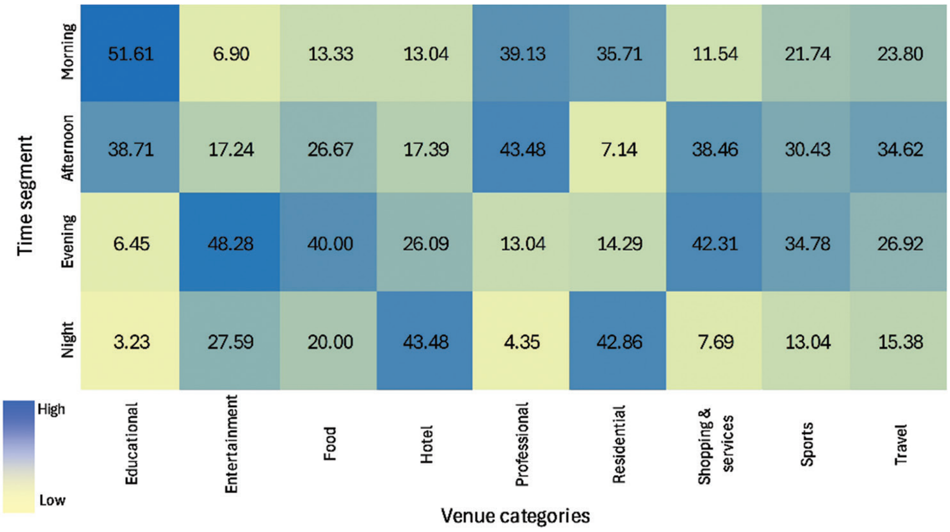


Fig. 4. Temporal distribution of user interaction across categories. Each cell represents the proportion of interactions within a specific venue category during different periods (morning, afternoon, evening, and night), highlighting peak usage patterns

48.28% and 27.59% activity, respectively, reflecting social and leisure behavior. Food establishments were most active during the evening, with 40% activity, and maintained significant activity of 26.67% and 20% in the afternoon and night, respectively, which corresponded to mealtime trends. Residential areas exhibit peak interactions of 42.86% and 35.71% at night and in the morning, respectively, as users returned home or began their day. Professional locations, as well as shopping and services, showed the highest engagement during traditional working and leisure hours, with professional venues showing high activity of 43.48% and 39.13% in the afternoon and morning, respectively. On the other hand, shopping and services venues reported high activity of 42.31% and 38.46% in the evening and afternoon, respectively. Hotels and travel-related venues exhibit steady activity, with hotels demonstrating the highest activity of 43.48% at night, while travel reporting consistent activity across all periods. These distributions underscore the importance of time-aware modeling in understanding user behaviors and optimizing LBSN-based recommendation systems.

4.3. Category-Wise Interaction Analysis

We also performed interaction analysis based on different venue categories, which is a vital part of understanding and utilizing user activities and preferences. Fig. 5 presents a comprehensive distribution of user interactions across different venue types, including food, shopping and services, travel, hotel, educational, entertainment, professional, sports, and residential domains. It represented the user

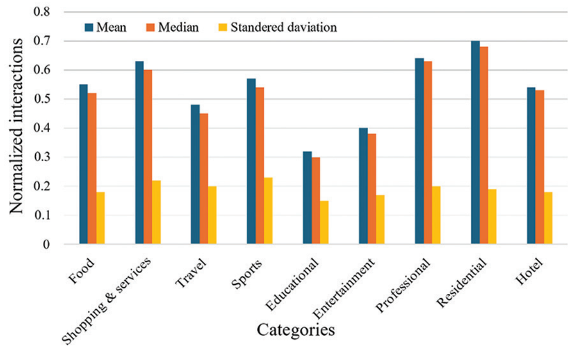


Fig. 5. Category-wise interaction analysis. The analysis displays normalized interaction values (mean, median, and standard deviation) across venue types, reflecting user engagement preferences

interaction patterns across these diverse categorical contexts.

