Vol. 9 No. 6
December 2025

Opportunity Identification & Problem Solving

INTERNATIONAL JOURNAL OF SYSTEMATIC INNOVATION

ISSN (Print): 2077-7973
ISSN (Online): 2077-8767
DOI: 10.6977/IJoSI.202512_9(6)

ACCSCIENCE PUBLISHING

ISSN (Print): 2077-7973
ISSN (Online): 2077-8767
DOI: 10.6977/IJoSI.202512_9(6)

The International Journal of Systematic Innovation

Publisher:

The Society of Systematic Innovation

Editorial Office:

- *International Journal of Systematic Innovation* Editorial Office
- 9 Raffles Place, Republic Plaza 1 #06-00 Singapore 048619
- Email: ijosi.office@accscience.sg
- Website: https://ijosi.org/journal/IJOSI
- Tel: +65 8182 1586

ISSN (Print): 2077-7973
ISSN (Online): 2077-8767
DOI: 10.6977/IJoSI.202512_9(6)

INTERNATIONAL JOURNAL OF SYSTEMATIC INNOVATION

CONTENTS

DECEMBER 2025 VOLUME 9 ISSUE 6

FULL PAPERS

DOI: 10.6977/IJoSI.202512_9(6).0001
K. Altun/Int. J. Systematic Innovation, 9(6), 1-16 (2025)

Contradiction-oriented exploration: A dual-track methodology combining OTSM–TRIZ and the Six-Box Scheme

Koray Altun[1,2]*

[1]Department of Industrial Engineering, Faculty of Engineering and Natural Sciences, Bursa Technical University, Bursa, Turkey

[2]ARDiMER Digital Innovation Ltd, Bursa, Turkey

*Corresponding author E-mail: koray.altun@btu.edu.tr

(Received 08 September 2025; Final version received 24 November 2025; Accepted 2 December 2025)

Abstract

Complex problems do not just ask for better answers; they ask for better ways of thinking. Accordingly, complex socio-technical design problems require integrated approaches that simultaneously address technical contradictions and human-centered processes. This study introduces contradiction-oriented exploration (COREX), a dual-track methodology designed to solve complex design problems involving both technical systems and human behavior. This approach combines two powerful tools: (i) The General Theory of Powerful Thinking–Theory of Inventive Problem Solving, which focuses on identifying and resolving system-level contradictions; and (ii) The Six-Box Scheme, which provides a user-centered, process-based framework for creative problem solving. By linking contradiction analysis with recursive exploration and real-world testing, this approach helps teams move from unclear user needs to structured innovations. The method was applied in a research and development setting focused on adaptive seat design. Participants followed a procedure that included problem modeling, contradiction identification, and inventive solution development. Results showed that COREX helped teams address design trade-offs more effectively than when using either method alone. The feedback cycles allowed for continuous improvement and system refinement. Overall, the methodology offers practical value for design teams working in emerging socio-technical domains by supporting both analytical thinking and creative ideation in an integrated process.

Keywords: Contradiction-Oriented Exploration, General Theory of Powerful Thinking–theory of Inventive Problem Solving, Six-Box Scheme

1. Introduction

General Theory of Powerful Thinking–Theory of Inventive Problem Solving (OTSM-TRIZ) excels in logical depth but is difficult to apply. In contrast, the Six-Box Scheme offers an intuitive, step-by-step process but lacks tools for handling contradictions. This reveals a core tension: Logic helps us go deep, while process helps us move forward, but each without the other remains incomplete.

The challenge mirrors a typical TRIZ contradiction (Fig. 1): increasing "ease of use" without reducing "complexity handling." Suggested inventive principles include IP1 (segmentation), IP5 (merging), IP13 (the other way round), and IP24 (intermediary).

To resolve this, we propose contradiction-oriented exploration (COREX), a unified innovation methodology that combines OTSM-TRIZ's contradiction reasoning with the Six-Box Scheme's structured flow. COREX merges logic and process, making problem-solving both rigorous and intuitive. Its core contribution is bridging two complementary paradigms—analytical and procedural—into a scalable method for addressing complex technical and behavioral challenges. COREX enables deeper insight, focused ideation, and iterative learning. It is not only a tool but also a systematic way of thinking within complex systems.

This article proceeds as follows: Section 2 reviews contradiction-based and cognitive models.

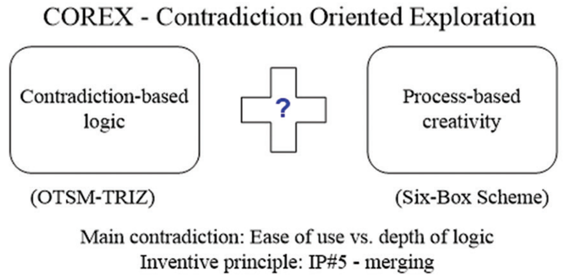


Fig. 1. Building blocks of the contradiction-oriented exploration
Abbreviations: OTSM-TRIZ: General Theory of Powerful Thinking–Theory of Inventive Problem Solving**.**

Section 3 introduces theoretical foundations. Section 4 presents COREX. Section 5 offers a case study. Section 6 compares results. Section 7 concludes with contributions and future directions.

2. Literature Review

Various systematic innovation processes (SIPs) have been proposed in the literature. Sheu and Lee (2011) introduced a phased SIP that integrates TRIZ and non-TRIZ tools, allowing structured transitions from opportunity discovery to implementation. Mann (2007) proposed the systematic creativity process, emphasizing the use of TRIZ tools across phases such a tool selection, idea generation, and evaluation. However, these models often address tactical rather than strategic innovation, and they typically lack recursive feedback mechanisms.

The W-model (Brandenburg, 2002) and the innovation value chain (Hansen and Birkinshaw, 2007) introduce strategic perspectives but fall short in problem-solving depth and contradiction resolution. Likewise, the accelerated innovation process, customer connection, and ecosystem of innovation (ACE) model incorporates big data analytics for dynamic product development but follows a linear process and lacks built-in adaptability (Zhan et al., 2017).

Some SIP models address emerging innovation needs more directly. Kruger et al. (2019) emphasize psychological factors such as inertia in problem solving. Roper et al. (2008) focus on knowledge transformation as a pathway to innovation, but both models lack mechanisms for contradiction resolution. In contrast, Sun et al. (2020) propose a TRIZ and OTSM-based SIP tailored for interdisciplinary research, offering recursive feedback and strong contradiction logic.

More recent models extend TRIZ applications further. Wang et al. (2024) propose the Radical Problem-Solving Model for breakthrough innovation, while Mann (2023) applies TRIZ to chaotic and high-risk environments using the "observe, orient, decide, and act" loop for rapid decision-making. These works highlight TRIZ's evolving flexibility, though most still rely on linear or semi-structured flows.

Recent developments such as cyclical TRIZ (Altun, 2025a) and TRIZ reverse (Dewulf et al., 2023; Cosgun and Altun, 2025) aim to address the limitations of linear thinking by introducing feedback loops, scenario-based learning, and layered innovation cycles. Cyclical TRIZ, inspired by the Mayan calendar, structures innovation into short-term, mid-term, and long-term cycles, allowing continuous strategic realignment. TRIZ reverse approaches contradictions retrospectively by analyzing how existing solutions emerged, thus offering insight into hidden design logic.

Recent efforts have also sought to enhance traditional TRIZ methodologies through hybrid approaches incorporating computational intelligence and environmental modeling. Notably, Mohammadi and Zeng (2025) proposed the environment-based design (EBD)-TRIZ-large language model (LLM) model, integrating TRIZ with EBD and LLMs to improve the generation and selection of inventive principles in context-aware scenarios. Their model systematically identifies environmental constraints and opportunities, enhancing the alignment between problem formulation and solution space exploration.

Classical TRIZ is often linear and focused on single-point contradictions, which limits its adaptability in complex and dynamic problem spaces. To overcome these limitations, OTSM-TRIZ was developed as a meta-level extension of TRIZ. It introduces problem networks, meta-contradictions, and fractal logic structures, enabling systemic exploration beyond isolated problems (Sun et al., 2020). OTSM-TRIZ supports recursive reasoning and predictive contradiction handling, but due to its formalism and abstraction, it often requires expert facilitation.

Meanwhile, the Six-Box Scheme proposes a process-oriented model that reflects natural human problem-solving behavior (Nakagawa, 2011, 2016a, 2016b, 2018). It divides the innovation process into six stages: From problem recognition to real-world implementation. This scheme is accessible and effective for interdisciplinary teams but lacks structural mechanisms to analyze contradictions or logical dependencies.

Although a variety of SIPs exist in the literature, a persistent gap remains: no single model fully integrates contradiction-based logic with a cognitive process structure in a recursive and scalable manner. To bridge this gap, COREX is proposed in this study. It introduces a dual-layered model: the inner layer, based on OTSM-TRIZ, structures and analyzes contradictions, while the outer layer, based on the Six-Box Scheme, provides a sequential and cognitively natural flow. This integration allows for both depth and usability.

As summarized in Table 1, COREX differentiates itself by enabling recursive learning, cognitive flow, and contradiction-centered exploration, making it a unique and integrated response to the limitations of existing SIP models.

3. Theoretical Foundations

3.1. OTSM-TRIZ

Classical TRIZ follows a largely linear approach, addressing one contradiction at a time (Altshuller, 1984, 1996). While effective for well-defined problems, this structure limits its adaptability in complex, multi-faceted situations where solving one contradiction often generates new, more intricate ones (Ilevbare et al., 2013). In complex systems, inventive design typically evolves non-linearly: resolving a contradiction may reshape the system and create additional conflicts. This increasing complexity presents significant challenges for classical TRIZ (Elmaraghy et al., 2012), as it typically addresses one contradiction at a time and lacks mechanisms to manage systemic interactions and layered networks of problems (NoPs) (Fiorineschi et al., 2015).

Recognizing this limitation, Altshuller proposed developing a more advanced methodology that could unify diverse problem types under a common problem-solving framework (Khomenko and

Table 1. A comparison of the existing SIP models/approaches

SIP model/approach	Process structure	Contradiction handling	Adaptability/ feedback	Innovation orientation
Classical TRIZ (Altshuller, 1984, 1996)	Linear	Focused on a single contradiction	Limited	Problem-solving
OTSM-TRIZ (Khomenko and Kucharavy, 2002)	Recursive/ fractal	Multi-level, networked contradictions	High	System transformation
Six-Box Scheme (Nakagawa, 2011, 2016a, 2016b, 2018)	Sequential	Absent	Moderate	Cognitive creativity
Cyclical TRIZ (Altun, 2025a)	Cyclical (3-tiered)	Embedded in each cycle	Built-in (short/mid/ long cycles)	Strategic and sustained
TRIZ reverse (Dewulf *et al*., 2023; Cosgun and Altun, 2025)	Inverse deductive	Extracted from solutions	Low	Retrospective learning
SIP (Sheu and Lee, 2011)	Phased	Moderate	Moderate	Cross-phase
Mann's systematic creativity (Mann, 2007)	Phased	Tool-driven	Low	Creative execution
W-model (Brandenburg, 2002)	Phased	Low	Low	Strategic planning
Innovation value chain (Hansen and Birkinshaw, 2007)	Linear	Absent	Low	Idea-to-market
Roper *et al*. (2008)	Phased	Absent	Low	Knowledge conversion
ACE model (Zhan *et al*., 2017)	Linear	None	Low	Data-driven cycles
Kruger *et al*. (2019)	Linear	Moderate	Low	Psychological enablers
IDR (Sun *et al*., 2020)	Recursive	Strong	High	Interdisciplinary
Radical TRIZ (Wang *et al*., 2024)	Linear	Strong	Low	Radical innovation
TRIZ-OODA (Mann, 2023)	Adaptive loop	Moderate	High	Crisis response
EBD-TRIZ-LLM (Mohammadi and Zeng, 2025)	Data-driven/ AI-supported	Context-aware contradiction suggestion via environment modeling	Moderate (via LLM-assisted iteration)	Environment-adaptive ideation
COREX (this study)	Dual-layered (process+logic)	Core mechanism (OTSM-enhanced)	Built-in recursion	Integrated thinking and systemic insight

Abbreviations: ACE: Accelerated innovation process, customer connection, and ecosystem of innovation; AI: Artificial intelligence; COREX: Contradiction-oriented exploration; EBD: Environment-based design; IDR: Interdisciplinary research; LLM: Large language models; OODA: Observe, orient, decide, and act; OTSM: General Theory of Powerful Thinking; SIP: Systematic innovation process; TRIZ: Theory of Inventive Problem Solving.

Ashtiani, 2007). This idea laid the groundwork for OTSM, a meta-level evolution of TRIZ introduced by Khomenko and Kucharavy in the 1980s (Khomenko and Kucharavy, 2002).

OTSM-TRIZ introduces several key conceptual tools, such as:

(i) NoP modeling to manage interdependent problems
(ii) Meta-contradiction analysis to reveal structural barriers to innovation
(iii) Fractal reasoning for recursive exploration of problem layers
(iv) Predictive evaluation to anticipate side effects of proposed solutions.

Cavallucci et al. (2015) applied OTSM-TRIZ in helicopter assembly processes to map system-wide decisions and anticipate the cascading impact of design choices. Khomenko et al. (2009) employed OTSM's NoP to support complex research and development (R&D) strategies in the energy sector, enabling interdisciplinary collaboration and long-term innovation planning.

Fiorineschi et al. (2015) compared OTSM-TRIZ with classical TRIZ in the conceptual design of a stratospheric gondola. Their study showed that OTSM-TRIZ's hierarchical decomposition of complex systems greatly improved the management of system-wide interrelations. Moreover, Borgianni et al. (2015) integrated OTSM-TRIZ with decision-making models to enhance the evaluation of design concepts in high-stakes innovation projects.

Together, these examples demonstrate that OTSM-TRIZ not only improves problem-solving efficiency but also strengthens strategic decision-making and innovation management across sectors.

3.1.1. Network of problems

The NoP framework maps the relationships between design parameters, intermediary parameters, and need parameters, forming a web of interactions where inconsistencies emerge as structural contradictions (Eltzer et al., 2006). These inconsistencies arise when the desired values of interconnected parameters conflict—improving one need parameter may compromise another. This systemic inconsistency constitutes the global problem, which cannot be resolved through linear reasoning alone (Fig. 2).

The network of problems offers a structured way to represent and explore these contradictions. Each node in the network corresponds to a parameter or function, and each edge represents a dependency or influence.

Through this structure, designers can identify:

(i) Where contradictions emerge (e.g., conflicting design goals)

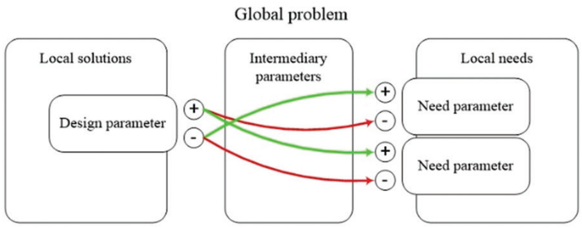


Fig. 2. Network of problems

(ii) How changes in one part of the system propagate elsewhere
(iii) Where inventive principles or separation strategies could be applied.

Khomenko and Ashtiani (2007) emphasize that NoP transforms loosely defined or chaotic design challenges into a logically organized architecture of problems. By formalizing system behavior through parameter relationships, the NoP model supports systematic problem decomposition, enabling a clearer definition of sub-problems and meta-contradictions.

Within the COREX methodology, NoP plays a central role in structuring problems and mapping contradictions. It forms the analytical backbone that enables teams to move beyond surface-level symptoms and uncover deeper logical tensions. This elevates COREX from a toolset to a comprehensive thinking system capable of structuring and transforming complex design challenges.

3.1.2. Meta-contradiction analysis

In many innovation scenarios, contradictions do not exist independently; they stem from deeper systemic constraints within the problem structure. Meta-contradictions refer to these higher-order conflicts that arise when multiple local contradictions interact or when a problem resists simplification into a single conflict (Khomenko and Ashtiani, 2007).

In OTSM-TRIZ, meta-contradiction analysis provides a strategic mechanism for identifying what makes a system inherently resistant to innovation. These contradictions often reflect competing system-level goals (e.g., maximizing customization while minimizing production complexity). Unlike classical TRIZ, which focuses primarily on technical or physical contradictions, OTSM-TRIZ uses meta-contradiction trees and problem parameter hierarchies to locate deeper innovation barriers.

This concept allows designers to go beyond direct contradiction resolution and target the root structural limits of the system. Fiorineschi et al. (2015) highlight that identifying meta-contradictions was essential in managing the layered constraints of high-altitude aerospace design. Thus, meta-contradiction analysis not only guides problem-solving but also redefines the boundaries of innovation.

3.1.3. Fractal reasoning

A major strength of OTSM-TRIZ is its use of "fractal reasoning" to handle complex, layered problem spaces (Khomenko and Kucharavy, 2002). In this approach, each problem is viewed as part of a larger system and may itself contain subsystems and nested contradictions. This mirrors the structure of complex real-world problems, where solving one issue often reveals a cascade of related sub-problems.

Rather than stopping after solving a single contradiction, OTSM-TRIZ promotes recursive exploration: testing solutions for hidden contradictions and tracing their implications across system layers. This principle reflects the non-linear nature of innovation, which is better understood as the ongoing reconfiguration of system constraints and opportunities.

3.1.4. Predictive evaluation

Predictive evaluation is based on cause–effect chain analysis and future scenario modeling (Khomenko et al., 2009). These tools help anticipate potential side effects and verify whether a solution resolves the core problem or merely shifts it elsewhere.

In the context of energy systems, Cavallucci et al. (2015) demonstrated how predictive analysis within OTSM-TRIZ/Inventive Design Method-TRIZ improved long-term decision robustness in helicopter assembly planning. Using NoP-based propagation models, they identified secondary contradictions before implementation.

3.2. Six-Box Scheme

The Six-Box Scheme, proposed by Toru Nakagawa as the foundation of his broader Creative Problem-Solving framework, offers a systematic structure for navigating the stages of inventive thinking. While classical TRIZ and many scientific problem-solving models rely on the Four-Box Scheme (specific problem, abstract problem, abstract solution, and specific solution), the Six-Box Scheme expands this into a more process-aware structure. As shown in Fig. 3, it provides a comprehensive map of the creative thinking journey, from initial problem recognition to real-world implementation (Nakagawa, 2011, 2016a, 2016b, 2018).

Unlike conventional flowchart-based approaches, which focus primarily on the order and execution of procedures, the Six-Box Scheme is based on a dataflow philosophy (Nakagawa, 2011). This means that the model emphasizes the types of information to be obtained, transformed, and delivered at each stage. Traditional flowcharts concentrate on "what to do and when," often losing sight of "what to know and why." In contrast, the Six-Box Scheme views innovation as an information-centered transformation process: from vague real-world observations to well-defined problems, from system analysis to idea generation, and from conceptual design to implementation. Each box corresponds not only to a step in the innovation journey but also to a knowledge state defining the type of insight that must be achieved at that stage (Nakagawa, 2016a).

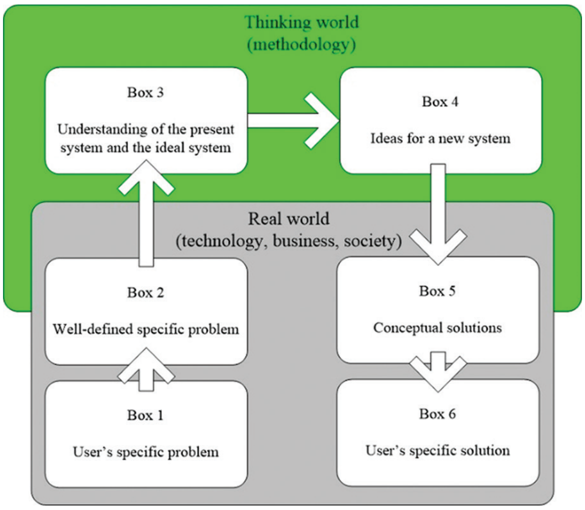


Fig. 3. The Six-Box Scheme

3.2.1. A dual domain perspective

A key conceptual advancement in the Six-Box Scheme is its separation of the innovation process into two cognitive domains:

(i) The "real world," where problems originate, and solutions are ultimately implemented (Boxes 1, 2, and 6),
(ii) The "thinking world," where abstract analysis and creative reasoning take place (Boxes 3, 4, and 5).

In this structure: Box 1 captures the user's initial complaint or observation. Box 2 formulates a well-defined problem. Boxes 3–5 reflect the process of structured idea development within the thinking world. Finally, Box 6 delivers actionable outputs back into reality.

This separation encourages practitioners to temporarily step away from immediate pressures and constraints and enter a focused, reflective mode of creative exploration. In collaborative environments, it also supports group ideation sessions, workshops, and structured co-creation.

3.2.2. The three macro processes

Nakagawa (2011) further identifies three macro-level processes within the Six-Box Scheme:

(i) Problem definition process (real world)
- Recognizing the user's concern (Box 1),
- Framing a solvable, structured problem (Box 2)

(ii) Creative problem-solving process (thinking world)
- Analyzing the present and the ideal system (Box 3)
- Generating new ideas (Box 4)

(iii) Constructing conceptual solutions (Box 5)
- Solution implementation process (real world)

(iv) Deploying the solution into actual systems or products (Box 6).

This clear division of phases enhances the clarity, teachability, and transferability of the model. Each stage aligns with a specific cognitive goal, allowing for modular adaptation across various real-life innovation workflows.