The professional and residential venue categories demonstrate the highest average and median interaction values, indicating the users' tendency to check in at these venues. Shopping and services, as well as food categories, exhibit substantial interaction levels, suggesting that these places are prominent venues for user engagement. This analysis reveals a remarkably consistent interaction pattern across most categories, as evidenced by the minimal standard deviation. However, the sports, as well as shopping and services categories, exhibit marginally higher variability, indicating potentially more heterogeneous user behaviors within these categories. Conversely, the educational and entertainment venue categories show noticeably lower mean and median interaction values, suggesting comparatively reduced user engagement in these specific venue types.

4.4. Temporal Behavior Distribution

Temporal analysis is an important factor in human behavior studies, which can be achieved by analyzing the users' activities during different periods of the day. We divided the time into four segments, including morning, afternoon, evening, and night. Fig. 6 shows the distribution of users' interactions over time across various categories. For educational venues, most interactions (51.61%) occurred in the morning, reflecting typical school and university hours, with a decline to 38.71% in the afternoon and minimal activity during the evening and night. In contrast, entertainment venues show high activity of 48.28% and 27.59% in the evening and night, respectively, corresponding to leisure and nightlife activities. Food establishments reported the highest engagement in the evening of 40%, followed by 26.67% and 20% in the afternoon and night, respectively, aligning with dining patterns. Hotels demonstrate the highest activity of 43.48% at night, indicating late check-ins or overnight stays. Professional locations show the most activity of 39.13% and 43.48% in the morning and afternoon, respectively, which is consistent with standard working hours. Residential areas are highly engaged at night and morning, with activity of 42.86% and 35.71%, respectively, reflecting the daily routines of starting and ending the day at home. Shopping and services venues are most active during the afternoon and evening, with interactions of 38.46% and 42.31%, respectively, reflecting shopping and errand behaviors. Sports activities are distributed across the day, with the highest activities of 34.78% and 30.43% in the evening and afternoon, respectively. Travel-related venues show moderate activity throughout the day, with peaks of 34.62% the afternoon and 23.80% in the morning. These results emphasize the temporal nature of user behavior in LBSNs, providing insights into category-specific trends that can be leveraged for time-aware recommendation systems.

4.5. Clustering Based on Preferences

The users' clustering approach enables the systematic identification of behavioral similarity within the user populations. By leveraging efficient clustering techniques such as *K*-means, researchers can effectively stratify users based on their mutual interaction preferences and spatial patterns. This approach reveals distinct user trends, ranging from individuals mostly interested in food venues to those demonstrating explicit engagement with entertainment or professional venues. The clustering methodology provides the mechanism for generating personalized recommendation strategies that align precisely with the characteristics of each identified user cluster.

Fig. 7 presents a comprehensive scatter plot illustrating the empirical outcomes of *K*-means clustering applied to analyze users' motivational factors for interaction across diverse venue categories, including food, travel, and other domains. Each data point represents an individual user, with the visualization classifying the user population into three distinct clusters, represented by different colors. The coordinate axes represent two critical dimensional components: food interaction, as well as shopping and services interaction, which serve as primary determinants in the cluster formation process in this figure. This visualization provides a graphical representation of user behavioral clusters, facilitating a deeper understanding of interaction patterns across various venue categories.

The clustering method presents user preferences through comprehensive behavioral and categorical interactions. User clusters are calculated based on distinctive venue types, such as clusters characterized by pronounced engagement with food, as well as

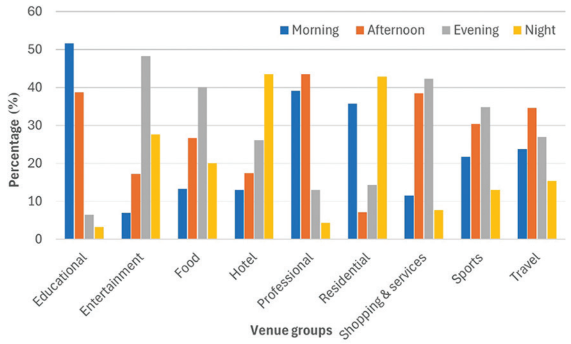


Fig. 6. Temporal behavior distribution across categories. This figure displays percentage interactions across categories segmented by time, revealing trends in check-in behavior