4. Proposed Methodology

4.1. A Dual-layered Structure

In COREX, OTSM-TRIZ serves as the logical engine responsible for mapping contradictions, modeling problem networks, and reasoning through recursive system behavior. Specifically, OTSM-TRIZ contributes to the proposed methodology in the following ways:

(i) Problem structuring (Box 2): The user's problem is not only redefined but also modeled through the NoP, enabling the identification of systemic inconsistencies and the decomposition of complex design goals

(ii) System analysis (Box 3): The present system and ideal system are explored through meta-contradiction analysis, fractal modeling, and parameter dependency mapping, providing structured abstraction and insight into the root logic of the problem

(iii) Idea generation (Box 4): Contradictions identified in earlier steps are resolved using inventive strategies drawn from TRIZ principles, enhanced by OTSM's predictive and recursive logic.

By embedding these mechanisms into COREX, contradiction handling becomes a continuous, feedback-driven process, rather than a single-point resolution effort.

While OTSM-TRIZ structures the internal reasoning process, the Six-Box Scheme forms the external procedural and cognitive structure through which users interact with the problem-solving journey. It does so by:

(i) Providing a stage-wise process flow aligned with human creative cognition, making COREX accessible to non-experts and cross-functional teams

(ii) Guiding practitioners through problem definition (boxes 1 and 2), structured analysis and creativity (boxes 3–5), and real-world implementation (Box 6) in an iterative fashion

(iii) Ensuring that each logical insight from OTSM-TRIZ is cognitively processed and practically translated into action within a broader innovation workflow.

The Six-Box Scheme allows COREX to function not merely as a logical tool but as a "thinking environment," supporting decision-making, team collaboration, and strategic innovation deployment.

4.2. Algorithm of the Proposed Methodology

Table 2 presents the COREX implementation algorithm, describing each phase, its function, and the tools involved. The approach is designed for real-world innovation teams seeking both conceptual clarity and logical depth when addressing complex, contradiction-rich challenges.

In this methodology, steps 1 and 2 are carried out primarily in the real world, reflecting direct user needs and constraints. Steps 3–5 occur within the thinking world, where abstraction, contradiction modeling, and ideation are guided by OTSM-TRIZ reasoning. Steps 6 and 7 transition back to the real world, where implementation and validation take place. Step 8 represents the recursive structure of COREX, allowing for continuous refinement and adaptation—an essential feature for complex and evolving systems.

5. Case Study

To illustrate the practical value of the COREX methodology, this section applies it to a real-world design challenge in the context of autonomous vehicles (Altun et al., 2022; Altun, 2023; Altun, 2025b; Kim, 2021). As driver responsibility decreases, the vehicle cabin must evolve to support new expectations of comfort, flexibility, and usability. However, these demands often introduce contradictions, especially between passenger comfort and system safety. The following use case illustrates how COREX can be applied step-by-step to structure, analyze, and resolve such contradictions.

5.1. Step 1 – Capture the Initial Problem (Box 1)

User observations and feedback from autonomous vehicle prototypes indicate a recurring concern: "I want to relax and recline fully during the ride, but I'm not sure if I will be safe in case of an accident."

This vague concern reflects a contradiction between comfort and safety (Fig. 4). Although no explicit failure has occurred, the user's hesitation signals an unresolved design problem. The goal at this step is to capture the user's discomfort and translate it into a design challenge suitable for structured exploration.

5.2. Step 2 – Define the Problem Structure (Box 2)

In this step, the vague user concern is transformed into a well-defined and analyzable design problem. To achieve this, the team applies NoP modeling to systematically map interactions among design parameters, intermediary parameters, and need parameters.

Table 2. Step-by-step contradiction-oriented exploration methodology

Steps	**Six-Box ref.**	**Cognitive objective**	**Tool (s) used**
Step 1: Capture the initial problem	Box 1	Recognize vague user concern	- Informal observation - User feedback
Step 2: Define the problem structure	Box 2	Formulate a well-defined and measurable problem	- NoP modeling - Parameter mapping
Step 3: Analyze present and ideal systems	Box 3	Explore both current and desired system states	- Meta-contradiction analysis - Fractal reasoning
Step 4: Identify core contradictions	Box 3 (continued)	Reveal systemic conflicts and inconsistencies	- TRIZ contradiction matrix - Root conflict analysis
Step 5: Generate inventive solutions	Box 4	Produce ideas to resolve core contradictions	- TRIZ inventive principles - Separation strategies
Step 6: Construct conceptual solutions	Box 5	Organize ideas into feasible concepts	- Constraint analysis - Predictive evaluation
Step 7: Implement and validate	Box 6	Deploy the solution in a real context	- Evaluation criteria (business/technical) - Feedback loop
Step 8: Recursive feedback loop	(Back to Box 1)	Reframe the problem if unresolved or evolving	- Recursive OTSM modeling - NoP update

Abbreviations: NoP: Network of problems; OTSM: General Theory of Powerful Thinking; TRIZ: Theory of Inventive Problem Solving.

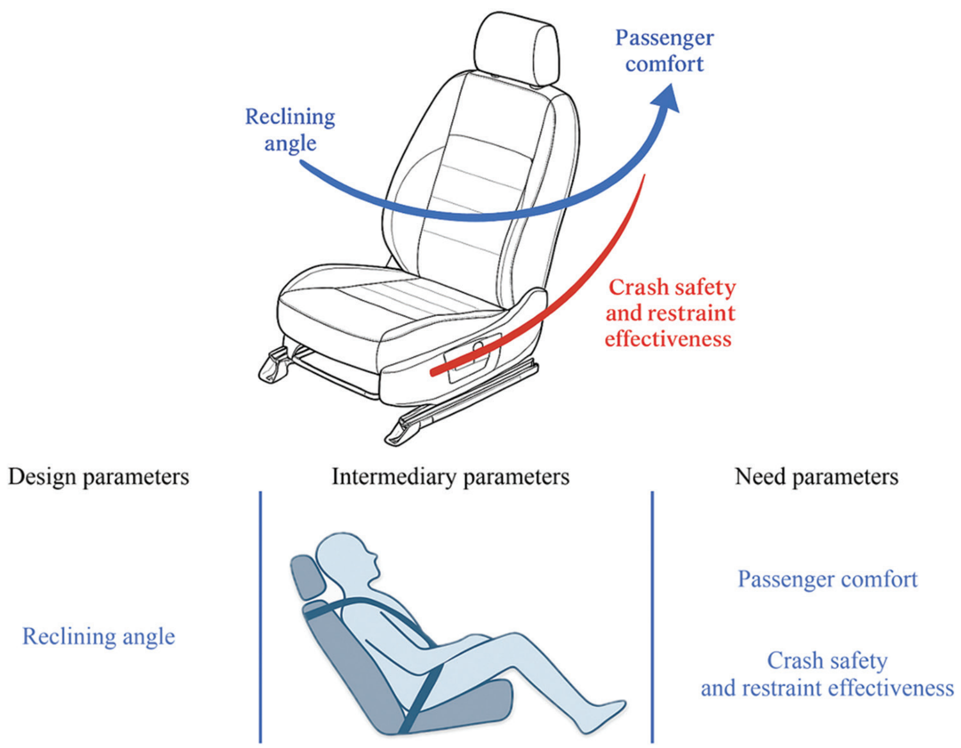


Fig. 4. Initial contradiction between comfort and safety

As illustrated in Fig. 5, the NoP of an automotive front-seat system reveals a complex web of interdependencies. Design parameters such as reclining angle, seat shape, headrest design, and seatbelt configuration influence intermediary attributes, including seat-back shape, foam density, and anchor points. These intermediary attributes, in turn, contribute to various need parameters such as comfort, safety, ergonomics, durability, and adjustability.

In the specific case of the reclining seat problem:

- Design parameter: Reclining angle
- Need parameters: Passenger comfort (positively affected), restraint system safety (negatively affected)
- Intermediary parameters: Seatbelt anchor position, headrest geometry, seat adjustability mechanism, seat foam density, and seat-back angle.

An increase in reclining angle improves comfort but compromises the alignment and effectiveness of seatbelt and airbag systems, thus reducing safety. This trade-off leads to a structural contradiction: optimizing one need parameter (comfort) degrades another (safety), with the conflict propagated through intermediary parameters.

This form of contradiction is non-local, meaning that it does not reside within a single parameter but emerges from cross-linked dependencies. As such, it cannot be resolved through linear optimization or isolated improvement. Instead, it requires multi-parametric adjustment and inventive restructuring, which COREX supports through its recursive and logic-based exploration process.

Using the NoP structure, as shown in Fig. 5, the design team is able to: (i) Make hidden parameter relationships visible, (ii) Identify and quantify the tension points within the system, and (iii) Prepare a structured basis for contradiction modeling in the following COREX step.

Although Fig. 5 visualizes the internal structure of the NoP, the resulting contradictions extend beyond isolated parameter interactions and form a hierarchical structure consistent with OTSM-TRIZ logic. At the first level, local contradictions arise between directly linked parameters. For example, increasing the reclining angle improves comfort but misaligns the seatbelt anchor, while softer foam density enhances pressure distribution but reduces structural stability. As these tensions propagate through intermediary attributes (such as seat-back shape, foam density, and anchor positioning), they evolve into technical contradictions, where improving one engineering attribute worsens another (e.g., increased adjustability reduces long-term durability; improved cushioning decreases crash stiffness). Some tensions span functional domains and become non-local

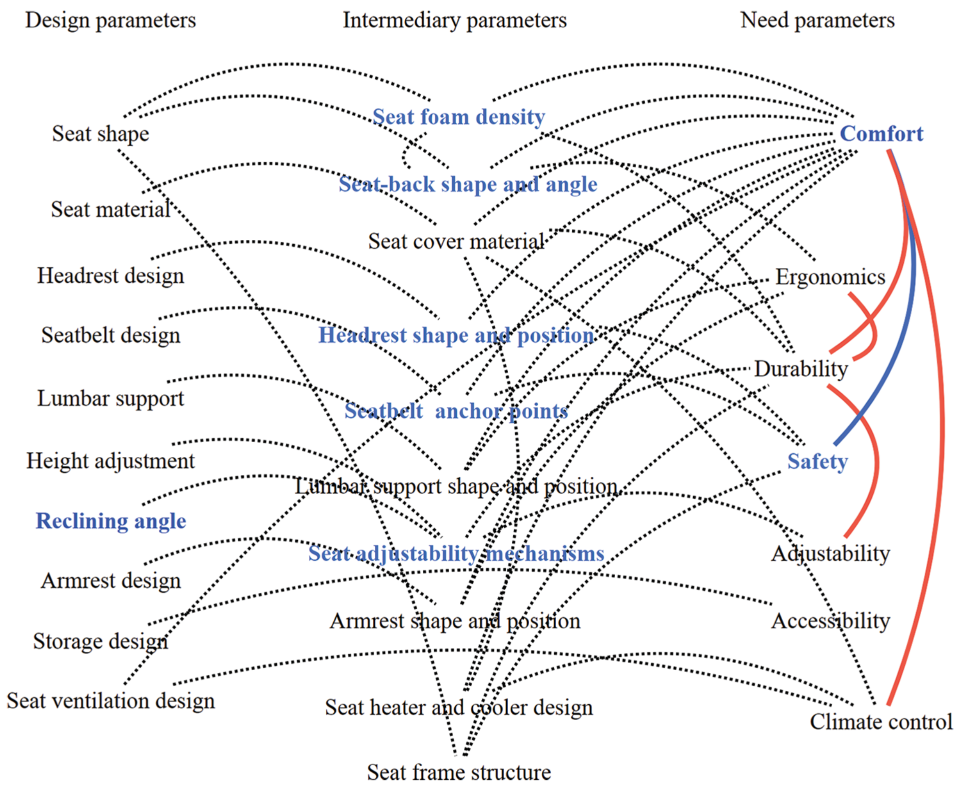


Fig. 5. Network of problems structure for front-seat systems in automotive design. Solid black arcs represent direct parameter dependencies; red arcs indicate aggregated conflict areas between key need parameters; blue elements represent the main influential parameters of the case study

contradictions, such as enhanced ergonomic contouring requiring additional sensors and thus increasing privacy intrusion. These accumulated tensions converge into a system-level meta-contradiction, as illustrated in Fig. 5 by the diverging need-parameter curves: Comfort requires softness, adaptability, and spatial freedom, whereas safety and durability require rigidity, reliable alignment, and controlled posture.

This layered structure clarifies how parameter-level dependencies within the seat system give rise to higher-order conflicts. The contradictions in this case can be grouped as follows:

(i) Local contradictions
- Reclining angle versus seatbelt anchor alignment
- Foam density/seat shape versus pressure distribution stability

(ii) Technical contradictions:
- Improved adjustability mechanisms reduce long-term durability
- Increased foam softness lowers crash stiffness
- Improved climate responsiveness increases energy use

(iii) Non-local contradictions
- Enhanced ergonomic shaping increases privacy intrusion (due to added sensing)
- Automated posture correction reduces perceived autonomy
- Improved deformation recovery reduces vibration comfort

(iv) Meta contradiction
- The comfort subsystem (e.g., reclining angle, foam softness, adaptability) conflicts with the safety-durability subsystem (e.g., structural rigidity, alignment stability, restraint geometry)

Similar multi-layered tensions were demonstrated in a previous OTSM-TRIZ-based front-seat design study (Altun, 2025c), reinforcing the relevance of network-driven contradiction mapping for automotive seating systems. Making these hierarchical relationships explicit strengthens the connection between the NoP structure in Fig. 5 and the contradiction modeling performed in the subsequent steps.

It should be noted that several contradictions discussed later (e.g., structural rigidity vs. flexibility; personal freedom vs. system control; sensor feedback vs. crash response time) represent higher-level abstractions derived from the combinations of the design and intermediary parameters, as shown in Fig. 5, rather than new or independent parameters.

5.3. Step 3 – Analyze Present and Ideal Systems (Box 3)

The present system is based on conventional seat designs optimized for upright posture, where seatbelts and airbags operate reliably. The ideal system would allow passengers to rest in any desired posture without compromising safety.

At this stage, meta-contradiction analysis is applied. The root contradiction lies not only between comfort and safety but also in the shared spatial domain: both demands are imposed on the same physical structure (the seat). Using fractal reasoning, the contradiction is decomposed into:

(i) Constraint conflicts within the seat design (structural rigidity vs. flexibility),
(ii) System-level conflicts (personal freedom vs. system control),
(iii) Cross-domain dependencies (sensor feedback vs. crash-response time).

This layered analysis allows the design team to reframe the problem beyond surface symptoms.

To clarify how the three conflict types emerge, a simplified root conflict analysis for the reclining-seat problem was conducted. The key issue—loss of restraint-system alignment during deep recline—was traced back to several first-level causes: fixed seatbelt anchor geometry, posture-dependent changes in headrest and airbag positions, and reduced structural stability when the backrest moves far from upright. All mechanisms point to a common root conflict: current safety systems are designed for a fixed posture, while comfort demands wide and dynamic posture variability. This explains the first conflict category: structural rigidity versus flexibility.

Fractal reasoning clarifies the remaining two conflict categories. At the system level, increasing recline freedom reduces the predictability required for safety systems—thus, personal freedom versus system control. At the cross-domain level, sensor quality, posture-detection accuracy, and required crash-response timing interact, creating reliability issues when posture deviates from the standard. These repeating tensions across physical structure, user behavior, and system timing show that the comfort–safety contradiction recurs at multiple scales. This confirms that the problem is systemic, requiring multi-level restructuring rather than isolated adjustments.

5.4. Step 4 – Identify Core Contradictions (Box 3 Continued)

The key contradiction is defined between "ease of use (comfort)" and "loss of information or system complexity" (safety systems handling crash data). The contradiction matrix suggests inventive principles (IPs): IP1 – Segmentation; IP24 – Intermediary; IP3 – Local quality; and IP13 – The other way round. Root conflict analysis further reveals that existing safety systems are built for static configurations and lack adaptability.

5.5. Step 5 – Generate Inventive Solutions (Box 4)

Several inventive ideas are generated based on the contradiction logic:

(i) IP24 – Intermediary: Introduce an active support module beneath the seat surface (Fig. 6A). When a crash is imminent (detected via sensor data), the module stiffens and repositions load-bearing regions (e.g., lumbar, side wings) to brace the passenger

(ii) IP3 – Local quality: Embed shape-memory alloys or air-cell actuators in selective contact zones (Fig. 6B), allowing local flexibility while maintaining rapid, localized rigidity

(iii) IP1 – Segmentation: Divide the seat into independently adjustable/lockable modules (e.g., headrest, torso, hip, leg) responding to real-time posture and crash dynamics (Fig. 6C)

(iv) IP13 – The other way round: Instead of adapting the seat to fixed safety systems, redesign the seatbelt and airbag geometry to follow the user's posture (Fig. 6D).

These concepts are not random ideation but systematically derived from the contradiction analysis conducted in prior steps.

5.6. Step 6 – Construct Conceptual Solutions (Box 5)

The most promising concepts are selected and integrated into a system-level solution:

(i) A smart seat system with posture sensors, real-time crash-prediction algorithms, shape-adaptive materials, and repositionable airbags

(ii) Constraint analysis ensures feasibility across mechanical, electrical, and timing domains

(iii) Predictive evaluation uses simulations to test the activation time of support modules relative to average crash speed and occupant behavior.

This stage yields a functional concept ready for prototyping and validation.

5.7. Step 7 – Implement and Validate (Box 6)

The solution is implemented in a concept vehicle cabin. Validation includes: (i) Crash simulations across

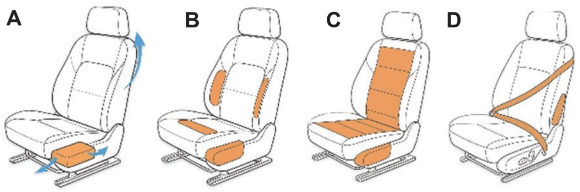


Fig. 6. Conceptual ideas generated based on the contradiction logic. (A) Active intermediary support; (B) Local adaptive contact zones; (C) Segmented modular seat structure; (D) Reversed safety adaptation (seatbelt–airbag following posture)

seating positions, (ii) comfort assessments in prolonged autonomous travel scenarios, and (iii) evaluation of cost, weight, and manufacturability.

Results show increased comfort with crash-dummy injury metrics remaining within acceptable thresholds.

5.8. Step 8 – Recursive Feedback Loop

Even after implementing the smart-seat solution, new contradictions emerge, indicating that the innovation process is ongoing. This step captures these emerging concerns and guides the system back into a new COREX cycle.

As adaptive systems gain autonomy, contradictions that arise are no longer confined to technical or ergonomic trade-offs but extend into the ethical domain, where values such as autonomy, privacy, consent, and trust frequently collide. In these situations, safety-enhancing automated actions may diminish perceived user control; data-driven personalization may improve comfort while raising concerns about monitoring and psychological acceptability. Treating these tensions as socio-technical contradictions allows COREX to articulate and formalize them (e.g., increasing automated intervention improves safety but reduces perceived autonomy) and convert them into new needs and intermediary parameters such as transparency, consent logic, or user override mechanisms. This ensures that ethical considerations are incorporated into the structured reasoning process rather than added informally, aligning the methodology with contemporary perspectives in design ethics and responsible innovation.

However, two major contradictions appear:

(i) User autonomy vs. system control – "What if I don't want the seat to move automatically?" This reflects a tension between user privacy and freedom, and the automatic system overrides required for safety

(ii) Ethical tension: Passive consent versus forced adjustment – raising the question of whether the system should forcibly adjust a passenger's posture prior to a crash, even if the user resists or is unaware. This introduces broader ethical considerations related to consent, trust, and the psychological acceptability of automated interventions.

Thus, the original comfort–safety contradiction evolves into behavioral and ethical contradictions. The problem must be reframed with new need parameters (user control, transparency, trust) and new intermediary parameters (consent logic, user override interface, behavior-prediction modules).

This reframing brings the process back to Box 1, where the updated problem can be captured as a new user concern. From here, a second iteration of COREX begins, applying the same step-by-step method but now targeting a different dimension of the challenge: not only technical performance, but human–system interaction. This recursive structure is what enables COREX to adapt to evolving systems—technically, socially, and ethically.

Although the COREX logic supports multiple recursive cycles, the present simulation study did not extend the process beyond the first loop. The aim of this work was not to exhaustively explore all possible iterations, but to compare methodological behavior under identical conditions. Given that the initial cycle already revealed a shift from a technical contradiction (comfort vs. safety) to a socio-technical one (autonomy, consent, and system control), the primary objective, demonstrating how COREX uncovers deeper layers of the problem, was achieved. In practice, a new COREX cycle is initiated only when emerging contradictions introduce new needs or intermediary parameters that alter the structure of the problem. If subsequent contradictions merely refine existing parameters without reshaping the conflict architecture, the process is considered converged, and further iterations are unnecessary.

6. Comparison of the Methodologies

To demonstrate how the proposed COREX methodology addresses complex design challenges more effectively than existing approaches, all three methods (OTSM-TRIZ, the Six-Box Scheme, and COREX) were applied to the same design problem: the development of an adaptive seat system for autonomous vehicles. This problem was intentionally selected due to its multi-domain nature.