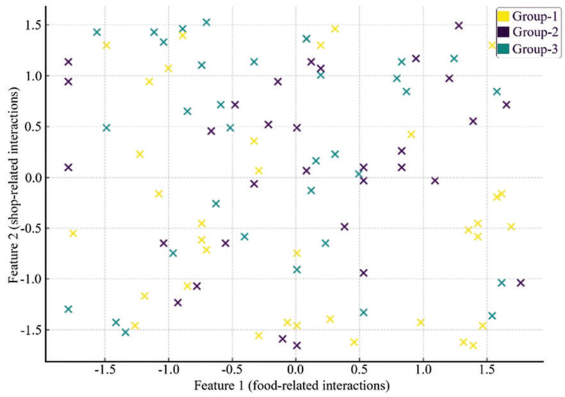


Fig. 7. Clustering based on user preference

DOI: 10.6977/IJoSI.202508_9(4).0008
N. U. Khan, R. Riaz, etc./Int. J. Systematic Innovation, 9(4), 106-122 (2025)

shopping and services categories. Similarly, we can compute other clusters demonstrating significant interaction patterns in travel and hotel venues. This clustering approach enables the recommendation system to generate highly contextualized recommendations tailored to the distinct preference profiles of each identified user group. The cluster-based recommendation strategy facilitates a more accurate and relevant user experience by aligning recommendation content with the observed behavioral and categorical preferences of each user segment.

4.6. Hybrid Recommendation System

The proposed model is implemented to determine its performance based on check-in data. In this section, we presented the results of various methods in comparison with our proposed model and assessed their effect and efficiency. In Fig. 8, user-based filtering identifies users with check-in patterns, achieving 68.5% precision and 70.2% recall. While effective, this method encountered limitations with users having sparse interaction data. It was most successful when analyzing users with highly similar preferences, with 65.9% of recommendations being relevant based on users with comparable experiences. Item-based CF slightly outperformed user-based, reaching a precision of 72.4% and a recall of up to 74.8%. It excelled in scenarios with consistent user behavior patterns, such as frequent visits to similar venue types. Recommending POIs based on item similarity particularly benefited users with clear, repetitive preferences. Applying SVD revealed interesting hidden relationships between users and POIs, and it is the most effective in sparse data environments, achieving 75.5% precision and 71.7% recall. The high efficiency demonstrated SVD's capability to uncover latent interaction patterns, generating improved recommendations for users.

The proposed hybrid model, integrating user-based CF, item-based CF, and SVD techniques with context-aware spatial and temporal features, demonstrated superior performance compared to individual recommendation methods. Achieving precision up to 80.6% and a recall up to 77.3%, with an F1 score achieving up to 78.5%, the hybrid approach consistently delivered the most accurate recommendations, as shown in Fig. 8.

Table 2 provides a detailed assessment of the recommendation system's accuracy across top-5, top-10, and top-20 recommendations, demonstrating the model's effectiveness in suggesting the most relevant items for users.

The proposed model demonstrates remarkable accuracy across different recommendation depths. For the top-5 recommendations, the system achieves 82.3% accuracy, indicating that over four-fifths of

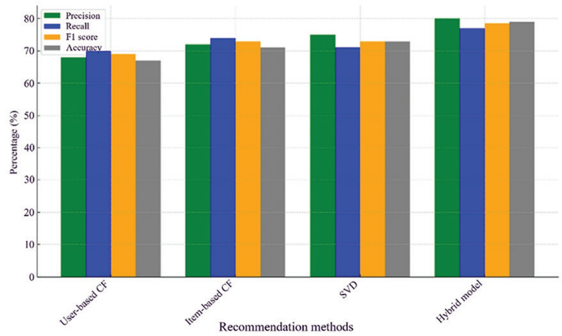


Fig. 8. Performance comparison of recommendation methods
Abbreviations: CF: Collaborative filtering; SVD: Singular value decomposition

Table 2. Top-N recommendation performance

Top-N	Accuracy (%)
Top-5	82.3
Top-10	90.6
Top-20	95.2

the initial suggestions are highly relevant to the user. The accuracy progressively improves, with top-10 recommendations reaching 90.6% and top-20 recommendations achieving an impressive 95.2% precision. This improvement reveals that as the number of recommended options increases, the likelihood of suggesting relevant POIs becomes significantly higher. The expanding recommendation set provides users with greater flexibility and choice, enhancing the overall recommendation experience. By combining different methodological strategies and incorporating contextual factors such as venue and temporal data, the system generates tailored and precise suggestions. Considering features such as users' venue and temporal aspect that enhance the appropriateness of the suggestions, resulting in improved user satisfaction. The results presented in Table 3 present the efficiency of our proposed hybrid model in comparison with baseline methods.

The baseline models demonstrate moderate performance, with accuracy ranging from 65% to 73.4% and respectable precision, recall, and F1 score metrics. The hybrid model significantly outperforms these baseline approaches, with precision up to 80.6%, recall achieving 77.3%, and accuracy of 79.1%. The performance improvement stems from the strategic integration of CF as a hybrid model with context-aware features and techniques, effectively mitigating individual method limitations. By leveraging user-item similarities and uncovering hidden relationships through SVD, the hybrid model generates more accurate recommendations. The approach proves particularly powerful in data-sparse environments,

Table 3. Comparison of the hybrid model with baseline methods

Method	Reference	Precision	Recall	F1 score	Accuracy
LBRS	Tao et al., 2021	68.60	60.80	64.80	67.50
GR-DELM with Gowalla	Zhao et al., 2018	77.84	81.23	79.50	-
TrustWalker	Ravi & Vairavasundaram, 2016	73.98	-	73.41	-
SVD	Rodpysh et al., 2023	75.50	71.70	73.60	73.40
CF	Liu et al., 2014	65.00	60.00	62.00	65.00
PARS	Ye et al., 2011	-	69.00	70.50	72.00
Hybrid model	(The current study)	80.60	77.30	78.50	79.10

Abbreviation: CF: Collaborative filtering; GR-DELM: Geographical-aware recurrent deep extreme learning machine; LBRS: Location-based recommendation system; PARS: Personalized adaptive recommendation system; SVD: Singular value decomposition.

enabling more accurate and relevant suggestions tailored to LBSN contexts. Ultimately, the proposed hybrid methodology emerges as the most reliable and sophisticated recommendation solution for location-based services.

5. Conclusion

In this research, we proposed a novel hybrid approach for group recommendations in LBSNs, addressing the unique contextual needs of these platforms. By incorporating spatial proximity and time-based patterns, our model effectively combines user-based CF, item-based CF, and SVD to enhance both accuracy and personalization. The integration of spatial and temporal factors significantly improves precision, as users frequently engage with nearby locations that align with their daily routines. Our evaluation demonstrated that this hybrid approach outperforms conventional methods, particularly in situations where interaction data are sparse. This model was able to achieve high accuracy and diversity in recommendations. However, limitations persist, especially with cold-start users and scalability as LBSNs expand in size. The system's reliance on sufficient historical interaction data poses a challenge for new or infrequent users, despite the mitigating effect of the hybrid method. Moreover, the hybrid model's computational complexity can limit responsiveness in large-scale, real-time applications, as the combination of user- and item-based filtering with SVD and contextual information may slow down recommendations in extensive datasets. In addition, while our model successfully accounts for user-item interactions, geographic proximity, and temporal behavior, it currently lacks real-time contextual adaptability factors, such as sudden location shifts or external conditions that are not fully captured, which may limit relevance in highly dynamic environments.

Future evaluations will consider deployment in real-time environments and validation on diverse LBSN datasets to assess scalability and generalization. It could address these limitations by enhancing cold-start handling with advanced embedding methods or social network analysis, which would incorporate user metadata or social connections to generate initial recommendations for new users or items. Improvements for real-time recommendations could involve integrating dynamic contextual data, such as weather or event information, to adapt recommendations to users' immediate surroundings and conditions. Furthermore, leveraging deep network-based models and attention mechanisms could improve the model's understanding of complex relationships between users, items, and context, thereby boosting both scalability and accuracy. Overall, this hybrid approach demonstrates strong potential for effectively meeting the dynamic and personalized demands of LBSNs.

Acknowledgment

None.

References

Arabi, H. (2018). *Collaborative and Content-Based Filtering Personalized Recommender System for Book*. University of Malaya, Kuala Lumpur, Malaysia. Available from: https://studentsrepo.um.edu.my/8978 [Last accessed on 2025 July 29].