To evaluate the practical performance of the methods, an empirical simulation study was conducted within the R&D department of a tier-1 automotive seat supplier. The purpose of this exercise was not to build an actual seat system but to assess and compare the methodologies under realistic design conditions. The teams worked with persona-based scenarios, system constraints, and structured design discussions. Thus, the insights presented here reflect methodological behavior rather than engineering feasibility testing.

The design task confronted the teams with multiple conflicting requirements. Passengers expected superior comfort, adaptive postural support, and temperature responsiveness during extended autonomous trips. These expectations conflicted with constraints related to structural durability, energy efficiency, manufacturability, and cost. Such tensions made the problem an ideal test bed for contradiction-oriented methods.

Table 3 presents a comparative overview of how each methodology approached the same problem and the specific steps taken during its application.

Table 3. Comparison of the methodological approaches

Stage	OTSM-TRIZ	Six-Box Scheme	COREX
Problem identification	Uses NoP to link discomfort to design parameters (seat hardness, thermal adaptation, load-bearing)	User complaints are framed in Box 1 and translated into clarified needs in Box 2	Integrates user discomfort from boxes 1 and 2 with NoP mapping to trace contradictions across physical and psychological needs
Problem modeling	Constructs a parameter network including design parameters, intermediary parameters, and need parameters	Performs present-vs-ideal system analysis (Box 3) for thermal and ergonomic mismatches	Runs NoP modeling and Box 3 analysis in parallel to capture layered, real-world tensions
Contradiction analysis	Identifies a meta-contradiction (e.g., adjustable softness reduces durability) and applies meta-level resolution tools	Contradictions are inferred from user needs but are not addressed through a systematic resolution process	Maps the contradiction matrix (e.g., speed vs. energy consumption) and augments it with meta-level insights added through structural contradiction mapping
Solution generation	Uses fractal reasoning and abstraction to propose concepts such as shape-memory foams or layered materials	Box 4 brainstorming produces ideas, including air-cushion systems and AI-controlled morphing surfaces	Applies inventive principles (e.g., IP35: parameter change; IP28: mechanical substitution) to develop adaptive layered structure
Evaluation	Predicts system impact via NoP updates, though real-world feedback mechanisms are limited	The transition from boxes 5 to 6 focuses on concept translation toward prototyping	Uses predictive evaluation tools and real-world constraints; identified failure points feed back into boxes 1 and 2 for recursion

Abbreviations: AI: Artificial intelligence; COREX: Contradiction-oriented exploration; NoP: Network of problems; OTSM-TRIZ: General Theory of Powerful Thinking–Theory of Inventive Problem Solving.

6.1. Study Design and Evaluation

This empirical study was conducted in collaboration with the R&D department of a tier-1 automotive seat supplier. Three interdisciplinary design teams were formed, each consisting of four members with complementary expertise in engineering design, ergonomics, and design thinking. Each team was assigned and formally trained in one of the three innovation methodologies evaluated in this study: OTSM-TRIZ, the Six-Box Scheme, and COREX. All teams worked on the same design problem: developing an adaptive seat system for autonomous vehicles capable of balancing comfort, safety, and energy efficiency.

The evaluation followed a structured four-step procedure:

(i) Step 1: A common scenario and design brief were introduced to all teams to ensure comparable starting conditions
(ii) Step 2: Teams applied their assigned methodology in a controlled setting over a four-hour design session
(iii) Step 3: Design outcomes were documented, including sketches, assumptions, contradiction formulations, intermediate models, and proposed system concepts
(iv) Step 4: Upon completion, each participant completed a structured assessment survey evaluating the methodology they used.

The assessment was based on six criteria reflecting both analytical and human-centered aspects of innovation work:

(i) Clarity of problem modeling – Ability to represent the design problem and contradictions
(ii) Contradiction resolution – Effectiveness in identifying and resolving systemic conflicts
(iii) Creativity and feasibility – Novelty and implementability of proposed concepts
(iv) User-centeredness – Degree to which user needs, personas, and scenarios informed the solution
(v) Cognitive load – Perceived mental effort required to apply the method
(vi) Integration capability – Ability to synthesize perspectives across domains (e.g., ergonomics, mechanics, materials, and user requirements).

Each criterion was rated individually on a 1–5 Likert scale (1 = very low, 5 = very high). Table 4 provides the ratings. To statistically examine differences between methods, a nonparametric analysis was conducted using Kruskal–Wallis omnibus tests followed by Mann–Whitney U post-hoc comparisons (Table 5). These analyses revealed significant methodological differences across most criteria, particularly in contradiction resolution, integration capability, user-centeredness, and cognitive load.

The results indicate that COREX combines the systemic depth of OTSM-TRIZ with the human-centered clarity of the Six-Box Scheme. COREX received the highest evaluations in contradiction resolution and integration—two key capabilities for socio-technical systems such as autonomous vehicle seating. While the Six-Box Scheme supported creativity and user-focused reasoning, it lacked analytical depth. OTSM-TRIZ exhibited strong modeling capability but required higher cognitive effort and provided limited support for user-integrated iteration. COREX therefore achieved a functional balance between analytical rigor and design usability, although future research may

Table 4. Evaluation scores across innovation methodologies

Model	Participant	Problem modeling	Contradiction resolution	Creativity	User-centeredness	Cognitive load	Integration
OTSM-TRIZ	O1	4	4	4	3	2.5	4
	O2	5	5	4	3	2	4
	O3	4	4	3	2	2	3
	O4	4	5	3.5	3	2.5	4
Six-Box Scheme	S1	4	3	4	5	4	3
	S2	3	3	4.5	5	4	4
	S3	3.5	2.5	4	4	4.5	3.5
	S4	3.5	3	4	4.5	4	3
COREX	C1	5	5	5	5	4	5
	C2	5	5	4	5	4	5
	C3	4.5	4	4	4.5	4	5
	C4	5	5	5	4	3.5	4.5

Note: "Cognitive load" was reverse-coded for clarity (higher = easier use).
Abbreviations: COREX: Contradiction-oriented exploration; OTSM-TRIZ: General Theory of Powerful Thinking–Theory of Inventive Problem Solving.

Table 5. Kruskal–Wallis and Mann–Whitney U test results across the three methodologies

Criterion	Comparison	Median difference	W	*p*-value	Adjusted *p*	Interpretation
Problem modeling _Kruskal–Wallis H (df=2): 8.17_ _p (adj): 0.017*_	COREX vs. OTSM	1.0	23.5	0.0745	0.0569	Marginal (COREX > OTSM)
	COREX vs.Six-Box	1.5	26.0	0.0152	0.0128*	Significant (COREX > Six-Box)
Contradiction resolution _Kruskal–Wallis H (df=2): 8.36_ _p (adj): 0.015*_	COREX vs.OTSM	0.0	20.0	0.3325	0.3042	No difference
	COREX vs.Six-Box	2.0	26.0	0.0152	0.0114*	Significant (COREX > Six-Box)
Creativity _Kruskal–Wallis H (df=2): 4.95_ _p (adj): 0.084_	COREX vs.OTSM	1.0	24.0	0.0562	0.0443	Slightly significant (COREX > OTSM)
	COREX vs.Six-Box	0.25	21.0	0.2352	0.2023	Not significant
User-centeredness _Kruskal–Wallis H (df=2): 7.82_ _p (adj): 0.020*_	COREX vs.OTSM	2.0	26.0	0.0152	0.0128*	Significant (COREX > OTSM)
	COREX vs.Six-Box	0.0	18.0	0.5000	0.5000	No difference
Cognitive load (inv.) _Kruskal–Wallis H (df=2): 9.02_ _p (adj): 0.011*_	COREX vs.OTSM	1.5	26.0	0.0152	0.0123*	Significant (COREX easier)
	COREX vs.Six-Box	0.0	14.5	n.s.	n.s.	No difference
Integration _Kruskal–Wallis H (df=2): 8.38_ _p (adj): 0.015*_	COREX vs.OTSM	1.0	26.0	0.0152	0.0114*	Significant (COREX > OTSM)
	COREX vs.Six-Box	1.5	26.0	0.0152	0.0128*	Significant (COREX > Six-Box)

Note: *indicates $p<0.05$.
Abbreviations: adj: Adjusted; COREX: Contradiction-oriented exploration; df: Degree of freedom; inv.: Inverted; n.s.: not significant; OTSM: General Theory of Powerful Thinking.

explore tool support or training aids to further reduce cognitive load.

6.2. Validity and Reliability of the Evaluation

To ensure the methodological soundness of the empirical comparison, particular attention was given to the validity and reliability of the evaluation process. The assessment involved 12 participants, organized into three independent design teams trained respectively in OTSM-TRIZ, the Six-Box Scheme, and COREX. Each team received the same design brief, scenario, and constraints, and all sessions were conducted in a controlled environment. Ratings were collected individually and anonymously immediately after task completion to minimize social desirability and conformity biases.

Construct validity was supported through the use of six well-defined evaluation criteria, each aligned with theoretical constructs widely used in innovation, design, and TRIZ research. To assess internal reliability, Cronbach's alpha was calculated across the six criteria, resulting in $\alpha = 0.699$. Considering the small sample size and the exploratory nature of design-team studies, this level of internal consistency is regarded as acceptable. Because Likert-scale data are ordinal and sample sizes were modest, inferential validity was strengthened through the use of nonparametric statistical analyses. Accordingly, Kruskal–Wallis omnibus tests and Mann–Whitney U *post hoc* comparisons were applied to identify significant differences among the three methodologies.

Potential sources of bias (e.g., varying familiarity with specific methods or individual cognitive differences) were mitigated through standardized training, consistent timing, homogeneous team structure, and independent evaluation rather than consensus-based scoring. Together, these measures reinforce the validity and reliability of the evaluation framework and support the robustness of the comparative findings as reported in Table 5.

6.3. OTSM-TRIZ

The OTSM-TRIZ method focused on constructing an NoP, mapping user discomfort to underlying design parameters such as foam stiffness, thermal regulation, and deformation memory.

It effectively identified a key meta-contradiction: "To improve comfort, the seat must be soft and adaptive; but to ensure durability, it must be firm and stable." This insight enabled the generation of high-level concepts (e.g., memory foams or adaptive materials). However, the approach remained largely abstract, lacked structured evaluation steps, and did not easily connect with user feedback loops or iterative prototyping stages.

6.4. Six-Box Scheme

The Six-Box Scheme structured the innovation process as a linear, intuitive flow. The user problem was first captured as a complaint (Box 1), then refined into a specific need (Box 2), and finally through a present-vs.-ideal system comparison (Box 3).

Creative brainstorming in Box 4 yielded concepts such as artificial intelligence (AI)-controlled morphing surfaces or air-cushion systems. While this approach supported ideation and facilitated team creativity, it lacked analytical tools to identify or resolve deep contradictions. In particular, it provided no formal mechanism to explain why certain needs conflicted or how trade-offs could be resolved systematically.

6.5. COREX

The COREX approach combined the strengths of both systems. It began by mapping user discomfort (boxes 1 and 2) into a structured NoP model, linking design parameters such as padding softness, thermal response, and pressure distribution.

In Box 3, COREX applied both the TRIZ contradiction matrix and meta-contradiction analysis to identify conflicts, including: "increased adjustability reduces long-term reliability" and "comfort vs. safety during crash scenarios."

Using IPs—specifically IP35 (parameter change) and IP28 (mechanics substitution)—the team generated a multi-layered adaptive seat system comprising a durable memory-foam base layer, a gel-based comfort top layer, and smart fabrics for climate control.

These concepts were validated using predictive evaluation tools (boxes 5 and 6) and tested through user scenarios. The solution improved comfort while maintaining energy consumption within targets. Importantly, real-world feedback (e.g., insufficient neck support limitations for older users) was fed back into Box 1 for system-level adjustments, demonstrating COREX's recursive power.

6.6. Critical Comparison and Theoretical Implications

The OTSM-TRIZ method and the Six-Box Scheme represent two distinct but incomplete approaches to early-stage problem structuring. OTSM-TRIZ provides rigorous tools for identifying systemic contradictions, yet its high level of abstraction and representational complexity limit its ability to incorporate user perspectives and scenario-based insights. In contrast, the Six-Box Scheme excels at capturing user needs and guiding intuitive reasoning but lacks formal mechanisms for navigating technical contradictions or linking user complaints to parameter-level trade-offs. These complementary limitations illustrate why neither approach, when used in isolation, is fully adequate for contemporary socio-technical design challenges.

Recent TRIZ developments exhibit similar patterns. Radical TRIZ (Wang et al., 2024) offers deep functional decomposition but requires extensive data processing and provides limited support for user integration. Cyclical TRIZ (Altun, 2025a) emphasizes iterative refinement without specifying how abstraction transitions should be managed. AI-assisted extensions such as EBD-TRIZ-LLM (Mohammadi and Zeng, 2025) broaden the search space but depend heavily on the quality of the underlying problem model and do not inherently resolve contradictions. COREX responds to these theoretical gaps by combining the structured contradiction modeling of OTSM with the user-grounded clarity of the Six-Box Scheme. This integration provides a balanced framework that maintains analytical depth while remaining cognitively accessible, offering a coherent methodological pathway for complex design problems.

7. Conclusion

This research contributes to the field of systematic innovation by demonstrating how classical contradiction-based logic can be integrated with cognitive innovation models. It supports a hybrid methodology that addresses both technical contradictions and human-centered needs, and it extends the application of meta-contradictions and predictive reasoning in early-stage design.

For R&D and innovation managers, COREX offers a practical roadmap for guiding complex projects. It helps teams to navigate ambiguity, clarify user and system needs, and generate solutions with reduced risk of early-stage failure. Its structured steps and recursive flow also improve cross-functional collaboration, particularly in industries where user–system interactions pose significant design challenges.

This study was conducted as a simulation within a real R&D setting. Although team discussions and feedback resembled authentic development conditions, neither market testing nor long-term adoption fell within the scope of investigation. The evaluation therefore focused on methodological effectiveness rather than engineering performance, leaving real-world validation of the final concepts as an open direction for future research.

Future studies may apply COREX in other industries and with more diverse teams. Integrating digital tools (e.g., AI-enhanced contradiction detection or simulation-based user testing) could further strengthen the methodology. Longitudinal research may also examine how COREX shapes innovation cycles and contributes to product success over time.

Acknowledgments

None.

Funding

None.

Conflict of Interest

The author declares no competing interests.

Author Contributions

This is a single-authored article.

Availability of Data

Data related to the study are available from the corresponding author upon reasonable request.

References

Altshuller, G.S. (1984). Creativity as an exact science: The theory of the solution of inventive problems. *Gordon and Breach Science Publishers*.

Altshuller, G.S. (1996). The innovation algorithm: TRIZ, systematic innovation and technical creativity. *Technical Innovation Center*.

Altun, K, Ozcan Berber, R, Kurt, R, Bektas, E, Turan, S, Korkmaz, V. (2022). Front seat development for autonomous driving: A case of innovative product development. *Journal of the Faculty of Engineering and Architectures of Gazi University*. 37 (3): 1441–1452. https://doi.org/10.17341/gazimmfd.936325

Altun, K. (2023). The synergy of multi-issue negotiation and OTSM-TRIZ: Inventive problem solving in automotive seat design. *ICSI 2023 – The 14th International Conference and Global Competition on Systematic Innovation*, October 13-15, Guangdong, PR China.

Altun, K. (2025a). A Mayan calendar-inspired cyclical TRIZ approach: Enhancing systematic innovation and long-term problem-solving. *International Journal of Systematic Innovation*. 9(3): 8-19. https://doi.org/10.6977/IJoSI.202506_9(3).0002

Altun, K. (2025b). A proof-of-concept for parameter manipulation in TRIZ: Automotive case study. *International Journal of Automotive Science and Technology*. 9(2): 166-173. https://doi.org/10.30939/ijastech.1592053

Altun, K. (2025c). Automotive front-seat design trade-offs: An OTSM-TRIZ-based approach. *International Journal of Automotive Science and Technology*. (accepted for publication).

Borgianni, Y., Frillici, F.S., Rotini, F. (2015). Integration of OTSM-TRIZ and Analytic Hierarchy Process for choosing the right solution. *Procedia Engineering*. 131: 388–400. https://doi.org/10.1016/j.proeng.2015.12.431

Brandenburg, F. (2002). Methodology for planning to technological product innovation. *Aachen: Shaker Verlag*.

Cavallucci, D., Fuhlhaber, S., Riwan, A. (2015). Assisting Decisions in Inventive Design of Complex Engineering Systems. *Procedia Engineering*. 131: 975–983. https://doi.org/10.1016/j.proeng.2015.12.409

Cosgun, M., Altun, K. (2025). Innovative solutions for convolutional neural network performance: A TRIZ-based reverse engineering approach. *International Journal of Systematic Innovation*. 9(3): 1-7. https://dx.doi.org/10.6977/IJoSI.202506_9(3).0001

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 (pp. 55-69). *Cham: Springer Nature Switzerland*.

Elmaraghy, W., Elmaraghy, H., Tomiyama, T., Monostori, L. (2012). Complexity in engineering design and manufacturing. *CIRP Annals - Manufacturing Technology.* 61(2): 793–814. https://doi.org/10.1016/j.cirp.2012.05.001

Eltzer, T., Cavallucci, D., Khomenkho, N., Lutz, P., Caillaud, E. (2006). Inventive Design Applied to Injection Molding, in: ElMaraghy, H.A., ElMaraghy, W.H. (Eds.), *Advances in Design, Springer Series in Advanced Manufacturing. Springer London, London*. 169–182.

Fiorineschi, L., Frillici, F.S., Rissone, P. (2015). A Comparison of Classical TRIZ and OTSM-TRIZ in Dealing with Complex Problems. *Procedia Engineering*. 131: 86–94. https://doi.org/10.1016/j.proeng.2015.12.350

Hansen, M.T., Birkinshaw, J. (2007). The innovation value chain. *Harvard Business Review*. 85 (6): 121-130.

Ilevbare, I.M., Probert, D., Phaal, R. (2013). A review of TRIZ, and its benefits and challenges in practice. *Technovation*. 33: 30–37. https://doi.org/10.1016/j.technovation.2012.11.003

Khomenko, N. and Kucharavy, D. (2002). OTSM-TRIZ problem solving process: solutions and their classification. *Proceedings of TRIZ Future Conference*. 6–8 November, Strasbourg, France.

Khomenko, N., Ashtiani, M. (2007). Classical TRIZ and OTSM as a scientific theoretical background for non-typical problem-solving instruments. *Frankfurt: ETRIA Future*.

Khomenko, N., De Guio, R., Cavallucci, D. (2009). Enhancing ECN's abilities to address inventive strategies using OTSM-TRIZ. *International Journal of Collaborative Engineering*. 1(1-2): 98-113. https://dx.doi.org/10.1504/IJCE.2009.027441

Kim, D. (2021). Analysis of automotive seating systems in automated driving vehicles according to the changes of the interior environment. *Master thesis of public administration.* Graduate School of Engineering Practice, Seoul National University, Seoul, South Korea.

Kruger, L.L.S.J., Pretorius, J.H.C., Erasmus, L.D. (2019). Towards a Comprehensive Systematic Innovation Model: A Literature review. *SAIEE Afr. Res. J.* 110: 39–46. https://doi.org/10.23919/SAIEE.2019.8643149

Mann, D. (2007). Hands-on systematic innovation. *Devon: IFR Press*.

Mann, D. (2023). TRIZ and chaos: First principles for emergency first responders, in: Cavallucci, D., Livotov, P., Brad, S. (Eds.), Towards AI-aided invention and innovation. *IFIP Advances in Information and Communication Technology*. Springer Nature Switzerland, Cham, pp. 453–464.

Mohammadi, A., Zeng, Y. (2025). Enhancing TRIZ through environment-based design methodology supported by a large language model. *Artificial Intelligence for Engineering Design, Analysis, and Manufacturing*. 39: e12, 1-18. https://doi.org/10.1017/S0890060425000083

Nakagawa, T. (2011). Education and training of creative problem-solving thinking with TRIZ/USIT. *Procedia Eng*. 9: 582–595. https://doi.org/10.1016/j.proeng.2011.03.144

Nakagawa, T. (2016a). CrePS (General Methodology of Creative Problem Solving) beyond TRIZ: what, why, and how? *Proceedings of TRIZCON 2016*. New Orleans, LA, March 2–5, 2016, TRIZ HP Japan.

Nakagawa, T. (2016b). USIT: a concise process for creative problem solving based on the paradigm of 'Six-Box Scheme' – USIT Manual and USIT Case Studies. *Proceedings of ETRIA TFC2015 (TRIZ Future Conference)*. Berlin, Germany, October 26–29, 2015, TRIZ HP Japan.

Nakagawa, T. (2018). TRIZ/CrePS approach to the social problems of poverty: 'liberty vs. love' is found the principal contradiction of the human culture. *In Advances and Impacts of the Theory of Inventive Problem Solving: The TRIZ Methodology, Tools and Case Studies (pp. 179-188).* Cham: Springer International Publishing.

Roper, S., Du, J, Love, J. (2008). Modeling the innovation value chain. *Research Policy*. 37(6-7): 962-977. https://doi.org/10.1016/j.respol.2008.04.005

Sheu, D. D., Lee, H. K. (2011). A proposed process for systematic innovation. *International Journal of Production Research*. 49(3): 847-868. https://doi.org/10.1080/00207540903280549

Sun, J., Li, H.-Y., Du, Y.J., Song, Z., Tan, R. (2020). A systematic innovation process oriented to inter-discipline, in: Cavallucci, D., Brad, S., Livotov, P. (Eds.), Systematic complex problem solving in the age of digitalization and open innovation. *IFIP Advances in Information and Communication Technology.* Springer International Publishing, Cham, pp. 257–267.