Barai, S., & Bhaumik P. (2015). *A Taxonomic Study of the Recent Security Concerns in Opportunistic Networks.* International Journal for Scientific Research and Development, Conference Special Issue (Computer Networking), p.34–46. Available from: https://ijsrd.com/articles/spcn028.pdf ijsrd.com [Last accessed on 2025 July 29].

Cho, E., Myers, S.A., & Leskovec J. (2011). Friendship and Mobility: User Movement in Location-Based

Social Networks. In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. p1082–1090. https://doi.org/10.1145/2020408.2020579

Choe, Y., Lee, C.K., Choi, J., Kim, M., & Sim, K.W. (2023). Identifying tourist spatial and temporal patterns using GPS and sequence alignment method. *Journal of Travel Research*, 62(6), 1181–1201. https://doi.org/10.1177/00472875221127685

Dadoun, A., Troncy, R., Ratier, O., & Petitti R. (2019). Location Embeddings for Next Trip Recommendation. In: *Companion Proceedings of the 2019 World Wide Web Conference*. p1–8. https://doi.org/10.1145/3308560.3316535

Davtalab, M., & Alesheikh, A.A. (2023). A multi-criteria point of interest recommendation using the dominance concept. *Journal of Ambient Intelligence and Humanized Computing*, 14, 6681–6696. https://doi.org/10.1007/s12652-021-03533-x

Deldjoo, Y., Schedl, M., Cremonesi, P., & Pasi, G. (2020). Recommender systems leveraging multimedia content. *ACM Computing Surveys*, 53(5), 1–38. https://doi.org/10.1145/3398685

Dietz, L.W., Sánchez, P., & Bellogín, A. (2025). Understanding the influence of data characteristics on the performance of point-of-interest recommendation algorithms. *Information Technology and Tourism*, 27, 75–124. https://doi.org/10.1007/s40558-024-00304-0

Dutta, P.K., Kumar, A., Sakici, Ş., & Mensah, B. (2025). Enhancing point-of-interest recommendation systems through multi-modal data integration in location-based social networks: Challenges and future directions. *EDRAAK*, 2025, 12–18. https://doi.org/10.70470/EDRAAK/2025/003

Eliyas, S., & Ranjana, P. (2022). Recommendation Systems: Content-Based Filtering vs Collaborative Filtering. In: *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. https://doi.org/10.1109/ICACITE54196.2022.9817512

Ezin, E. (2024). *Towards Context-Aware Recommender Systems for Tourists*. University of Bristol, UK. Available from: https://research-information.bris.ac.uk/en/studentTheses/towards-context-aware-recommender-systems-for-tourists [Last accessed on 2025 July 29].

Hu, Z.H., Li, Z., Deng, Y., & Ma, X. (2019). Examining collaborative filtering algorithms for clothing recommendation in E-commerce. *Textile Research Journal*, 89(14), 2821–2835. https://doi.org/10.1177/0040517519836282

Kanfade, M.M., Ambade, S.D., & Bhagat, A.P. (2018). Location-Based Notification System. In: *2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)*. https://doi.org/10.1109/RICE.2018.8509064

Liu, H., Hu, Z., Mian, A., Tian, H., & Zhu, X. (2014). A new user similarity model to improve the accuracy of collaborative filtering. *Knowledge-Based Systems*, 56, 156–166. https://doi.org/10.1016/j.knosys.2013.11.002

Mahajan, P., & Kaur, P.D. (2023). 3T-IEC*: A context-aware recommender system architecture for smart social networks. *Journal of Intelligent Information Systems*, 60(1), 199–233. https://doi.org/10.1007/s10844-022-00743-3

Nian, C.C. (2021). *Recommender System on Social Networking Site with Domain Specific and Sparse Data*. Toronto Metropolitan University. Available from: https://rshare.library.torontomu.ca/articles/thesis/14661834 [Last accessed on 2025 July 29].

Papadakis, H., Papagrigoriou, A., Panagiotakis, P., Kosmas, E., & Fragopoulou, P. (2022). Collaborative filtering recommender systems taxonomy. *Knowledge and Information Systems*, 64(1), 35–74. https://doi.org/10.1007/s10115-021-01628-7

Ravi, L., & Vairavasundaram, S. (2016). A collaborative location-based travel recommendation system through enhanced rating prediction for the group of users. *Computational Intelligence and Neuroscience*, 2016, 1291358. https://doi.org/10.1155/2016/1291358

Redondo, R.P.D., Garcia-Rubio, C., Vilas, A.F., Campo, C., & Rodriguez-Carrion, A. (2020). A hybrid analysis of LBSN data to early detect anomalies in crowd dynamics. *Future Generation Computer Systems*, 109, 83–94. https://doi.org/10.1016/j.future.2020.03.010

Rodpysh, K.V., Mirabedini, S.J., & Banirostam, T. (2023). Employing singular value decomposition and similarity criteria for alleviating cold start and sparse data in context-aware recommender systems. *Electronic Commerce Research*, 23(2), 681–707. https://doi.org/10.1007/s10660-022-09679-8

Sánchez, P., & Bellogín, A. (2022). Point-of-interest recommender systems based on location-based social networks: A survey from an experimental perspective. *ACM Computing Surveys*, 54(11s), 1–37. https://doi.org/10.1145/3520490

Sun, Y., Tan, W., Huang, L., Xie, N., Ruan, L., &

Xu, L.D. (2022). An integrated collaborative filtering framework with location-aware graph embedding in intelligent internet of things systems. *IEEE Transactions on Aerospace and Electronic Systems*, 58(6), 5089–5104. https://doi.org/10.1109/TAES.2022.3213631

Tao, X., Sharma, N., Delaney, P., & Hu, A. (2021). Semantic knowledge discovery for user profiling for location-based recommender systems. *Human Centric Intelligent Systems*, 1(1), 32–42. https://doi.org/10.2991/hcis.k.210704.001

Teoman, H.A. (2022). *Location Recommendation for Groups on Location-Based Social Networks*. Middle East Technical University. Available from: https://hdl.handle.net/11511/123456 [Last accessed on 2025 July 29].

Tourinho, I.A.D.S., & Rios, T.N. (2021). FACF: Fuzzy areas-based collaborative filtering for point-of-interest recommendation. *International Journal of Computational Science and Engineering*, 24(1), 27–41. https://doi.org/10.1504/IJCSE.2021.10035701

Wachyuni, S.S., & Kusumaningrum, D.A. (2020). The effect of COVID-19 pandemic: How are the future tourist behavior? *Journal of Education Society and Behavioural Science*, 33(4), 67–76. https://doi.org/10.9734/jesbs/2020/v33i430328

Wang, H., Terrovitis, M., & Mamoulis, N. (2013). Location Recommendation in Location-Based Social Networks Using User Check in Data. In: *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. p374–383. https://doi.org/10.1145/2525314.2525362

Wang, S., Zhang, X., Wang, Y., & Ricci, F. (2024). Trustworthy recommender systems. *ACM Transactions on Intelligent Systems and Technology*, 15(4),1–20. https://doi.org/10.1145/3627826

Wang, Y., Sun, H., Zhou, Y., Zhou, W., & Zhu, S. (2019). A heterogeneous graph embedding framework for location-based social network analysis in smart cities. *IEEE Transactions on Industrial Informatics*, 16(4), 2747–2755. https://doi.org/10.1109/TII.2019.2910370

Wirastuti, N., Verlin, L., Mkwawa, I.H., & Samarah, K.G. (2023). Implementation of geographic information system based on google maps api to map waste collection point using the haversine formula method. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, 9(3), 731–745. https://doi.org/10.26555/jiteki.v9i3.26792

Ye, M., Yin, P., Lee, W.C., & Lee, D.L. (2011). Exploiting Geographical Influence for Collaborative Point-of-Interest Recommendation. In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. p325–334. https://doi.org/10.1145/2009916.2010125

Yuan, Z., & Chen, C. (2017). Research on Group POIs Recommendation Fusion of Users' Gregariousness and Activity in LBSN. In: *2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. p102–107. https://doi.org/10.1109/ICCCBDA.2017.7951955