Wang, F., Tan, R., Wang, K., Cen, S., Peng, Q. (2024). Innovative product design based on radical problem solving. *Computers & Industrial Engineering*. 189: 109941. https://doi.org/10.1016/j.cie.2024.109941

Zhan, Y., Tan, K.H., Ji, G., Chung, L., Tseng, M. (2017). A big data framework for facilitating product innovation processes. *Business Process Management Journal*. 23: 518–536. https://doi.org/10.1108/BPMJ-11-2015-0157

AUTHOR BIOGRAPHIES

Koray Altun is an assistant professor of industrial engineering at TU Bursa and a research collaborator at Glorad. He is also the founder and manager of a startup focused on digital innovation and software consultancy. He holds a Ph.D. and a B.Sc. degree in industrial engineering from Gaziantep University and Erciyes University, respectively. He has published papers in well-respected journals and has led consulting engagements in reputable companies with a focus on research and development (R&D), technology, and innovation management. His recent research interests include the innovation excellence model, systematic innovation, technology and innovation management, and R&D management.

DOI: 10.6977/IJoSI.202512_9(6).0002
W. Shin, Y. Choi, etc./Int. J. Systematic Innovation, 9(6), 17-26 (2025)

Enhancing problem-solving and data protection through the integration of function-oriented search and ChatGPT

Won-Shik Shin[1], Youngjoon Choi[2], Yong-Won Song[3]*

[1]Department of Tourism, Business and Economic Research Institute, Jeju National University, Jeju, Jeju, South Korea

[2]Department of International Office Administration, Ewha Womans University, Seoul, South Korea

[3]Department of Semiconductor Engineering, Tech University of Korea Siheung, Gyeonggi, South Korea

*Corresponding author E-mail: ywsong@tukorea.ac.kr

(Received 19 June 2024; Final version received 09 December 2025; Accepted 10 December 2025)

Abstract

As a large language model, ChatGPT's ability to learn from big data and respond to diverse user queries makes it a powerful tool for research and development. Despite the potential benefits of using ChatGPT, there are risks concerning users' data protection. To address this issue, this study proposes utilizing Function-Oriented Search (FOS), a methodology based on Theory of Inventive Problem Solving (TRIZ). FOS provides an innovative approach to problem-solving by functionally defining a problem and generating solutions from areas where the function can be optimally performed. Thus, this study argues that applying FOS when using ChatGPT can ensure accurate results while mitigating the exposure of sensitive information. Although implementing FOS requires specialized training and sufficient hands-on experience to identify and conceptualize problem focus areas, ChatGPT can serve as an efficient tool for developers adopting this methodology. For both experts and novices in FOS, ChatGPT enables users to conduct efficient and comprehensive problem explorations and devise solutions. By demonstrating the application of FOS in practical cases, the study's findings support the potential benefits of ChatGPT as a dynamic collaborator in problem-solving. The findings also indicate that FOS can guide the use of ChatGPT to generate suitable solutions while maintaining the protection of personal or corporate information. Overall, this study contributes to the emerging field of artificial intelligence by illustrating the possible synergy between TRIZ-based FOS and ChatGPT, a large language model.

Keywords: ChatGPT, Data Protection, Function-Oriented Search, Large Language Model, Prompt Engineering, Theory of Inventive Problem Solving, TRIZ-Informed Prompt Engineering

1. Introduction

The field of artificial intelligence (AI) has developed rapidly in recent years. In particular, the Generative Pre-trained Transformer (ChatGPT) has garnered significant attention due to its large-scale language model with a conversational interface. ChatGPT's ability to generate human-like text not only provides users with personalized interactions but also has numerous applications in areas, such as customer service, education, and healthcare (Aljanabi et al., 2023; Bang et al., 2023). It can also serve as a versatile tool for complex problem-solving, research, and development (Sinha et al., 2023; Tafferner et al., 2023).

Despite these benefits, ChatGPT has certain limitations, including privacy and security risks, bias, misuse, and the potential for misinformation (Borji, 2023; Dwivedi et al., 2023; Oviedo-Trespalacios et al., 2023; Wach et al., 2023). In particular, this study focuses on the risk of unintentional exposure of sensitive data, which can range from privacy breaches to corporate confidentiality leaks and intellectual property theft. Therefore, there is an urgent need to explore methodologies that leverage ChatGPT's capabilities while protecting personal or sensitive information.

The main purpose of this study is to apply the Theory of Inventive Problem Solving (TRIZ)'s Function-Oriented Search (FOS) as a countermeasure to these vulnerabilities. Specifically, this study has two main objectives. The first objective is to investigate the potential of implementing FOS, a methodology within TRIZ's systematic problem-solving framework. FOS emphasizes a function-oriented perspective, defining problems functionally and generating solutions across diverse areas. By applying FOS to data protection challenges, this study aims to elucidate whether it can ensure accurate and relevant responses from ChatGPT without compromising sensitive data. The second objective is to examine how the integration of FOS and ChatGPT enables efficient and comprehensive problem exploration and solution development. This approach aims to improve users' problem-solving abilities, both for FOS experts and for developers unfamiliar with FOS, demonstrating a synergistic collaboration between ChatGPT and TRIZ's FOS.

2. Literature Review

2.1. ChatGPT and its Limitations

Developed by OpenAI, ChatGPT is a generative pre-trained transformer widely recognized for its advanced machine learning algorithms, capabilities, and interactive applications. Its remarkable ability to understand and generate human-like text from vast amounts of data allows it to provide meaningful responses to diverse prompts, making it applicable in fields, such as coding, nursing, tourism, writing, and publishing (Aljanabi et al., 2023; Bang et al., 2023; Dwivedi et al., 2023).

However, ChatGPT has limitations. One major limitation is that it generates responses based on patterns learned from training data. Given that ChatGPT cannot yet fully capture the nuances of human language and social communication, its understanding of context is limited, potentially resulting in inappropriate or meaningless responses in complex or ambiguous situations (Aljanabi et al., 2023). Data protection is also a critical concern. Due to the way ChatGPT learns from accumulated data at OpenAI, there is a risk of inadvertently compromising privacy or exposing sensitive information (Dwivedi et al., 2023). This vulnerability presents a significant obstacle to wider application, which this study addresses using TRIZ's FOS methodology.

2.2. TRIZ and FOS

The TRIZ, developed by Genrich Altshuller in the 1940s, is a systematic methodology for understanding and solving complex problems (Altshuller, 1984; 1999). TRIZ transforms a specific problem into a generalized problem model through abstraction and formalization, and identifies corresponding general solution models. It then provides systematic methods, such as solving specific problems through analogy and interpreting the general solution model (Cameron, 2010; Haines-Gadd, 2016; Ilevbare et al., 2013; Orloff, 2017). One of the core techniques in modern TRIZ is FOS.

FOS translates the key problem into a functional language, formulating generalized functions to identify technical solutions from other areas and transferring existing technologies to address the initial problem (Litvin, 2005). As shown in Fig. 1, FOS is based on TRIZ, and its problem-solving process is analogous to that of TRIZ (Choi et al., 2012).

The FOS procedure, based on TRIZ's general problem-solving process, is as follows (Wang et al., 2023; Zhang et al., 2023): (i) Identify the target problem; (ii) abstract the target problem into a conceptual problem and generalize it into a functional language; (iii) search the existing solutions using a function-based technology database (leading area); and (iv) apply the existing solutions to the target problem.

By defining and abstracting the essence of a problem as a function that needs to be performed, FOS expands the solution search to areas where the identified function can be optimally executed. FOS deviates from conventional problem-solving methodologies, which largely depend on pre-existing solutions within the immediate problem domain. While many prevailing strategies draw from cases addressing analogous challenges, FOS characterizes the problem in functional terms. This broadened search can lead to innovative, cross-disciplinary solutions that might not have been recognized if the problem remained contextualized within its original field, seeking resolutions in alternative fields where the function is optimally performed.

Several studies have attempted to implement FOS using data mining techniques (Wang et al., 2023; Zhang et al., 2023). These approaches are noteworthy because they provide algorithmic methods for FOS, but they can present barriers for users who are unfamiliar with the technology in this area.

Theory of Inventive Problem Solving methodologies, including FOS, requires extensive training and practice. In this study, we suggest an approach to implementing FOS by leveraging ChatGPT, regardless of the user's level of expertise.

3. Methodology

FOS solves problems by abstracting them using functional terms instead of specific terms that might reveal product brands or problem targets. This strategy

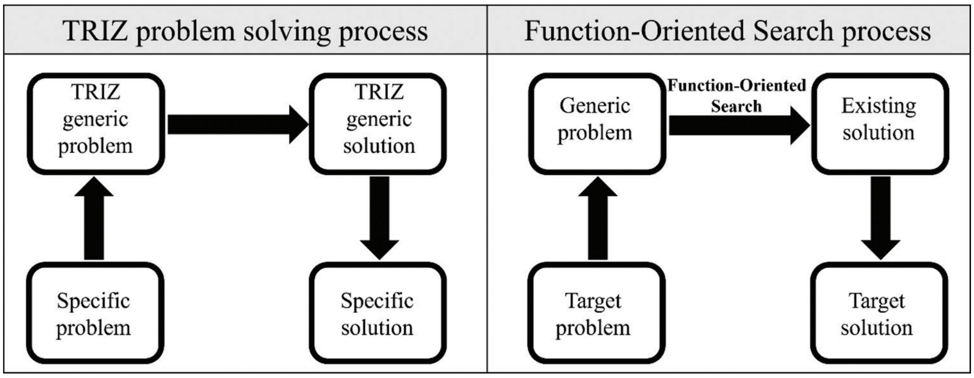


Fig. 1. Comparison of TRIZ's general problem-solving process and FOS's process (reprinted with permission from Choi et al., [2012]; Copyright © 2012 Elsevier Ltd.)
Abbreviations: FOS: Function-Oriented Search; TRIZ: Theory of Inventive Problem Solving

not only broadens the scope of problem-solving but also protects sensitive details from being exposed. In addition, FOS's tendency to look beyond the immediate problem context, combined with ChatGPT's ability to access a wide range of knowledge, creates a synergy that improves both problem-solving efficiency and data security.

For this study, we designed an FOS implementation using ChatGPT. We examined the applicability of ChatGPT's solutions from the perspective of TRIZ's FOS and validated the approach by comparing it with actual development cases.

In large language models, prompt engineering refers to the systematic formulation and refinement of input prompts to guide the model's behavior toward outputs that are useful, reliable, and contextually appropriate. This is achieved by designing structured instructions and interactions that clearly articulate goals, constraints, roles, and reasoning styles.

In this study, prompt formulation was conceptualized and implemented as TRIZ-informed prompt engineering. Specifically, FOS was employed to represent the target problem in functional terms. These resulting functional descriptions were subsequently used to structure the prompts provided to ChatGPT, guiding its problem-solving process while minimizing the risk of disclosing sensitive information.

The procedure followed in this study is as follows:

(i) Case description: Outline the target engineering problem, its constraints, and the core function to be achieved
(ii) Prompt design and implementation: Develop ChatGPT prompts from a FOS perspective, using TRIZ-informed functional abstraction and interaction patterns to translate the problem into function-oriented prompts
(iii) Evaluation: Assess the results generated by ChatGPT and compare them to real-world problem solutions to demonstrate the effectiveness of FOS implementation.

All procedures were performed using ChatGPT Plus (GPT-4, OpenAI, US).

4. Case Study

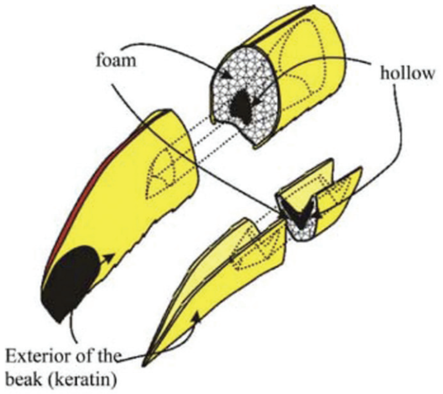
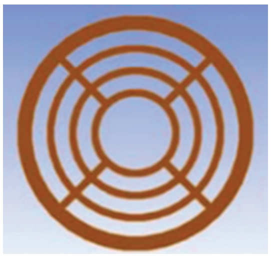
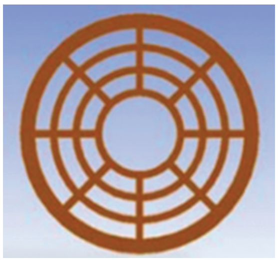
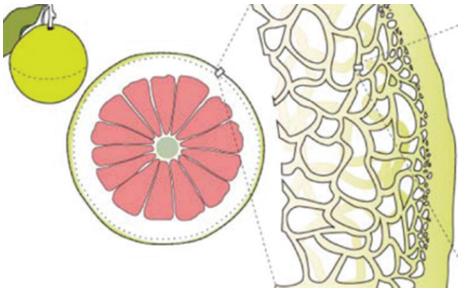
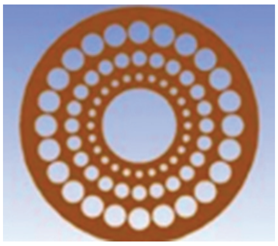
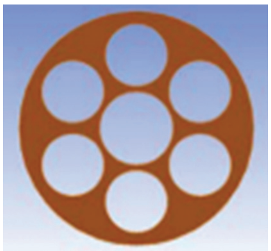
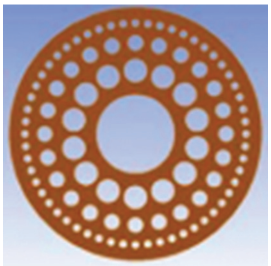
4.1. Description of Case Study

As a case study, we consider research on side-door impact beams by Shaharuzaman et al. (2020). The side-door impact beam was first introduced by General Motors in the late 1960s to safeguard the passenger compartment from external collisions. Side-door impact beams, which improve the strength, stiffness, and energy absorption of a vehicle door in a side impact, come in a variety of designs, with ribs for the beam being a common feature, and a circular cross-section being the most widely used design type.

With the growing emphasis on sustainable design and environmentally friendly products, automakers are increasingly adopting natural fiber composites (NFC), which are characterized by specific strength, rigidity, recyclability, and their appeal as natural materials, in place of synthetic fiber composites. The use of NFC is a significant trend in the automotive industry due to its lightweight properties, which enhance fuel efficiency and offer eco-friendly options for recycling and disposal. As natural fibers exhibit diverse mechanical and material properties, thorough analysis during material selection is essential in the early stages of product design and development.

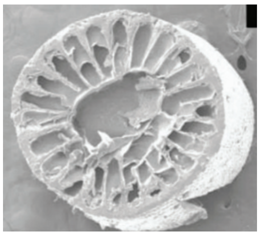
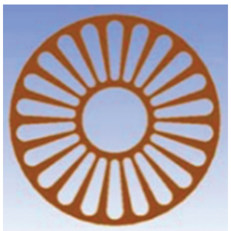
As shown in Table 1, Shaharuzaman et al. conducted FOS and biometric methods using ideas obtained from asknature.org and derived eight design proposals through finite element analysis. Afterward, one of the pomelo peel models (B-03) and one of the hedgehog spine models (C-02) were selected as the optimal choices for the side-door impact beam using VIseKriterijumska Optimizacija I Kompromisno Resenje.

Table 1. Key ideas proposed by Shaharuzaman et al. (2020)

Biomimetic strategies	**Design**	**Name**
 Toucan beak		A-01
		A-02
		A-03
 Pomelo peel		B-01
		B-02
		B-03

(Cont'd...)

Table 1. *(Continued)*

Biomimetic strategies	Design	Name
 Hedgehog spine		C-01
		C-02

4.2. Design and Implementation of ChatGPT Prompt Engineering from a FOS Perspective

To design an NFC side-impact beam from an FOS perspective, this study used the persona pattern and flipped interaction pattern during prompt engineering (White et al., 2023). For the persona pattern, we assigned ChatGPT the role of an experienced TRIZ (FOS) expert and incorporated biomimetics following the research method of Shaharuzaman et al. (2020). Afterward, we explained the usual FOS procedure and allowed ChatGPT to proceed accordingly. This study also applied the flipped-interaction pattern, which gives ChatGPT the right to ask questions so that it can provide more accurate answers (Fig. 2).

We asked ChatGPT to design a "circular lateral force absorber–distributor made of natural fiber composite" using FOS biomimetics (Fig. 3).

ChatGPT generated a response to the request, as shown in Fig. 4. Due to space constraints, only a partial view of the content is presented here; the full content can be accessed at the following link: https://chat.openai.com/share/bfda4e5d-4ee7-4046-8069-db2a3acdee9d.

The problem was expressed in the functional language "circular lateral force absorber–distributor made of natural fiber composite" for a side-door impact beam made of NFCs. The conversion of a specific description into a functional language was chosen after discussion with the research participants. Of course, it is also possible to utilize ChatGPT, in which case we took measures to protect against the unintentional disclosure of information, such as using a third-party account. When this description was entered into the prompt, ChatGPT provided several answers. To better emphasize functionality, we further simplified the functional language from "circular lateral force absorber–distributor made of natural fiber composite" to "absorb and distribute forces."

As presented in Fig. 5, ChatGPT suggested several biomimetic solutions: Bamboo, spider silk, the woodpecker's beak and head, and shell structures.

If another response is desired, the user can simply click the "Regenerate" button. Fig. 6 shows the second response generated after clicking the button. In this response, ChatGPT attempted to provide a more accurate answer by asking clarifying questions. Each time the user clicks "Regenerate," ChatGPT can produce different, yet related, responses that offer the user additional knowledge and insights.

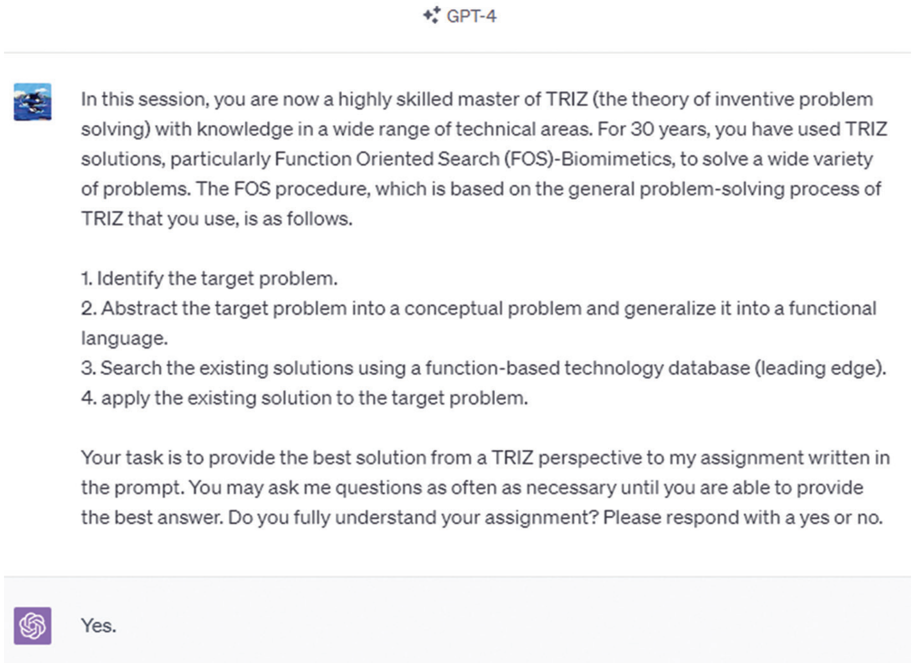
4.3. Evaluation of ChatGPT's Response and Comparison with the Solutions of Shaharuzaman et al.

This study also compared the answers obtained by running FOS on ChatGPT with the study of Shaharuzaman et al. (2020). Identifying the similarities and differences between the two studies serves as the primary means of judging ChatGPT's validity and its potential to provide insights for problem-solving.

4.3.1. Relevance to the research of Shaharuzaman et al. (2020)

The findings indicated that ChatGPT's response contained ideas similar to those presented by Shaharuzaman et al. (2020). Both results referenced designs inspired by the same woodpecker order, including the toucan and the woodpecker. The idea of using pomelo shells presented in their study can also be inferred from the shell-based structures proposed by ChatGPT.

DOI: 10.6977/IJoSI.202512_9(6).0002
W. Shin, Y. Choi, etc./Int. J. Systematic Innovation, 9(6), 17-26 (2025)


GPT-4

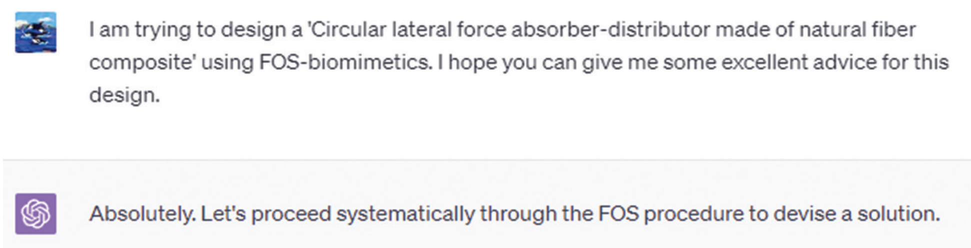
In this session, you are now a highly skilled master of TRIZ (the theory of inventive problem solving) with knowledge in a wide range of technical areas. For 30 years, you have used TRIZ solutions, particularly Function Oriented Search (FOS)-Biomimetics, to solve a wide variety of problems. The FOS procedure, which is based on the general problem-solving process of TRIZ that you use, is as follows.