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52(1), 1–38. https://doi.org/10.1145/3285029

Zhao, X., Ma, Z., & Zhang, Z. (2018). A novel recommendation system in location-based social networks using distributed ELM. *Memetic Computing*, 10(3), 321–331. https://doi.org/10.1007/s12293-018-0273-3

Zhao, X., Zhang, Z., Bi, X., & Sun, Y. (2023). A new point-of-interest group recommendation method in location-based social networks. *Neural Computing and Applications*, 35, 12945–12956. https://doi.org/10.1007/s00521-020-04979-4

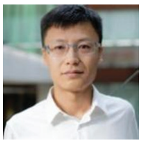
Zheng, Y. (2022). Context-aware collaborative filtering using context similarity: An empirical comparison. *Information*, 13(1), 42. https://doi.org/10.3390/info13010042

Zheng, Y., & Zhou, X. (2024). Modeling multi-factor user preferences based on transformer for next point of interest recommendation. *Expert Systems with Applications*, 255, 124894. https://doi.org/10.1016/j.eswa.2023.124894

Zhou, X., Wang, Z., Liu, X., Liu, Y., & Sun, G. (2024). An improved context-aware weighted matrix factorization algorithm for point of interest recommendation in LBSN. *Information Systems*, 122, 102366. https://doi.org/10.1016/j.is.2024.102366

AUTHOR BIOGRAPHIES

Dr. Naimat Ullah Khan Received his Bachelor's and Master's degrees in Computer Sciences from Pakistan. He worked as a lecturer at Preston University Pakistan. He completed his PhD in Communication and Information Systems from Shanghai University in 2024 and is currently pursuing his PhD in Computer Systems from University of Technology Sydney, Australia on collaborative research program. His current research interest includes internet of things, machine learning, artificial intelligence, and data sciences.



Dr. Xianzhi Wang is currently a Senior Lecturer in the School of Computer Science and a Core Member of the Australian Artificial Intelligence Institute (AAII) at the University of Technology Sydney. His research focuses on fundamental research and application issues with data mining, spatiotemporal analysis, and recommender systems. These underpin elevated operational efficiency and cost-effectiveness in transportation, e-commerce, and healthcare sectors. The long-term ambition is to devise robust techniques to solve computing challenges in complex scenarios, for example, citywide transit.

Rubina Riaz Received her Bachelor's degree in Computer Sciences from University of Azad Jammu and Kashmir, Pakistan and Master's degree in Computer Sciences from Preston University, Pakistan. Currently, she is pursuing her PhD from Dalian University of Technology, China. Her current research interest includes internet of things, sensor networks, machine learning, image processing, artificial intelligence, security, and big data.

Dr. Angela Huo is currently an Associate Professor at the School of Computer Science in the Faculty of Engineering and Information Technology at the University of Technology Sydney (UTS) working in the fields of Data Analysis, Cybersecurity, and Privacy. In particular, her research focuses on utilizing AI and data analysis techniques to improve the efficiency and security of modern software systems. Her recent interest lies in recommendation system, anomaly detection, and privacy-preserving through data mining and deep learning.

Dr. Jutao Hao is currently an assistant professor and master's supervisor at the School of Electronic Information Engineering, Shanghai DianJi University. His research focuses on industrial big data architecture and multimodal operational condition anomaly detection, which significantly enhances industrial equipment reliability by integrating multi-source data to achieve full-dimensional monitoring, early fault warning, and precise root cause localization.

INSTRUCTIONS TO AUTHORS

Submission of papers

The International Journal of Systematic Innovation is a refereed journal publishing original papers four times a year in all areas of SI. Papers for publication should be submitted online to the IJoSI website (http://www.ijosi.org) In order to preserve the anonymity of authorship, authors shall prepare two files (in MS Word format or PDF) for each submission. The first file is the electronic copy of the paper without author's (authors') name(s) and affiliation(s). The second file contains the author's (authors') name(s), affiliation(s), and email address(es) on a single page. Since the Journal is blind refereed, authors should not include any reference to themselves, their affiliations or their sponsorships in the body of the paper or on Figs and computer outputs. Credits and acknowledgement can be given in the final accepted version of the paper.

Editorial policy

Submission of a paper implies that it has neither been published previously nor submitted for publication elsewhere. After the paper has been accepted, the corresponding author will be responsible for page formatting, page proof and signing off for printing on behalf of other co-authors. The corresponding author will receive one hardcopy issue in which the paper is published free of charge.

Manuscript preparation

The following points should be observed when preparing a manuscript besides being consistent in style, spelling, and the use of abbreviations. Authors are encouraged to download manuscript template from the IJoSI website, http://www.ijosi.org.

1. *Language.* Paper should be written in English except in some special issues where Chinese may be accepTable Each paper should contain an abstract not exceeding 200 words. In addition, three to five keywords should be provided.
2. *Manuscripts.* Paper should be typed, single-column, double-spaced, on standard white paper margins: top = 25mm, bottom = 30mm, side = 20mm. (The format of the final paper prints will have the similar format except that double-column and single space will be used.)
3. *Title and Author.* The title should be concise, informative, and it should appear on top of the first page of the paper in capital letters. Author information should not appear on the title page; it should be provided on a separate information sheet that contains the title, the author’s (authors') name(s), affiliation(s), e-mail address(es).
4. *Headings.* Section headings as well as headings for subsections should start front the left-hand margin.
5. *Mathematical Expressions.* All mathematical expressions should be typed using Equation Editor of MS Word. Numbers in parenthesis shall be provided for equations or other mathematical expressions that are referred to in the paper and be aligned to the right margin of the page.
6. *Tables and Figs.* Once a paper is accepted, the corresponding author should promptly supply original copies of all drawings and/or tables. They must be clear for printing. All should come with proper numbering, titles, and descriptive captions. Fig (or table) numbering and its subsequent caption must be below the Fig (or table) itself and as typed as the text.
7. *References.* Display only those references cited in the text. References should be listed and sequenced alphabetically by the surname of the first author at the end of the paper. For example:

Altshuller, G. (1998). *40 Principles: TRIZ Keys to Technical Innovation*, Technical Innovation Center.
Sheu, D. & Lee, H. (2011). A Proposed Process for Systematic Innovation, International Journal of Production Research, Vol. 49, No. 3, 2011, 847-868.

The International Journal of Systematic Innovation
Journal Order Form

Organization Or Individual Name	
Postal address for delivery	
Person to contact	Name: e-mail: Position: School/Company:
Order Information	**I would like to order ___ copy(ies) of the** *International Journal of Systematic Innovation***:** **Period Start: 1st/ 2nd half ____, Year:____(Starting 2010)** **Period End : 1st/ 2nd half ____, Year:** **Price:** **Institutions: US $150 (yearly) / NT 4,500 (In Taiwan only)** **Individuals: US $50 (yearly) / NT 1500 (In Taiwan only)** (Local postage included. International postage extra) **E-mail to**: IJoSI@systematic-innovation.org or fax: +886-3-572-3210 Air mail desired □ (If checked, we will quote the additional cost for your consent)
Total amount due	**US$**

Payment Methods:

1. **Credit Card (Fill up the following information and e-mail/ facsimile this form to The Journal office indicated below)**
2. **Bank transfer**
3. **Account:** The Society of Systematic Innovation
4. **Bank Name:** Mega International Commercial BANK
5. **Account No:** 020-53-144-930
6. **SWIFT Code:** ICBCTWTP020
7. **Bank code**：017-0206
8. **Bank Address:** No. 1, Xin'an Rd., East Dist., Hsinchu City 300, Taiwan (R.O.C.)

VISA / Master/ JCB/ AMERICAN Cardholder Authorization for Journal Order

Card Holder Information

Card Holder Name	(as it appears on card)		
Full Name (Last, First Middle)			
Expiration Date	/ (month / year)	Card Type	□ VISA □ MASTER □ JCB
Card Number	□□□□-□□□□-□□□□-□□□□	Security Code	□□□ 
Amount Authorized		Special Messages	
Full Address (Incl. Street, City, State, Country and Postal code)			

Please Sign your name here _________________ (same as the signature on your card)

The Society of Systematic Innovation

6 F, #352, Sec. 2, Guanfu Rd,
Hsinchu, Taiwan, 30071, R.O.C.