1. Identify the target problem.
2. Abstract the target problem into a conceptual problem and generalize it into a functional language.
3. Search the existing solutions using a function-based technology database (leading edge).
4. apply the existing solution to the target problem.

Your task is to provide the best solution from a TRIZ perspective to my assignment written in the prompt. You may ask me questions as often as necessary until you are able to provide the best answer. Do you fully understand your assignment? Please respond with a yes or no.

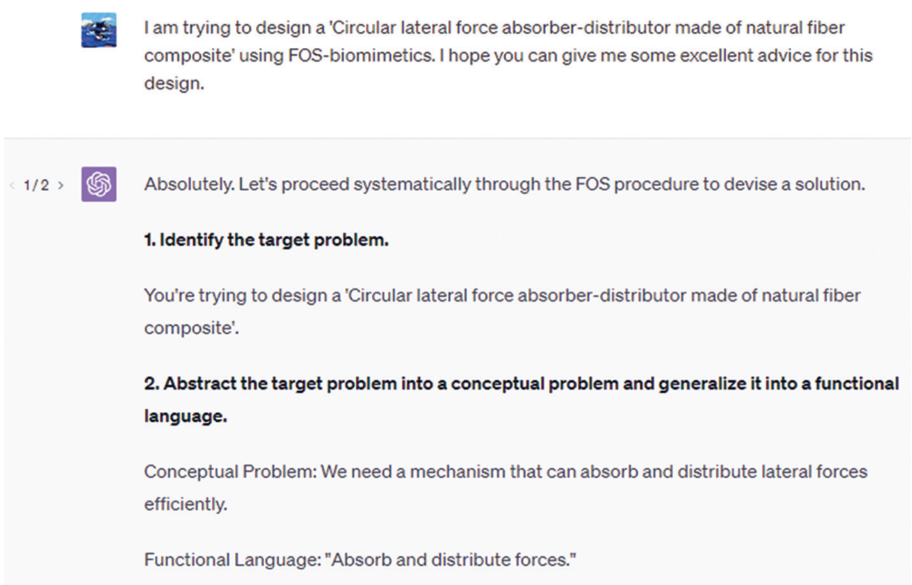
Yes.

Fig. 2. Prompt engineering of the Function-Oriented Search-biomimetics perspective
Abbreviation: TRIZ: Theory of Inventive Problem Solving


I am trying to design a 'Circular lateral force absorber-distributor made of natural fiber composite' using FOS-biomimetics. I hope you can give me some excellent advice for this design.

Absolutely. Let's proceed systematically through the FOS procedure to devise a solution.

Fig. 3. Prompt provided to ChatGPT
Abbreviation: FOS: Function-Oriented Search


I am trying to design a 'Circular lateral force absorber-distributor made of natural fiber composite' using FOS-biomimetics. I hope you can give me some excellent advice for this design.

1/2

Absolutely. Let's proceed systematically through the FOS procedure to devise a solution.

1. Identify the target problem.

You're trying to design a 'Circular lateral force absorber-distributor made of natural fiber composite'.

2. Abstract the target problem into a conceptual problem and generalize it into a functional language.

Conceptual Problem: We need a mechanism that can absorb and distribute lateral forces efficiently.

Functional Language: "Absorb and distribute forces."

Fig. 4. Response from ChatGPT
Abbreviation: FOS: Function-Oriented Search

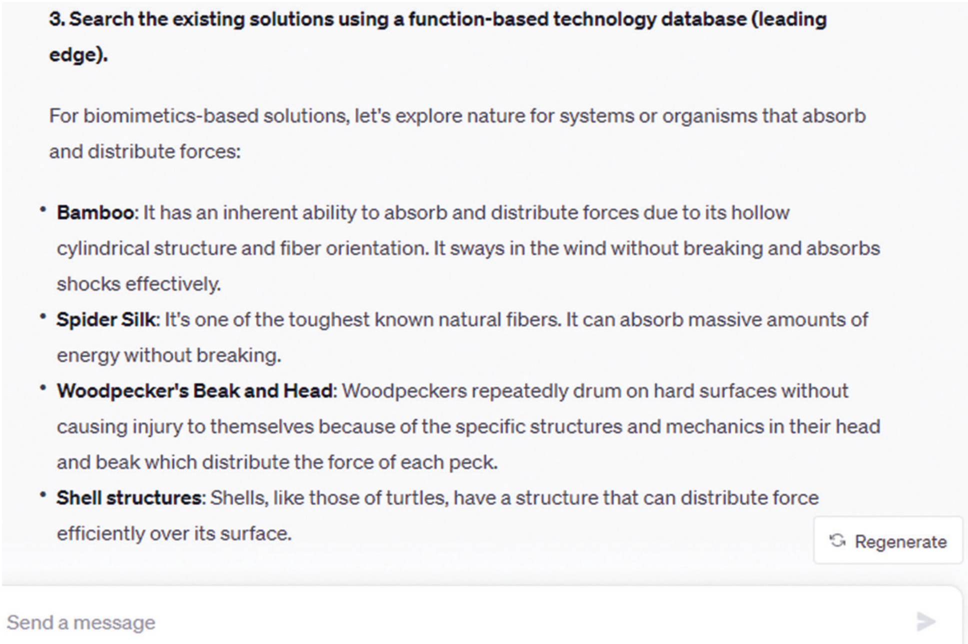


Fig. 5. Biomimetic suggestions generated by ChatGPT

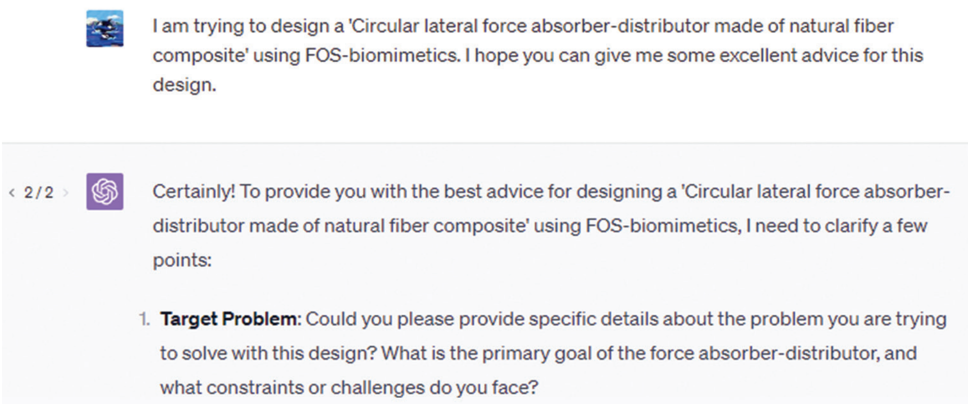


Fig. 6. Second response from ChatGPT

4.3.2. Biomimetics inspiration

The solutions proposed by ChatGPT were based on biomimetics, derived from the prompted input. ChatGPT did not simply identify biological analogues but also provided explanatory information that could inspire design ideas—for example, bamboo's concentric layer structure, the strength properties of spider webs, woodpecker-inspired cushioning mechanisms, and segmented shell structures.

4.3.3. Dynamic interaction design

This study utilized the flipped-interaction pattern to allow ChatGPT to ask users additional questions and respond based on their specific intentions. This created a dynamic interaction, rather than a simple one-time input/output exchange, supporting collaborative engagement between the user and ChatGPT.

When we clicked the "Regenerate" button, ChatGPT generated more specific questions about what user goals, constraints, and requirements. This enabled ChatGPT to provide more contextually relevant and personalized advice. Such dynamic interaction supports a user-centered approach to generating results.

4.3.4. Evaluation

The results of Shaharuzaman et al. (2020) were compared with ChatGPT's outputs to investigate the feasibility and effectiveness of applying FOS through ChatGPT. The similar outcomes suggest that ChatGPT can provide valid FOS-based results. Furthermore, ChatGPT provided various additional suggestions not found in Shaharuzaman et al.'s study (2020) and proposed customized solutions through interactive dialogue. This demonstrates both the effectiveness and utility of implementing FOS via ChatGPT. In addition, although traditional FOS requires substantial learning and practical experience to use effectively, this study found that even inexperienced FOS users can apply it easily through the simple prompt-engineering strategies demonstrated here.

5. Conclusion

In recent years, collaboration between humans and AI has garnered significant interest. Contrary to earlier apprehensions, AI is increasingly recognized not as a detractor from human creativity but as an enhancer of human cognitive capabilities, enabling new scientific and artistic pathways (Colton et al., 2009; Wingström et al., 2022). Cropley et al. (2022) underscored the potential integration of human and AI creativity and advocated for a deeper understanding and development of the synergy at their intersection.

The findings of this study contribute to this evolving field by examining the feasibility and effectiveness of implementing FOS within ChatGPT. The specific contributions of this study include:

i. Effective implementation of FOS through ChatGPT
As demonstrated in the case study on side-door impact beam design, this study showed empirically that FOS can be effectively implemented in ChatGPT via TRIZ-informed prompt engineering. Among the outputs generated by ChatGPT, woodpecker-inspired structures and shell structures aligned with the results of Shaharuzaman et al. (2020), supporting the validity of this study. In addition, ChatGPT continuously generated user-centered solutions through interactive dialogue, providing users with efficient and diverse design concepts.

ii. Mitigating data security vulnerabilities
Using functional language in FOS abstracts sensitive information and minimizes the exposure of important details. This enables users to obtain ideas and insights from ChatGPT while reducing the risk of data disclosure.

iii. Broadening FOS accessibility
Effective application of FOS traditionally requires significant learning and experience; however, ChatGPT makes it easier for a broader range of users to apply FOS. This allows individuals with varying levels of expertise to analyze complex problems or create innovative solutions by integrating FOS with ChatGPT.

Nevertheless, this study is not without its limitations. A key limitation is its reliance on a single in-depth case study. Therefore, to move beyond the constraints of a single case and evaluate the generalizability of the proposed approach, future research should undertake a broader range of case studies. Furthermore, although the synergy between FOS and ChatGPT appears promising, its implementation requires further refinement for wider applicability. Future progress will benefit from developing user-friendly tools, such as plugins, apps, or interfaces to facilitate the use of FOS through ChatGPT.

Since its release in November 2022, ChatGPT has received an overwhelming response, and ongoing efforts continue to explore its potential across various sectors. This study represents one step in that direction and aims to serve as a foundation for future research.

Conflict of Interest

The authors declare that they have no competing interests.

References

Aljanabi, M., Ghazi, M., Ali, A., & Abed, S.A. (2023). ChatGPT: Open possibilities. *Iraqi Journal For Computer Science and Mathematics*, 4(1), 62–64. https://doi.org/10.52866/20ijcsm.2023.01.01.0018

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

Altshuller, G.S. (1999). *The Innovation Algorithm: TRIZ, Systematic Innovation and Technical Creativity*. Technical Innovation Center, Inc., Somayampalayam.

Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Holy, B.W., et al. (2023). *A Multitask, Multilingual, Multimodal Evaluation of Chatgpt on Reasoning, Hallucination, and Interactivity*; [arXiv Preprint].

Borji, A. (2023). *A Categorical Archive of Chatgpt Failures*. [arXiv Preprint].

Cameron, G. (2010). *Trizics*. Available from: https://www.com/trizics.com

Choi, S., Kang, D., Lim, J., & Kim, K. (2012). A fact-oriented ontological approach to SAO-based function modeling of patents for implementing function-based technology database. *Expert Systems with Application*, 39(10), 9129–9140. https://doi.org/10.1016/j.eswa.2012.02.041

Colton, S., Mántaras, R., De., & Stock, O. (2009). Computational creativity: Coming of age. *AI Magazine*, 30(3), 11–14. https://doi.org/10.1609/aimag.v30i3.2257

Cropley, D.H., Medeiros, K.E., & Damadzic, A. (2022). The intersection of human and artificial creativity. *In: Creative Provocations: Speculations on the Future of Creativity, Technology and Learning.* Springer International Publishing, Cham, p19–34.

Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kumar Kar, A., et al. (2023). “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Haines-Gadd, L. (2016). *TRIZ for Dummies*. John Wiley and Sons Inc., United States.

Ilevbare, I.M., Probert, D., & Phaal, R. (2013). A review of TRIZ, and its benefits and challenges in practice. *Technovation*, 33(2-3), 30–37. https://doi.org/10.1016/j.technovation.2012.11.003

Litvin, S. (2005). New TRIZ-based tool-function-oriented search (FOS). *The TRIZ Journal*.

Orloff, M.A. (2017). *ABC-TRIZ*. Springer International Publishing, Berlin.

Oviedo-Trespalacios, O., Peden, A., Science, T.C., Costantini, A., Haghani, M., Rod, JE., et al. (2023). The risks of using chatgpt to obtain common safety-related information and advice. *Safety Science*, 167, 106244. https://doi.org/10.1016/j.ssci.2023.106244

Shaharuzaman, M.A., Sapuan, S.M., Mansor, M.R., & Zuhri, M.Y.M. (2020). Conceptual design of natural fiber composites as a side-door impact beam using hybrid approach. *Journal of Renewable Materials*, 8(5), 549–563. https://doi.org/10.32604/jrm.2020.08769

Sinha, R., Roy, A., Kumar, N., & Mondal, H. (2023). Applicability of ChatGPT in assisting to solve higher order problems in pathology. *Cureus*, 15(2), e35237. https://doi.org/10.7759/cureus.35237

Tafferner, Z., Illés, B., Krammer, O., & Géczy, A. (2023). Can chatGPT help in electronics research and development? A case study with applied sensors. *Sensors (Basel)*, 23(10), 4879. https://doi.org/10.3390/s23104879

Wach, K., Duong, C.D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., et al. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2), 7–24. https://doi.org/10.15678/EBER.2023.110201

Wang, J., Zhang, Z., Feng, L., Lin, K.Y., & Liu, P. (2023). Development of technology opportunity analysis based on technology landscape by extending technology elements with BERT and TRIZ. *Technological Forecasting and Social Change*, 191(122481), 122481. https://doi.org/10.1016/j.techfore.2023.122481

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., et al. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with Chatgpt. In: PLoP ‘23: Proceedings of the 30th Conference on Pattern Languages of Program*.

Wingström, R., Hautala, J., & Lundman, R. (2022). Redefining creativity in the Era of AI perspectives of computer scientists and new media artists. *Creativity Research Journal*, 36, 1–17. https://doi.org/10.1080/10400419.2022.2107850

Zhang, D., Wu, X., Liu, P., Qin, H., & Sciences, W.Z. (2023). Identification of product innovation path incorporating the FOS and BERTopic model from the perspective of invalid patents. *Applied Sciences*, 13(13), 7987. https://doi.org/10.3390/app13137987

AUTHOR BIOGRAPHIES

Won-Shik Shin is a Research Professor at the Tourism, Business, and Economic Research Institute of Jeju National University. He earned his Ph.D. in the Faculty of Data Science for Sustainable Growth, majoring in Management Information Systems at Jeju National University. His research integrates artificial intelligence, data science, and TRIZ to investigate analytical and computational mechanisms that support structured innovation and creative problem-solving in complex environments. His current work focuses on advancing problem-solving capability through the convergence of TRIZ and generative AI. He is leading the project "Development of a Data–AI Convergence Model for Creative Problem-Solving Mechanisms," which explores how data-driven intelligence and methodological frameworks can be combined to enhance creativity, scalability, and practical applicability in modern innovation systems.

Youngjoon Choi is an Associate Professor in the Department of International Office Administration at the College of Science and Industry Convergence at Ewha Womans University since September 2021. From August 2015 to July 2021, he worked as an Assistant Professor in the School of Hotel and Tourism Management at the Hong Kong Polytechnic University. He received his BS in Tourism Management from Kyung Hee University. He then obtained his MSc and PhD in Recreation, Park & Tourism Management from The Pennsylvania State University, USA, in 2011 and 2014, respectively. His main research interests include MICE (meeting, incentive, convention, and exhibition) management, event tourism, technology innovation in hospitality, and destination marketing.

Yong Won Song is currently working as a professor at the Tech University of Korea, Republic of Korea. He received his M.S. degree in Materials Science and Engineering from Korea University, Korea, in 1988, and Ph.D. degree in Physics from Moscow State University, Russia, in 1995. He has published several books and more than 40 articles on semiconductor technology and the Theory of Inventive Problem Solving (TRIZ). His research interests include TRIZ, future prediction, and advanced coating technology.

DOI: 10.6977/IJoSI.202512_9(6).0003
S. Upadhyay, H. K. Soni/Int. J. Systematic Innovation, 9(6), 27-43 (2025)

Automatic text summarization framework for multi-text and multilingual documents using an ensemble of HIN-MELM-AE and improved DePori model

Sunil Upadhyay*, Hemant Kumar Soni

Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Madhya Pradesh, Gwalior, Madhya Pradesh, India

*Corresponding author E-mail: supadhyay@gwa.amity.edu

(Received 28 November 2024; Final version received 10 November 2025; Accepted 02 December 2025)

Abstract

Automatic text summarization (ATS) has gained increasing significance in recent years due to the rapid growth of textual data across digital platforms. The main objective of ATS is to generate a concise, informative summary from a lengthy document. Multi-document and multilingual summarization has been largely underexplored in previous research. This study presents an improved ensemble learning-based ATS system with slang filtering, using the Hyperfan-IN multilayer extreme learning machine-based autoencoder (HIN-MELM-AE) and the improved Dehghani poor-and-rich optimization algorithm (DePori). The original text undergoes comprehensive preprocessing, after which slang is detected and removed using DePori. Subsequently, the clean text is processed through info-squared C-means clustering, latent Dirichlet allocation-based topic modeling, term frequency–inverse document frequency weighting, and frequent-term extraction. Next, part-of-speech (POS) tagging is performed using a sememe similarity-induced hidden Markov model, and key entities are extracted from the transformed and POS–tagged data. Distilled bidirectional encoder representations from transformers (DBERT) are used to convert these entities into vectors. The final summary is generated through a combination of HIN-MELM-AE, stack autoencoder, variational autoencoder, and DBERT models, followed by cosine similarity calculation, voting-based fusion, re-ranking, and selection of the optimal sentences. Experimental results indicate that the proposed framework achieves superior performance 97.92% of the time, outperforming existing ATS methods.

Keywords: Hyperfan-IN Multilayer Extreme Learning Machine Auto Encoder, Info-Squared Fuzzy C-Means Clustering, Latent Dirichlet Allocation, Parts of Speech, Sentence Bidirectional Encoder Representations from Transformers, Sememe Similarity-Induced Hidden Markov Model, Term Frequency–Inverse Document Frequency, Variational Auto Encoder

1. Introduction

The development of the internet and big data has led to an enormous amount of textual content, which grows exponentially every day. Users spend a considerable amount of time searching for information and may read irrelevant portions of the content (El-Kassas et al., 2021). Therefore, summarizing text resources has become essential and increasingly important. However, manual text summarization is time-consuming and requires significant effort and cost (Yadav et al., 2022). To address this, automatic text summarization (ATS) can save users time and effort by generating summaries automatically. Typically, ATS creates summaries composed of significant sentences and relevant information from the original document (Widyassari et al., 2022). The summaries generated by ATS consider aspects such as readability, information diversity, and sentence ordering (Wahab et al., 2024).

The applications of text summarization include legal document abstraction and clinical text summarization (Zhang et al., 2020).

In general, ATS consists of two main methods: extractive and abstractive. The extractive method identifies the most important parts of the text based on scoring criteria and then combines them to produce a summary. In contrast, the abstractive method is more complex; it involves paraphrasing the text to generate a concise version using words and sentences that differ from the original text document (Syed et al., 2021).

In recent years, many approaches have been developed to perform ATS for single- or multi-document summarization. For a single document, ATS produces a concise summary of the document (Payak et al., 2020). In multi-document summarization, users can quickly gain comprehensive knowledge of a topic by reviewing multiple documents (Awasthi et al., 2021). In existing studies, K-means clustering and Gensim Word2Vec have been used for ATS; however, these approaches require high memory (Haider et al., 2020). Moreover, certain models employ deep reinforcement learning to enhance ATS performance; however, they encounter issues with computational complexity (Alomari et al., 2022).

Similarly, the term frequency–inverse document frequency (TF-IDF) algorithm has been used in several existing studies for ATS; however, it is not suitable for abstractive summarization (Manjari et al., 2020). Several researchers also employed a combined latent semantic analysis (LSA) and bidirectional encoder representations from transformers (BERT) model for ATS—they reported that the model is highly sensitive to term variability (Gupta & Patel, 2021). In addition, sequence-to-sequence Seq2Seq with a recurrent neural network (RNN) has been used to perform ATS; however, these models are affected by vanishing or exploding gradient problems (Prasad et al., 2020). In addition, several studies employed latent Dirichlet allocation (LDA) topic modeling and soft-cosine similarity to summarize text documents effectively; however, these approaches consider limited linguistic features (Jain & Rastogi, 2020; Onah et al., 2022). Moreover, relatively few studies have addressed ATS for multi-document and multilingual summaries. To address these gaps, this study presents an improved ensemble learning-based ATS with slang filtering, using the Hyperfan-IN multilayer extreme learning machine-based autoencoder (HIN-MELM-AE) and the Dehghani poor-and-rich optimization algorithm (DePori) techniques.

Conventional methods have rarely focused on ATS for multi-document and multilingual summaries. Therefore, the proposed model explicitly targets ATS in this multi-document and multilingual setting. In existing studies, handling unknown or out-of-vocabulary (OOV) words was challenging (El-Kassas et al., 2020). In addition, using neural network-based word embeddings for the semantic representation of sentences remained difficult in ATS. Previous studies used less effective part-of-speech (POS) tagging techniques, especially for classifying word classes such as nouns, verbs, and adjectives (Hailu et al., 2020). Owing to noisy and incomplete data, conventional methods exhibited performance stability issues (Jiang et al., 2021). Furthermore, due to the lower score values of existing score-based text summarization techniques, certain important sentences might be excluded from the summary. Most existing ATS methods attained low accuracy and efficiency when summarizing long texts.

The proposed model performs ATS for multi-text documents and multilingual summaries. By incorporating key steps such as preprocessing, slang identification, filtering, and data transformation, it enables effective summarization in these settings. For semantically effective sentence representation, distilled BERT (DBERT)-based entity vectorization is performed. The sememe similarity-induced hidden Markov model (SemSim-HMM) is introduced to perform POS tagging and more accurately classify word categories. Preprocessing steps—including natural language processing (NLP), metadata removal, language identification and translation, as well as uniform resource locator (URL), symbols, and emotions removal—are applied to improve model's stability.

To achieve effective ATS, ensemble methods—including HIN-MELM-AE, stack autoencoder (SAE), variational autoencoder (VAE), and DBERT—are utilized. Data transformation steps, such as info-squared fuzzy C-means (InS-FCM)-based clustering, LDA-based topic modeling, TF-IDF analysis, and frequent term selection, are performed to accurately summarize the longer texts.

In this paper, Section 2 presents the existing works, Section 3 describes the proposed methodology, Section 4 details the materials and methods used in this study, Section 5 discusses the results and discussion, and finally, Section 6 concludes the proposed work with future scope.

2. Literature Review

Hailu et al. (2020) introduced an ATS and evaluation framework based on word embedding. They employed a word embedding-based text summarization technique, and the cosine similarity between each sentence in the document and the keywords was evaluated. The model effectively identified the top-*n* sentences of a source text. However, the study used a less effective POS tagging technique, particularly for classifying word categories such as nouns, verbs, and adjectives.

Jiang et al. (2021) utilized an attention-based bidirectional long short-term memory (LSTM) for hybrid ATS. Four ATS methods—enhanced semantic network, decoder attention based on a pointer network (DA-PN), DA-PN with a coverage mechanism (DA-PN + cover), and mixed learning objective (MLO) function combined with DA-PN + cover (DA-PN + cover + MLO)—were employed. The model effectively addressed the problem of OOV words; however, its performance stability was affected by noisy and incomplete data.

Hernandez-Castaneda et al. (2023) proposed a model for the automatic generation of an objective function for text summarization. In this approach, heuristic functions were automatically generated using genetic programming for ATS. In addition, the model automatically derived an orientation function, resulting in higher-quality summaries. However, the challenge of increased computational complexity with larger text datasets remained.

Zhong & Wang (2022) analyzed ATS for domain adaptation. In this work, a multi-task learning method was employed for ATS, in which bidirectional and auto-regressive transformers were integrated as shared text encoding layers. The model effectively performed ATS across multiple domains; however, it was time-consuming, which introduced significant time complexity issues.

Abo-Bakr and Mohamed (2023) proposed an automatic multi-document text summarization model. Initially, the whole text was preprocessed through sentence segmentation, word tokenization, stop-word removal, and stemming. Then, the large-scale sparse multi-objective optimization algorithm was employed to perform ATS. The model effectively extracted a small set of sentences from a large multi-document text. However, the generated summaries lacked cohesion and semantics, thereby reducing the efficiency of the model.

Hosseinabadi et al. (2022) explored an ATS model based on iterative sentence scoring and extraction schemes. They performed processes such as preprocessing, graph representation, sentence clustering, cluster scoring, sentence selection and elimination, and final rearrangement for ATS. They achieved better results in terms of precision and recall. However, they did not take into account the broader context and the relationship between sentences, which may result in disjointed summaries.

Belwal et al. (2021) employed a topic-based vector space model and a semantic measure for text summarization. They used a combination of topic vector and individual topic vector methods to generate the topic vector from the given document. Eventually, the summarized text obtained from this model was closer to human-generated summaries. The model, however, may lose fine-grained context and subtleties of individual words.

Muniraj et al. (2023) introduced a model for hybrid text summarization of transliterated news articles. They performed ATS using a hybrid Seq2Seq model—the encoder contained three bidirectional LSTM units, and the decoder contained one LSTM unit. The model provided high performance and produced higher Recall-Oriented Understudy Gisting Evaluation (ROUGE) values. However, the hybrid Seq2Seq model required large datasets for training and was prone to overfitting issues, thereby reducing accuracy.

Alami Merrouni et al. (2023) explored a text summarization method for generating both extractive and abstractive summaries. They applied statistical and semantic scoring methods, along with a graph-based approach, were employed for text summarization. The model effectively removed non-essential information and produced coherent, grammatically correct abstractive and extractive summaries. Nevertheless, the model lacked key NLP operations, which affected the quality of the final results.

Kouris et al. (2024) introduced a text summarization model based on semantic graphs. The semantic graph representation, along with reinforcement learning and transformer-based models, was employed for ATS. The proposed model demonstrated promising performance and high robustness. However, the model struggled to handle unknown or OOV words, thereby reducing its effectiveness.

Onan and Alhumyani (2024a) explored an extractive TypeScript framework utilizing fuzzy topic modeling (FuzzyTM) and BERT. They applied fuzzy logic to enhance topic modeling, resulting in more accurate modeling of word-topic relationships. This method achieved ROUGE-1 and ROUGE-2 scores of 45.3774 and 24.1808, respectively. Although this approach improved the quality of text summarization, the performance of the framework was compromised due to its interpretability.

Onan and Alhumyani (2024b) present a novel approach to extractive text summarization (ETS) by integrating large language models with hierarchical positional encoding, which supports deep semantic understanding and improved context awareness in summary generation. However, while the framework shows enhanced performance in capturing and preserving contextual relationships within texts, it incurs high computational costs, which may pose challenges for large-scale or real-time applications.

Hassan et al. (2024) proposed an ETS approach using NLP with an optimal deep learning (ETS-NLPODL) model. They reported that the

ETS-NLPODL model achieved strong performance compared to other models across multiple evaluation measures.

Liu et al. (2024) improved TextRank by using deep contextual embeddings and K-means clustering, yielding more coherent and less redundant summaries with higher ROUGE scores than baseline methods. Although this method is computationally slower and sensitive to cluster initialization, it offers an effective upgrade to classic TextRank for quality-oriented extractive summarization.

Yang et al. (2025) introduced a novel generative adversarial network-based framework that combines transductive LSTM with Distil-BERT embeddings and reinforcement learning to fine-tune sentence selection. The model alleviates greedy biases typical of conventional extractive methods using the discriminator as a reward in a policy gradient setting, and it obtained high ROUGE-1 (52.45), ROUGE-2 (26.46), and ROUGE-L (44.85) scores on standard datasets such as CNN/Daily Mail. The method produces more coherent and concise summaries and offers value for domains where time efficiency and operational costs are critical, such as engineering and healthcare. However, the model's reliance on generative adversarial networks and reinforcement learning makes it computationally expensive, thereby limiting its use to environments with ample computing resources.

Divya and Sripriya (2025) developed a semantics-driven extractive approach utilizing transformer-based multilingual clustering and hybrid clustering algorithms for Tamil and English texts. The system converts text into embeddings using multilingual BERT variants, after which it applies density-based and centroid clustering to identify the most relevant sentences. Evaluations on the constructed bilingual corpora showed that the proposed method outperformed baselines such as LexRank and TextRank in terms of precision and informativeness, resulting in concise summaries that preserve cross-lingual nuances. They reported that their research is instrumental for low-resource language development and that it improved F1-scores for Tamil texts by up to 15%. However, the model may require language-specific adjustments to achieve optimal performance, and it faces challenges in identifying outliers during clustering, which may lead to fragmented summaries in noisy datasets.

Mandale-Jadhav (2025) investigated a multifaceted extractive summarization pipeline that integrated graph-based ranking, deep learning classifiers, and hybrid linguistic-neural features for scalable processing. The system combines TF-IDF for the initial weighting step with graph spreading (using eigenvector centrality) and a small neural component that scores sentence embeddings, enabling efficient processing of benchmarks such as Multi-News. It extends the capabilities of existing methods while achieving a balanced ROUGE performance (for example, a ROUGE-1 value of approximately 48%), and the method is computationally efficient enough for deployment on edge devices. The use of language heuristics also enhances performance across different genres, from technical reports to social media content. Nevertheless, domain-specific tuning is often required to achieve optimal results in specialized fields such as legal or medical domains, and the evaluation revealed issues with handling ambiguous pronouns, which may reduce coherence in lengthy documents.

3. Proposed Novel Ensemble Learning-Based ATS Framework

In the proposed ATS model, the ensemble methods, which include HIN-MELM-AE, SAE, VAE, and DBERT, are used to perform ATS. In addition, the newly improved DePori algorithm is employed to perform slang identification and filtering. The structural diagram of the proposed model is displayed in Fig. 1.

3.1. Text Document

In the first step, the text documents are obtained from the DUC 2004 dataset, and they are used as input to the proposed system. The total m number of text documents (N_M^{text}) is expressed as:

$$N_M^{text} \rightarrow N_1^{text}, N_2^{text}, N_3^{text}, \ldots\ldots, N_m^{text}$$
$$where\ M = (1, 2, \ldots (m) \quad (1)$$

where $M = (1,2,\ldots,m)$ indicates the total number of (N_M^{text}).

3.2. PreProcessing

Next, the (N_M^{text}) are preprocessed (or cleaned) to remove textual noise and improve the incomplete text. The proposed model contains preprocessing operations such as standard NLP procedures, removal of URLs, symbols, and emojis, elimination of metadata, and language identification, followed by translation when necessary.

3.2.1. NLP operations

In the proposed model, NLP operations, such as stop-word removal, stemming, and lemmatization, are performed for the (N_M^{text}). First, the stop-word removal eliminates the common or unwanted words (e.g., "the," "and," "is," and "of") from the text documents (N_M^{text}). Thus, the stop-word–removed text is indicated as ς.

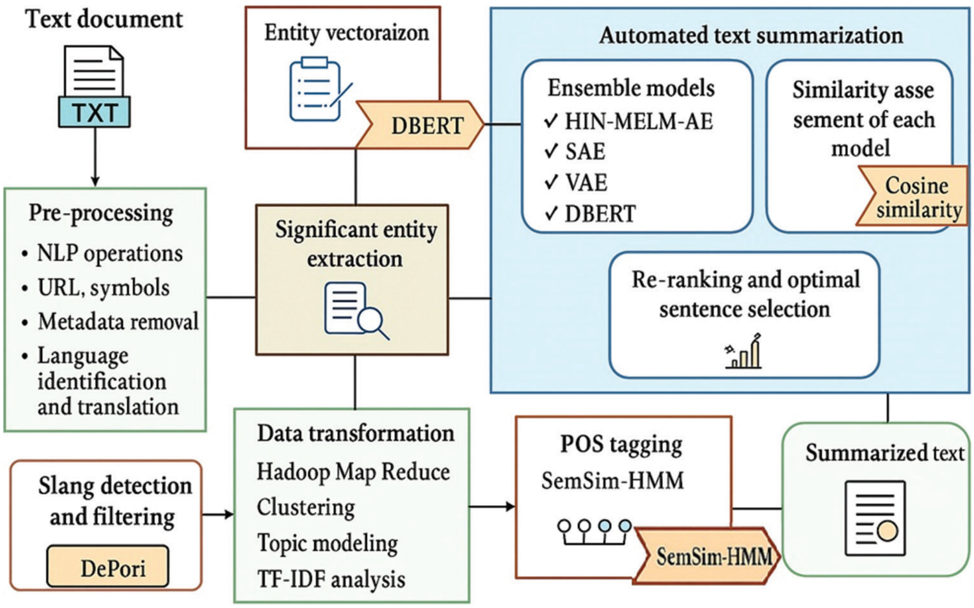


Fig. 1. Structural diagram of the proposed model
Abbreviations: DBERT: distilled bidirectional encoder representations from transformers; DePori: Dehghani poor-and-rich optimization algorithm; HIN-MELM-AE: Hyperfan-IN multilayer extreme learning machine-based autoencoder; NLP: Natural language processing; POS: Parts of speech; SAE: Stack autoencoder; SemSim-HMM: Sememe similarity-induced hidden Markov model; TF–IDF: Term frequency–inverse document frequency; URL: Uniform resource locator; VAE: Variational autoencoder

Next, the last few characters from are removed by the stemming process, which is represented as Ξ. Finally, the lemmatization process converts the word into its meaningful base form. Thus, the lemmatized text is denoted as λ.

3.2.2. URL, symbol, emotion, and metadata removal

Next, the URL, symbols (e.g., @, #, and/), and emotions are removed from the lemmatized text (λ), which is expressed as δ. Then, the metadata removal is performed on the URL-, symbol-, and emotion-removed text (δ). Here, the hypertext markup language tags are removed; thus, the metadata-removed text is denoted as Γ.

3.2.3. Language identification and translation

After that, the language of the Γ is identified, and the text is translated into the English language. Thus, the English-translated text is defined as ε.

Finally, the preprocessed text ($P_{\partial}^{\bullet}$) is expressed as:

$$P_{\partial}^{\bullet}(\varepsilon) \xrightarrow{Preprocessed} \left[P_1^{\bullet} + P_2^{\bullet} + P_3^{\bullet} + \ldots\ldots + P_D^{\bullet}\right] \tag{2}$$

where $P_D^{\bullet}$ specifies the Dth preprocessed text.

3.3. Slang Identification and Filtering

After preprocessing, slang recognition and filtering are performed. Words or expressions in languages other than the source language that are present in the text documents are identified and removed to make the text summarization process more accurate. The improved DePori algorithm is employed for slang identification and filtering. The poor and rich optimization (Pori) algorithm effectively differentiates words into slang and normal words based on the discrimination between poor and rich. Nevertheless, Pori suffers from low convergence speed and local optima issues while solving very complex optimization problems. To address these issues, the Dehghani method (DM) is incorporated into Pori to determine the best member of the population by utilizing information on population positions.

3.3.1. Initialization

Primarily, the initial population is created randomly with uniform distribution between upper-bound and lower-bound values. Here, the initialized population is considered as the preprocessed text ($P_{\partial}^{\bullet}$). In DePori, the main population contains two subpopulations: poor and rich. The main population (J_{main}) is defined as follows,

$$J_{main}\left(P_{\partial}^{\bullet}\right) = J_p + J_r \tag{3}$$

where J_p indicates the initialized population of poor, and J_r denotes the initialized population of rich. Here, (J_{main}) is stored in ascending order, where the first part corresponds to the rich population and the second part corresponds to the poor population. In DePori, all rich population members have better positions than the poor, which is expressed as:

$$asc_1 < asc_2 < asc_3 < \cdots < asc_o < asc_{o+1} < asc_{o+2} < asc_{o+3} < \cdots < asc_{\phi} \tag{4}$$

where asc_o specifies the total number of rich populations, and asc_{ϕ} implies the total number of the main population.

3.3.2. Fitness function

Afterward, the fitness function is calculated based on the lexicon dictionary words (*Lex*) to differentiate the slang and non-slang words. Thus, the fitness ($\Im it$) is defined as:

$$Fit = Lex \times Fit(J_{main}) \tag{5}$$

$$J_{best} = Best(Fit) \tag{6}$$

The best fitness value (*Fit*) obtained corresponds to the best candidate solution J_{best}.

3.3.3. Updating the position of each rich population member

Here, the position of every rich population member is changed. The DM is used to locate the best member of the population using information on population positions, thereby improving the performance of the optimization algorithm. In addition, DM ensures that all members of the population, even the worst, can contribute to population development. Thus, the position update of each rich population member based on DM is given by:

$$\overrightarrow{J_{r,n}^{new}} = \overrightarrow{J_{r,n}^{old}} + I_{DM}^{r}\left[\overrightarrow{J_{r,n}^{old}} - \overrightarrow{J_{p,best}^{old}}\right] \tag{7}$$

$$I_{DM}^{r} = \left[i_{best}^{1} \cdots i_{n}^{r} \cdots i_{best}^{p}\right] \tag{8}$$

where

(i) $\overrightarrow{J_{r,n}^{new}}$ specifies the new value of the rich population's n^{th} position.

(ii) $\overrightarrow{J_{r,n}^{old}}$ denotes the present value of the rich population's n^{th} position.

(iii) I_{DM}^{r} implies the DM for the rich population.

(iv) $\overrightarrow{J_{p,best}^{old}}$ represents the best member of the poor population's present position.

(v) i_{n}^{r} defines the current value of the rich population's n^{th} position based on I_{DM}^{r}.

(vi) i_{best}^{p} signifies the best member of the poor population's present position based on I_{DM}^{r}.

In fact, $\overrightarrow{J_{p,best}^{old}}$ is the best member of the poor population; when a rich population member increases the distance from $\overrightarrow{J_{p,best}^{old}}$, its distance from all poor population members is also increased.

3.3.4. Updating the position of each poor population member

The position update of each poor population member is given in Eq. (9). Here, DM is employed to locate the best member of the population utilizing information on population positions.

$$\overrightarrow{J_{p,n}^{new}} = \overrightarrow{J_{p,n}^{old}} + \left[I_{DM}^{p} - \overrightarrow{J_{p,n}^{old}}\right] \tag{9}$$

$$I_{DM}^{p} = \left[i_{r,best}^{1} \cdots i_{r,mean}^{old} \cdots i_{r,worst}^{old}\right] \tag{10}$$

where:

(i) $\overrightarrow{J_{p,n}^{new}}$ specifies the new value of the poor population member's n^{th} position.

(ii) $J_{p,n}^{old}$ indicates the present value of the poor population's n^{th} position.

(iii) I_{DM}^{p} represents DM for the poor population.

(iv) $i_{r,best}^{1}$ is the best member of the rich population based on I_{DM}^{p}.

(v) $i_{r,mean}^{old}$ signifies the average position of rich population members based on I_{DM}^{p}.

(vi) $i_{r,worst}^{old}$ represents the worst position of rich population members based on I_{DM}^{p}.

Getting rich varies for each person, where the I_{DM} causes changes in their positions.

3.3.5. Mutation

The mutation for each rich and poor population is formulated as:

$$Mu = \begin{cases} if\ ran < \Pr_{mut} & \overrightarrow{J_{r,n}^{new}} = \overrightarrow{J_{r,n}^{new}} + ran \\ if\ ran < \Pr_{mut} & \overrightarrow{J_{p,n}^{new}} = \overrightarrow{J_{p,n}^{new}} + ran \end{cases} \tag{11}$$

where M_u indicates the mutation, Pr_{mut} denotes the mutation probability, and *ran* represents the value obtained from the normal distribution with an average of 0 and variance of 1.

This position updating continues until the best position is obtained. Thus, the total *l* number of non-slang texts (γ_z) is defined as,

$\gamma_z \rightarrow [\gamma_1, \gamma_2, \gamma_3, \ldots\ldots s\gamma_l]$ Here $z = (1 \text{ to } l)$ (12)

where γ_1 specifies the first (γ_z) and γ_l indicates the last (γ_z). The pseudocode for the DePori algorithm is depicted below:

After identifying and filtering, the slang (γ_z) is transformed for effective ATS.

3.4. Data Transformation

Next, data transformation is performed on (γ_z) by employing the Hadoop MapReduce for handling large volumes of text and multiple documents effectively. Here, the data transformation comprises steps such as clustering, topic modeling, TF-IDF analysis, and frequent term selection.

Algorithm 1. Improved pseudocode for Dehghani's poor and rich optimization algorithm

Input: Preprocessed text $\left(\wp_{\partial}^{\bullet}\right)$
Output: Non-slang text $\left(\gamma_z\right)$
Begin
  Initialize main population
  $J_{main}\left(\wp_{\partial}^{\bullet}\right) = J_p + J_r$
  Store J_{main} in ascending order
  $asc_1 < asc_2 < asc_3 < \cdots < asc_o < asc_{o+1} < asc_{o+2} < asc_{o+3} < \cdots < asc_{\phi}$
While $iter \le iter_{\max}$
  For each $\left(J_{main}\right)$
    Compute fitness function
    $\Im it = Lex * \Im it\left(J_{main}\right)$
Update position of each rich population member
$\overrightarrow{J_{r,n}^{new}} = \overrightarrow{J_{r,n}^{old}} + I_{DM}^{r}\left[\overrightarrow{J_{r,n}^{old}} - \overrightarrow{J_{p,best}^{old}}\right]$
Discover Dehghani Method for rich population
$I_{DM}^{r} = \left[i_{best}^{1} \cdots i_{n}^{r} \cdots i_{best}^{p}\right]$
Update position of each poor population member
$\overrightarrow{J_{p,n}^{new}} = \overrightarrow{J_{p,n}^{old}} + \left[I_{DM}^{p} - \overrightarrow{J_{p,n}^{old}}\right]$
Estimate Dehghani Method for poor population
$I_{DM}^{p} = \left[i_{r,best}^{1} \cdots i_{r,mean}^{old} \cdots i_{r,worst}^{old}\right]$
  Find mutation
$Mu = \begin{cases} if\ ran < \Pr_{mut} & \overrightarrow{J_{r,n}^{new}} = \overrightarrow{J_{r,n}^{new}} + ran \\ if\ ran < \Pr_{mut} & \overrightarrow{J_{p,n}^{new}} = \overrightarrow{J_{p,n}^{new}} + ran \end{cases}$
  End For
  End While
  Obtain Non-slang text $\left(\gamma_z\right)$
End

3.4.1. Clustering

First, the clustering process is applied (γ_z) to create the text document clusters. Here, the InS-FCM model is employed for clustering. In general, FCM performs better for data with complex structures or overlapping class boundaries. If outliers exist in the data, they affect the clustering outcomes significantly. If the problem of outliers is avoided, the results of clustering will be more accurate. To solve this problem, the info-squared technique is incorporated into FCM, effectively mitigating the effect of outliers. The clustered data ($Z_{r'}^{cl}$) is formulated as:

$$Z_{r'}^{cl} \Rightarrow \left(Z_1^{cl}, Z_2^{cl}, Z_3^{cl}, \ldots\ldots, Z_{J'}^{cl}\right) \tag{13}$$

Where $r' = (1,2,\ldots,J')$ specifies the clustered data.

3.4.2. Topic modeling

Next, the topic modeling is applied to each clustered of data ($Z_{r'}^{cl}$) for creating the cluster topics and terms belonging to each cluster topic. Here, the LDA approach is used for topic modeling. LDA effectively identifies the hidden topic structure in a set of documents. This can be defined as:

$$\log j(Z \mid a,b) \ge E_{ELB} \left[\log f\left(Z,V,\varpi,\varphi \mid a,b\right) - \log k\left(\varpi,\varphi,V\right)\right] \tag{14}$$

Thus, a good approximation of the posterior distribution is determined by iteratively updating the σ and v. Finally, the topic modeling outcomes are represented as T_c.

3.4.3. Term frequency–inverse document frequency analysis

The TF-IDF analysis is performed for T_c to identify the globally frequent terms from the collection of multiple text documents. The TF-IDF analysis reduces the computational complexity of the model. First, the TF(κ) is calculated by the frequency of the specific term present in T_c:

$$\kappa = \left\| \Omega\left(\tilde{d}\right) \Big/ \tau\left(\tilde{d}, T_c\right) \right\| \tag{15}$$

where $\Omega\left(\tilde{d}\right)$ denotes the number of times, the term $\tilde{d}$ is present in T_c, and $\tau\left(\tilde{d}, T_c\right)$ represents the total number of terms in T_c. The IDF (U) is known as the frequency of tokens (*to*) containing the term. Thus, the IDF is defined as:

$$U = \log\left|\frac{\Omega(to)}{to(\tilde{d})}\right| \tag{16}$$

where Ω(to) denotes the number of tokens present in T_c, and $to(\tilde{d})$ denotes the tokens that contain the term $\tilde{d}$. Next, the TF-IDF can be calculated by multiplying the TF and IDF values:

$$\tilde{F} \Rightarrow (\kappa \times U) \tag{17}$$

where $\tilde{F}$ indicates the obtained TF-IDF score.

3.4.4. Frequent term selection

After the TF-IDF analysis ($\tilde{F}$), the frequent terms are selected. Here, sentence filtering is performed on each individual input text document based on the frequent and semantic similar terms generated from the previous stage, which is indicated as c'. Finally, the total λ' number of transformed data ($Y_{q'}$) is given by:

$$Y_{q'} \rightarrow [Y_1 + Y_2 + Y_3 + \cdots \ldots Y_{\lambda'}] \tag{18}$$

where $Y_{\lambda'}$ specifies the λ'^{th} transformed data.

3.5. POS Tagging

Then, POS is tagged from the ($Y_{q'}$). The SemSim-HMM is used for POS tagging. HMM effectively reduces ambiguities in the sentences and improves tagging accuracy. However, HMM achieves low accuracy for unknown words because the emission probability tends to be zero in such cases. Therefore, the sememe similarity is incorporated with HMM, which evaluates the semantic relationship between sentences through the horizontal semantic relation.

Finally, the tagged POS (Z_ℓ^{POS})is expressed as:

$$Z_\ell^{POS} \rightarrow \left[Z_1^{POS} + Z_2^{POS} + \ldots\ldots + Z_{\rho'}^{POS}\right]$$
$$where\ \ell = (1, 2, ss\rho') \tag{19}$$

where ρ' denotes the number of tagged POS. The pseudocode for SemSim-HMM is depicted below:

Thus, SemSim-HMM effectively performs POS tagging.

Algorithm 2. Pseudocode for sememe similarity-induced hidden Markov model

Input: Transformed data $(Y_{q'})$

Output: Tagged POS $(\Im_\ell^{POS})$

Begin

Initialize $(Y_{q'})$

For each $(Y_{q'})$

Compute set of $\tilde{M}$ hidden state (H_{ab})

Estimate a sequence of $\tilde{P}$ observations (O_{pq})

Derive transition probability matrix

$$W[l', m'] = \Pr(\mu_{t'+1} = H_v \mid \mu_{t'} = H_u)$$

Evaluate emission probability

$$\tilde{V}[v, l'] = \Pr(\varphi'_{t'} = O_{l'} \mid \mu_{t'} = H_v) + \Im'$$

Discover sememe similarity

$$\Im' = \begin{bmatrix} O_1\ Simil\ (H_1, H_2) + O_2\ Simil(H_1, H_2) \\ +O_3\ Simil\ (H_1, H_2) \end{bmatrix}$$

Find initial state probability

$$\theta_u = \Pr(\mu_1 = H_u)$$

Compute SemSim-HMM

$$\delta = (H, O, W, \tilde{V}, \theta)$$

End for

Obtain tagged POS $(\Im_\ell^{POS})$

End

3.6. Significant Entity Extraction

Next, the significant entities, namely statistical and linguistic features, including code quantity principle, noun and verb phrases, content words, proper nouns, cue-phrase, biased words, positive keywords, negative keywords, and sentence length, are extracted from the ($Y_{q'}$) and (Z_ℓ^{POS}). The extracted significant entities ($\beta^\circ_{\varpi'}$) are given by:

$$\beta^\circ_{\varpi'} \Rightarrow \left[\beta^\circ_1, \beta^\circ_2, \beta^\circ_3, \ldots\ldots v\beta^\circ_{\delta'}\right] \tag{20}$$

where $\beta^\circ_{\delta'}$ denotes the number of ($\beta^\circ_{\varpi'}$).

3.7. Entity Vectorization

After that, ($\beta^\circ_{\varpi'}$)undergoes entity vectorization for effective ATS. In this case, the DBERT algorithm is used for entity vectorization. DBERT efficiently handles complex language structures and captures the contextual meaning of words within a sentence.

Initially, ($\beta^\circ_{\varpi'}$) is tokenized into υ tokens $\tilde{\tau}_{to}$. The $\tilde{\tau}_{to}$ is then fed into the DBERT model to obtain *i*' number of token embeddings $\tilde{e}_{to}$.

The next step is to obtain a fixed-sized sentence embedding by applying a pooling operation on the

DBERT model. Thus, the sentence embedding (S) is calculated as:

$$S = \frac{1}{\varepsilon'} \sum_{em=1}^{\varepsilon'} \tilde{C}_{em} \tag{21}$$

where ε' indicates the number of contextualized embeddings.

Subsequently, $\tilde{\tau}_{to}$ are handled by the DBERT model, and it produces contextualized embeddings $\tilde{C}_{em}$.

Then, the embeddings of the two sentences S_1 and S_2 are obtained by using the Siamese network structure. This can be written as:

$$\phi' = DBERT(S_1)$$
$$\kappa' = DBERT(S_2) \tag{22}$$

where ϕ' and κ' represent the embeddings. Then, the cosine similarity (*Sim*) is evaluated between two embeddings ϕ' and κ', which is given by:

$$Sim(\phi', \kappa') = \frac{\phi' \cdot \kappa'}{\|\phi'\| \, \|\kappa'\|} \tag{23}$$

Next, a suitable loss function is chosen for training the BERT model. Thus, the contractive loss function (*Loss*) is defined as:

$$Loss(\phi', \kappa', \nu') = \nu' \cdot Sim(\phi', \kappa') + (1 - \nu') \cdot \max(0, ma - Sim(\phi', \kappa')) \tag{24}$$

where *ma* signifies the hyperparameter, and if S_1 and S_2 are similar, then ν' is 1; otherwise, ν' is 0. Finally, the entity vectorization outcomes are indicated as $X_{\Gamma'}$.

3.8. ATS

Then, ($X_{\Gamma'}$) are subjected as input to the ATS. In this stage, ensemble-based voting, similarity evaluation, and re-ranking with optimal sentence selection are performed to obtain the summarized text.

3.8.1. Ensemble models

The ATS is performed based on the ($X_{\Gamma'}$). In the proposed model, the HIN-MELM-AE is employed to generate the summary text. In general, AEs are used to convert a given input into a lower-dimensional representation. The decoder reconstructs the input from the encoded version. However, AEs are computationally expensive and provide poor performance for text summarization.

To address these issues, multilayer extreme learning machine (ELM)-based AEs (MELM-AEs) are employed for ATS. ELM offers several advantages, such as fast training speed and versatility in handling diverse data types. MELM-AE effectively learns feature representations by adopting a singular value decomposition. Hyperfan-IN initialization is used in MELM-AE for generating the weights and biases of the network to enhance the learning of machine.

Along with the proposed HIN-MELM-AE, SAE, VAE, and DBERT are employed for ATS. Fig. 2 presents the structural visualization of the ensemble models.

3.8.1.1. Hyperfan-IN multilayer extreme learning machine-based AE

The HIN-MELM-AE consists of the input layer, encoder, MELM, decoder, and output layer.

(i) Hyperfan-IN initialization: Here, the weights (ω) and biases (B) of the network are generated by employing Hyperfan-IN initialization, which improves the learning efficiency of ELM. This is expressed as:

$$\omega = \frac{2^1 \operatorname{Re} LU}{2\upsilon'_{\theta'} \, Var\left(e[2]^{X_{\Gamma'}}\right)} \tag{25}$$

$$B = \frac{2^1 \operatorname{Re} LU}{2\upsilon'_{\theta'} \, Var\left(e[2]^{X_{\Gamma'}}\right)} \tag{26}$$

where $\upsilon'_{\theta'}$ specifies the constant, and *Re L U* denotes the rectified linear unit function.

(ii) Encoder phase: Primarily, $X_{\Gamma'}$ are subjected to the input layer I_y, which forwards the data to the multiple hidden layers (L_g). This can be determined as:

$$L_g = \xi(\omega \, I_y + B) \tag{27}$$

where ξ indicates the sigmoid activation function.

(iii) Decoder phase: Finally, in the decoder phase, the output layer reconstructs the summarized text. Thus, the output layer operation is defined by:

$$N_y(X_{\Gamma'}) = (\omega \, A_g + B) \tag{28}$$

Thus, the obtained sentences are denoted as λ_Γ.

3.8.1.2. SAE

SAE effectively learns the codings of unlabeled data. The SAE contains two parts: an encoder and a decoder.

Firstly, the $X_{\Gamma'}$ is subjected as an input to the input layer In_{AE} which forwards the data to the encoder. Then, the In_{AE} is mapped by the encoder to a latent space representation, which is performed by utilizing

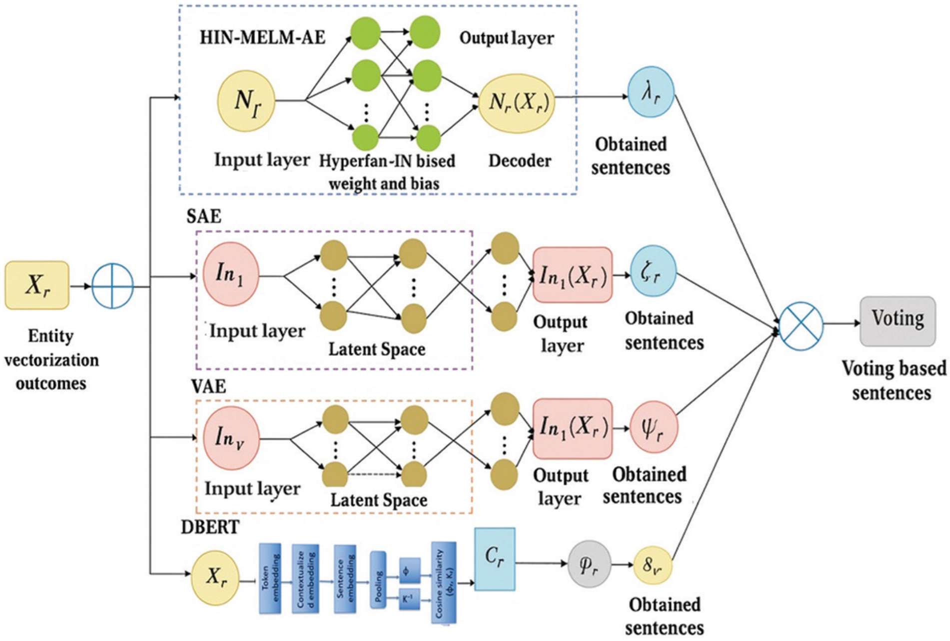


Fig. 2. Structural representation of ensemble models
Abbreviations: DBERT: distilled bidirectional encoder representations from transformers; HIN-MELM-AE: Hyperfan-IN multilayer extreme learning machine-based autoencoder; SAE: Stack autoencoder; VAE: Variational autoencoder

one or more hidden layers. Thus, the encoder operation is written as:

$$\hbar = \sigma\left(\tilde{W}_{en} In_{AE} + \tilde{b}_{en}\right) \tag{29}$$

where $\tilde{W}$ signifies the weight of SAE, $\tilde{b}$ indicates the bias term of SAE, σ denotes the activation function, and $\hbar$ represents the latent space representation. Then, the $\hbar$ is mapped back to the original input space by the decoder. Thus, the decoder operation is given by:

$$In_{AE} = \sigma\left(\tilde{W}_{de} \hbar + \tilde{b}_{de}\right) \tag{30}$$

Next, the decoded information In_{AE} is passed to the output layer, which delivers the outcomes. Eventually, the obtained sentences are represented as $\chi_{n''}$.

3.8.1.3. VAE

A VAE combines a neural network with variational inference. The VAE consists of an encoder, a latent state distribution, and a decoder.

Initially, the $X_{\Gamma''}$ is fed into the input layer $\tilde{I}_{VAE}$, which forwards the data to the encoder. Then, the $\tilde{I}_{VAE}$ is mapped by the encoder to a latent space distribution (la). Thus, the encoder operation is given by:

$$\vartheta_{pa}\left(la \mid \tilde{I}_{VAE}\right) = \tilde{G}\left(la; \mu'_{pa}(\tilde{I}_{VAE}), \alpha_{pa}\left(\tilde{I}_{VAE}\right)^2\right) \tag{31}$$

where $\vartheta_{pa}\left(la \mid \tilde{I}_{VAE}\right)$ specifies the posterior distribution, signifies the encoder parameters, $\tilde{G}$ implies the Gaussian distribution, and μ' and α are the mean and standard deviation, respectively. The latent variable is determined as follows:

$$\tilde{E}(la) = \tilde{G}(0, Id) \tag{32}$$

where Id is the identity matrix. Then, is mapped back to the original input space by the decoder. This is given by:

$$\tilde{E}_{co}\left(\tilde{I}_{VAE} \mid la\right) = \tilde{G}\left(\tilde{I}_{VAE};\ \mu_{co}(la),\ \sigma_{co}(la)^2\right) \tag{33}$$

where $\tilde{E}_{co}\left(\tilde{I}_{VAE} \mid la\right)$ defines the likelihood, and co signifies the decoder parameters. Then, the decoded information $\tilde{I}_{VAE}$ is passed to the output layer, which delivers the outcomes. Eventually, the obtained sentences are represented as $\psi_{s'}$.

3.8.1.4. DBERT

The $X_{\Gamma'}$ is fed to the DBERT model for ATS. The processes involved in the DBERT model are explained in Section 3.7. Lastly, the obtained sentences are indicated as ζ'_{mn}.

Finally, the voting-based sentences ($S_{\partial'}^{vot}$) are obtained, which are represented as:

$$S_{\partial'}^{vot} \rightarrow S_1^{vot}, S_2^{vot}, S_3^{vot}, \ldots\ldots, S_{\tilde{\rho}}^{vot} \; where \; \partial' = (1, 2, bt\tilde{\rho}) \tag{34}$$

where $\partial' = (1, 2, \ldots, \tilde{\rho})$ specifies the number of $S_{\partial'}^{vot}$.

3.8.2. Similarity evaluation of each model

A similarity between sentences is estimated from the voting-based sentences ($S_{\partial'}^{vot}$). Here, cosine similarity is used to calculate the similarity between sentences. Thus, the cosine similarity (*Cos*) is given by:

$$Cos = \frac{S_1^{vot} \cdot S_2^{vot}}{\left\| S_1^{vot} \right\| \left\| S_2^{vot} \right\|} \tag{35}$$

The cosine similarity is calculated for all voting-based sentences. Thus, the evaluated similarity between all the voting-based sentences is indicated as $\Theta_{l'}$.

3.8.3. Re-ranking and optimal sentence selection

After similarity evaluation ($\Theta_{l'}$), the voting-based fusion is performed. Then, the sentences are rearranged, and optimal sentences are selected to obtain the summarized text. Finally, the obtained summarized text ($R_{\tau'}^{text}$) is formulated as:

$$R_{\tau'}^{text} \Rightarrow \left[R_1^{text}, R_2^{text}, R_3^{text}, \ldots\ldots, R_{\mu'}^{text} \right] \tag{36}$$

Where $R_{\mu'}^{text}$ specifies the μ'^{th} summarized text. Thus, the proposed model effectively performs ATS with slang filtering.

4. Materials and Methods

4.1. Software Requirements

The proposed model was implemented in the Python platform. Python is a widely used general-purpose and high-level programming language designed to emphasize code readability. The syntax of Python allows developers to define concepts in fewer lines of code. In addition, Python efficiently integrates with systems and enables rapid development. Python is a multipurpose programming language and has been employed in many applications, such as machine learning, scientific computing, and automation. Furthermore, Python's extensive standard library is equipped with modules and functions for a variety of frequently occurring tasks.

4.2. Hardware Requirements

The hardware requirements for the proposed model are as follows:

(i) Processor: Intel Core i5/Core i7
(ii) Central processing unit speed: 3.20 GHz
(iii) Operating system: Windows 10
(iv) System type: 64-bit
(v) Random-access memory: 4 GB.

4.3. Dataset Description

The DUC 2004 dataset was used to assess the proposed model. This dataset was obtained from publicly available sources. The DUC 2004 dataset consists of 1,358 text documents, of which 80% (*n* = 1,087) were employed for training, and the remaining 20% (*n* = 271) were employed for testing.

5. Results and Discussion

In this section, the performance analysis and comparative assessment of the proposed model are discussed to demonstrate the model's reliability.

5.1. Performance Evaluation of the Proposed HIN-MELM-AE Model

The performance of the proposed model and existing techniques, such as AE, sentence bidirectional encoder representations from transformers (SBERT), LSTM, and RNN, was evaluated to demonstrate the model's reliability.

Fig. 3 illustrates the graphical analysis of the proposed HIN-MELM-AE and existing techniques in terms of accuracy, precision, recall, F-measure, sensitivity, ROUGE 1, and ROUGE 2. The proposed HIN-MELM-AE achieved high accuracy, precision, recall, F-measure, sensitivity, ROUGE 1, and ROUGE 2 of 97.92%, 99.16%, 98.01%, 98.24%, 98.01%, 0.935145, and 0.911371, respectively.

In contrast, the existing techniques achieved lower average accuracy, precision, recall, F-measure, sensitivity, ROUGE 1, and ROUGE 2 of 91.59%, 93.05%, 90.21%, 91.59%, 90.21%, 0.833706, and 0.761878, respectively. This indicates that the proposed HIN-MELM-AE utilized the MELM-based AE to achieve superior performance for text summarization.

A comparative assessment of the proposed model and existing techniques regarding execution time and error is presented in Table 1. Here, the proposed model utilizes the Hyperfan-IN initialization for generating the weights and biases of the network, which enhances the learning process. The proposed HIN-MELM-AE obtained a low execution time and error of 51,015 ms

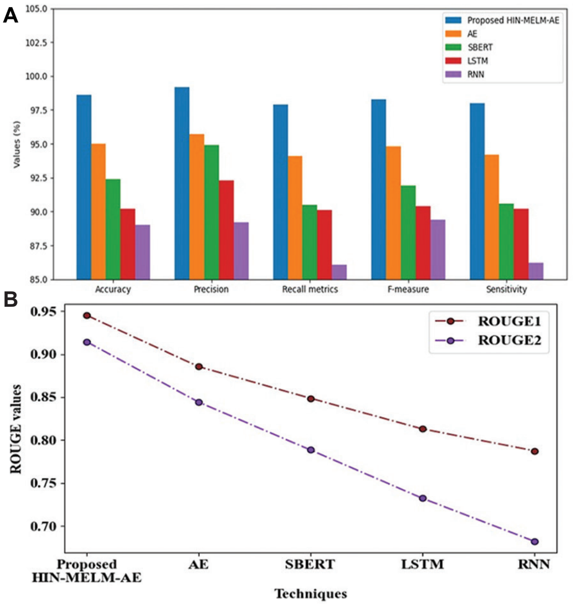


Fig. 3. Graphical analysis in terms of (A) accuracy, precision, recall, F-measure, and sensitivity, and (B) ROUGE 1 and ROUGE 2
Abbreviations: AE: Autoencoder; IN-MELM-AE: Hyperfan-IN multilayer extreme learning machine-based autoencoder; LSTM: Long short-term memory; RNN: Recurrent neural network; ROUGE: Recall-oriented understudy gisting evaluation; SBERT: Sentence bidirectional encoder representations from transformers

and 0.011766, respectively. However, the existing SBERT and RNN achieved a higher execution time of 62,008 ms and 83,012 ms, respectively. In addition, the existing AE and LSTM exhibited higher errors of 0.051064 and 0.097872, respectively. These results indicate that the proposed model performs better for text summarization.

5.2. Performance Estimation of Ensemble Models

The performance estimation of ensemble models was conducted to assess the model's reliability.

Fig. 4 displays the performance evaluation of ensemble models in terms of summary size versus recall and summary size versus ROUGE. The proposed HIN-MELM-AE achieved a high recall and ROUGE of 0.9857 and 0.9132, respectively, for a summary size of 50. On the other hand, AE, VAE, and SBERT achieved a low recall of 0.8942, 0.8215, and 0.7745, respectively, for a summary size of 50.

In addition, AE, VAE, and SBERT achieved lower ROUGE values of 0.8412, 0.7916, and 0.7195, respectively, for a summary size of 50. This indicates that the proposed HIN-MELM-AE achieved superior

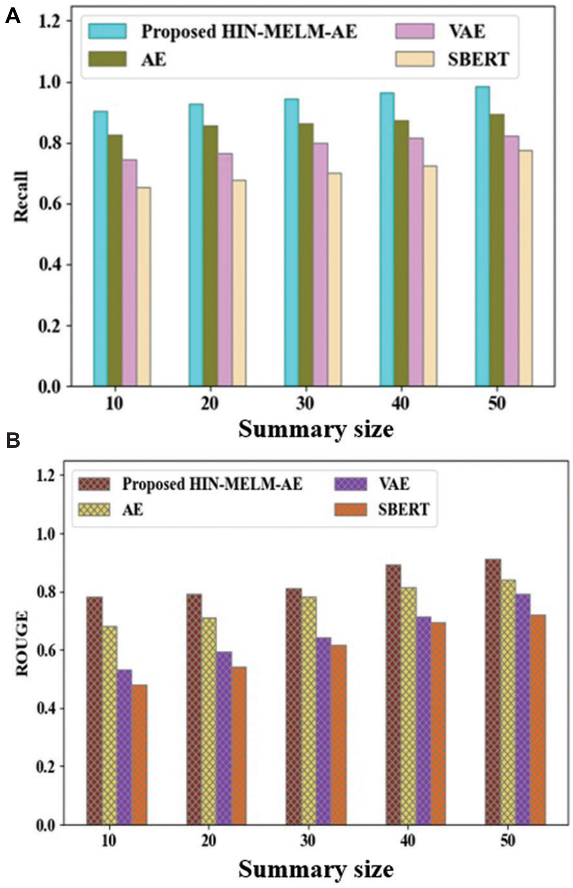


Fig. 4. Performance evaluation of ensemble models in terms of (A) summary size versus recall and (B) summary size versus ROUGE
Abbreviations: AE: Autoencoder; HIN-MELM-AE: Hyperfan-IN multilayer extreme learning machine-based autoencoder; ROUGE: Recall-oriented understudy gisting evaluation; SBERT: Sentence bidirectional encoder representations from transformers; VAE: Variational autoencoder

Table 1. Comparative assessment of the proposed model and existing techniques

Methods	Execution time (ms)	Error
Proposed HIN-MELM-AE	51,015	0.011766
Existing SAE	57,006	0.051064
Existing SBERT	62,008	0.076596
Existing LSTM	68,018	0.097872
Existing RNN	83,012	0.110638

Abbreviations: HIN-MELM-AE: Hyperfan-IN multilayer extreme learning machine-based autoencoder; LSTM: Long short-term memory; RNN: Recurrent neural network; SAE: Stack autoencoder; SBERT: Sentence bidirectional encoder representations from transformers.

performance in text summarization due to the use of MELM-based AE and Hyperfan-IN initialization.

5.3. Performance Assessment of the Proposed DePori

The performance assessment of the proposed DePori was compared with conventional techniques, such as Pori, honey badger algorithm (HBO), fennec fox optimization (FFO), and sail fish optimizer (SFO).

Fig. 5 displays the comparison of fitness versus iteration for the proposed model and traditional methods, including Pori, HBO, FFO, and SFO. The proposed model employed DM to locate the best member of the population using information on population location. The proposed DePori achieved a high fitness of 88.12 in the 10th iteration and 96.37 in the 40th iteration.

In contrast, HBO achieved lower fitness values of 85.26 in the 10th iteration and 91.26 in the 40th iteration. Other existing methods also obtained low fitness values. This indicates that the proposed model effectively performs slang identification and filtering.

5.4. Performance Analysis of the Proposed InS-FCM

The performance analysis of the proposed InS-FCM and existing techniques was performed to assess the model's performance.

Table 2 presents the performance validation of the proposed InS-FCM and conventional methods, including FCM, K-means, K-medoids, and balanced iterative reducing and clustering using hierarchies (BIRCH) in terms of clustering time and silhouette score. The proposed InS-FCM obtained a low clustering time and a high silhouette score of 6,134 s and 0.9085, respectively. On the other hand, the existing techniques achieved a high mean clustering time of 13,382.25 s and a low silhouette score of 0.7737. In this approach, the info-squared method is combined with FCM, which effectively mitigates outlier-related issues.

Fig. 6 illustrates the clustering accuracy of the proposed model and the existing techniques. The proposed InS-FCM achieved a high clustering accuracy of 96.90%. On the other hand, the existing techniques, such as FCM, K-means, K-medoids, and BIRCH, obtained a low clustering accuracy of 93.89%, 92.49%, 90.35%, and 87.32%, respectively. These results indicate the superiority of the model.

5.5. Comparative Analysis of the Proposed Model

The comparative analysis of the proposed model and related works was performed to assess the model's reliability.

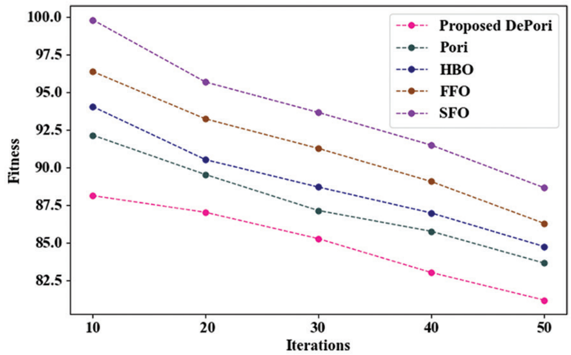


Fig. 5. Fitness versus iteration analysis. Abbreviations: DePori: Dehghani poor and rich optimization algorithm; FFO: Fennex fox optimization; HBO: Honey badger algorithm; Pori: Poor and rich optimization algorithm; SFO: Sail fish optimizer

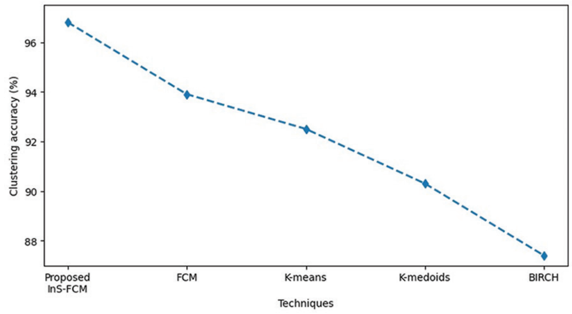


Fig. 6. Graphical representation of clustering accuracy. Abbreviations: BIRCH: Balanced iterative reducing and clustering using hierarchies; FCM: Fuzzy C-means clustering; InS-FCM: Info-squared fuzzy C-means clustering

Table 2. Performance validation in terms of clustering time and silhouette score

Techniques	Clustering time (s)	Silhouette score
Proposed InS-FCM	6,134	0.9085
FCM	9,219	0.8475
K-means	12,237	0.7954
K-Medoids	14,389	0.7542
BIRCH	17,684	0.6978

Abbreviations: BIRCH: Balanced iterative reducing and clustering using hierarchies; FCM: Fuzzy C-means clustering; InS-FCM: Info-squared fuzzy C-means clustering.

Table 3 presents the comparative analysis of the proposed HIN-MELM-AE model and related works. The proposed HIN-MELM-AE model, which utilizes the MELM-based AE and Hyperfan-IN for initialization, achieved a high ROUGE 1 and ROUGE 2

Table 3. Comparative analysis

Techniques	Dataset	ROUGE 1	ROUGE 2	References
HIN-MELM-AE	DUC 2004 dataset	0.94014	0.90437	Proposed model
FA		0.4782	0.2295	Tomer and Kumar (2022)
DNN		0.42813	0.23668	Onan and Alhumyani (2024a)
USE transformer		0.4286	0.1425	Lamsiyah et al. (2021)
Karci's summarization approach		0.58669	0.31804	Hark & Karcı (2020)

Abbreviations: DNN: Deep neural network; FA: Firefly algorithm; HIN-MELM-AE: Hyperfan-IN multilayer extreme learning machine-based autoencoder; MMR: Maximal marginal relevance; ROUGE: Recall-oriented understudy gisting evaluation; USE: Universal sentence encoder.

values of 0.94014 and 0.90437, respectively. However, the existing firefly algorithm achieved a low ROUGE-1 of 0.4782 using the DUC 2004 dataset. Furthermore, the existing maximal marginal relevance method and PageRank algorithm achieved a low ROUGE-1 score of 0.684. Moreover, the existing deep neural network, universal sentence encoder transformer, and Karci summarization approach achieved low ROUGE-1 and ROUGE-2 scores, potentially reflecting their higher computational complexity. These findings indicate that the proposed model performs ATS more effectively.

6. Conclusion

This study presents an improved ensemble learning-based ATS with slang filtering using HIN-MELM-AE and DePori. The DUC 2004 dataset was used to train the proposed model. In this study, processes—such as text document acquisition, preprocessing, slang identification and filtering, data transformation, POS tagging, significant entity extraction, entity vectorization, ensemble models, similarity evaluation of each model, and re-ranking and optimal sentence selection—were performed. The proposed HIN-MELM-AE achieved high accuracy, precision, and recall of 97.92%, 99.16%, and 98.01%, respectively. In addition, the proposed InS-FCM achieved a low clustering time and high clustering accuracy of 6,234 s and 96.90%, respectively. Furthermore, the proposed DePori achieved a high fitness of 88.12 in the 10th iteration and 99.81 in the 50th iteration. These findings indicate that the proposed model performs ATS with slang filtering more effectively. However, the proposed model is limited to summarizing text from documents and does not handle context in videos. Future studies could focus on developing techniques for summarizing context in videos to further enhance the model's applicability.

Acknowledgments

The authors would like to thank the supervisor and mentors for their invaluable guidance, support, and encouragement throughout the research. The authors are also grateful to the institution for offering the essential facilities and resources needed to complete this work.

Funding

None.

Conflict of Interest

The authors declare that they have no competing interests.

Author Contributions

Conceptualization: Sunil Upadhyay
Data curation: Sunil Upadhyay
Methodology: Sunil Upadhyay
Writing – original draft: Sunil Upadhyay
Writing – review & editing: All authors

Availability of Data

The dataset used in this study was obtained from https://www.kaggle.com/datasets/usmanniazi/duc-2004-dataset.

References

Abo-Bakr, H., & Mohamed, S.A. (2023). Automatic multi-documents text summarization by a large-scale sparse multi-objective optimization algorithm. *Complex and Intelligent Systems*, 9(4), 4629–4644 https://doi.org/10.1007/s40747-023-00967-y

Alami Merrouni, Z., Frikh, B., & Ouhbi, B. (2023). EXABSUM: A new text summarization approach for generating extractive and abstractive summaries. *Journal of Big Data*, 10(1), 1–34. https://doi.org/10.1186/s40537-023-00836-y

Alomari, A., Idris, N., Sabri, A.Q.M., & Alsmadi, I. (2022). Deep reinforcement and transfer learning for abstractive text summarization: A review.

Computer Speech and Language, 71, 1–43. https://doi.org/10.1016/j.csl.2021.101276

Awasthi, I., Gupta, K., Bhogal, P.S., Anand, S.S., & Soni, P.K. (2021). Natural Language Processing (NLP) Based Text Summarization-A Survey. In: *Proceedings of the 6th International Conference on Inventive Computation Technologies*, p1310–1317. https://doi.org/10.1109/icict50816.2021.9358703

Belwal, R.C., Rai, S., & Gupta, A. (2021). Text summarization using topic-based vector space model and semantic measure. *Information Processing and Management*, 58(3), 102536. https://doi.org/10.1016/j.ipm.2021.102536

Divya, S., & Sripriya, N. (2025). Semantic based extractive document summarization using deep learning model. *ICTACT Journal on Soft Computing*, 15(4), 3669–3681. https://doi.org/10.21917/ijsc.2025.0509

El-Kassas, W.S., Salama, C.R., Rafea, A.A., & Mohamed, H.K. (2021). Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*, 165, 1–46. https://doi.org/10.1016/j.eswa.2020.113679

Gupta, H., & Patel, M. (2021). Method of Text Summarization Using Lsa and Sentence Based Topic Modelling with Bert. In: *International Conference on Artificial Intelligence and Smart Systems*, p511–517. https://doi.org/10.1109/ICAIS50930.2021.9395976

Haider, M.M., Hossin, M.A., Mahi, H.R., & Arif, H. (2020). Automatic Text Summarization Using Gensim Word2Vec and K-Means Clustering Algorithm. In: *IEEE Region 10 Symposium, TENSYMP 2020*, p283–286. https://doi.org/10.1109/TENSYMP50017.2020.9230670

Hailu, T.T., Yu, J., & Fantaye, T.G. (2020). A framework for word embedding based automatic text summarization and evaluation. *Information (Switzerland)*, 11(2), 1–23. https://doi.org/10.3390/info11020078

Hark, C., & Karci, A. (2020). Karci summarization: A simple and effective approach for automatic text summarization using Karcı entropy. *Information Processing and Management*, 57(3), 1–16. https://doi.org/10.1016/j.ipm.2019.102187

Hassan, A.A., Al-Onazi, B.B., Maashi, M., Darem, A.A., Abunadi, I., & Mahmud, A. (2024). Enhancing extractive text summarization using natural language processing with an optimal deep learning model. *AIMS Mathematics*, 9(5), 12588–12609.

Hernandez-Castaneda, A., Garcia-Hernandez, R.A., & Ledeneva, Y. (2023). Toward the automatic generation of an objective function for extractive text summarization. *IEEE Access*, 11, 51455–51464. https://doi.org/10.1109/access.2023.3279101

Hosseinabadi, S., Kelarestaghi, M., & Eshghi, F. (2022). ISSE: A new iterative sentence scoring and extraction scheme for automatic text summarization. *International Journal of Computers and Applications*, 44(6), 1–6. https://doi.org/10.1080/1206212X.2020.1829844

El-Kassas, Wafaa S., Cherif R. Salama, Ahmed A. Rafea, and Hoda K. Mohamed. (2020). EdgeSumm: Graph-Based framework for automatic text summarization. *Information Processing and Management,* 57(6), 102264. https://doi.org/10.1016/j.ipm.2020.102264.

Jain, M., & Rastogi, H. (2020). Automatic Text Summarization using Soft-Cosine Similarity and Centrality Measures. In: *Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology*, p. 1021–1028. https://doi.org/10.1109/iceca49313.2020.9297583

Jiang, J., Zhang, H., Dai, C., Zhao, Q., Feng, H., Ji, Z., et al. (2021). Enhancements of attention-based bidirectional LSTM for hybrid automatic text summarization. *IEEE Access*, 9, 123660–123671. https://doi.org/10.1109/access.2021.3110143

Kouris, P., Alexandridis, G., & Stafylopatis, A. (2024). Text summarization based on semantic graphs: An abstract meaning representation graph-to-text deep learning approach. *Journal of Big Data*, 11(1), 1–39. https://doi.org/10.1186/s40537-024-00950-5

Lamsiyah, S., Mahdaouy, E., Espinasse, B., & Ouatik, A. (2021). An unsupervised method for extractive multi-document summarization based on centroid approach and sentence embeddings. *Expert Systems with Applications*, 167, 114152.

Liu, W., Sun, Y., Yu, B., Wang, H., Peng, Q., Hou, M., et al. (2024). Automatic text summarization method based on improved TextRank algorithm and K-Means clustering. *Knowledge-Based Systems*, 287, 111447. https://doi.org/10.1016/j.knosys.2024.111447

Mandale-Jadhav, A. (2025). Text summarization using natural language processing. *Journal of Electrical Systems*, 20(11), 3410–3417. https://doi.org/10.52783/jes.8095

Manjari, K.U., Rousha, S., Sumanth, D., & Devi, J.S. (2020). Extractive Text Summarization from Web pages using Selenium and TF-IDF algorithm. In: *Proceedings of the 4th International Conference on Trends in Electronics and Informatics*, p648–652. https://doi.org/10.1109/icoei48184.2020.9142938

Muniraj, P., Sabarmathi, K.R., Leelavathi, R., &

Balaji, B.S. (2023). HNTSumm: Hybrid text summarization of transliterated news articles. *International Journal of Intelligent Networks*, 4, 53–61. https://doi.org/10.1016/j.ijin.2023.03.001

Onah, D.F.O., Pang, E.L.L., & El-Haj, M. (2022). A Data-driven Latent Semantic Analysis for Automatic Text Summarization using LDA Topic Modelling. *IEEE International Conference on Big Data (Big Data)*, 2771–2780. https://doi.org/10.1109/BigData55660.2022.10020259

Onan, A., & Alhumyani, H.A. (2024a). DeepExtract: Semantic-driven extractive text summarization framework using LLMs and hierarchical positional encoding. *Journal of King Saud University - Computer and Information Sciences*, 36(8), 1–19. https://doi.org/10.1016/j.jksuci.2024.102178

Onan, A., & Alhumyani, H.A. (2024b). FuzzyTP-BERT: Enhancing extractive text summarization with fuzzy topic modeling and transformer networks. *Journal of King Saud University - Computer and Information Sciences*, 36(6), 102080.

Payak, A., Rai, S., Shrivastava, K., & Gulwani, R. (2020). Automatic Text Summarization and Keyword Extraction using Natural Language Processing. In: *Proceedings of the International Conference on Electronics and Sustainable Communication Systems*, p98–103. https://doi.org/10.1109/icesc48915.2020.9155852

Prasad, C., Kallimani, J.S., Harekal, D., & Sharma, N. (2020). Automatic Text Summarization Model using seq2seq Technique. In: *Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud*, p599–604. https://doi.org/10.1109/I-SMAC49090.2020.9243572

Syed, A.A., Gaol, F.L., & Matsuo, T. (2021). A survey of the state-of-the-art models in neural abstractive text summarization. *IEEE Access*, 9, 13248–13265. https://doi.org/10.1109/access.2021.3052783

Tomer, M., & Kumar, M. (2022). Multi-document extractive text summarization based on firefly algorithm. *Journal of King Saud University - Computer and Information Sciences*, 34(8), 6057–6065. https://doi.org/10.1016/j.jksuci.2021.04.004

Wahab, M.H.H., Ali, N.H., Hamid, N.A.W.A., Subramaniam, S.K., Latip, R., & Othman, M. (2024). A review on optimization-based automatic text summarization approach. *IEEE Access*, 12, 4892–4909. https://doi.org/10.1109/access.2023.3348075

Widyassari, A.P., Rustad, S., Shidik, G.F., Noersasongko, E., Syukur, A., Affandy, A., et al. (2022). Review of automatic text summarization techniques and methods. *Journal of King Saud University - Computer and Information Sciences*, 34(4), 1029–1046. https://doi.org/10.1016/j.jksuci.2020.05.006

Yadav, D., Katna, R., Yadav, A.K., & Morato, J. (2022). Feature based automatic text summarization methods: A comprehensive state-of-the-art survey. *IEEE Access*, 10, 133981–134003. https://doi.org/10.1109/access.2022.3231016

Yang, J., Wang, H., Qin, H., Sun, Y., Khan, A.A., Por, L.Y., et al. (2025). A generative adversarial network-based extractive text summarization using transductive and reinforcement learning. *IEEE Access*, 13, 65490–65509. https://doi.org/10.1109/access.2025.3558266

Zhang, M., Zhou, G., Yu, W., & Liu, W. (2020). A Survey of Automatic Text Summarization Technology Based on Deep Learning. In: *Proceedings - International Conference on Artificial Intelligence and Computer Engineering*, p211–217. https://doi.org/10.1109/icaice51518.2020.00047

Zhong, J., & Wang, Z. (2022). MTL-DAS: Automatic text summarization for domain adaptation. *Computational Intelligence and Neuroscience*, 2022, 1–10. https://doi.org/10.1155/2022/4851828

AUTHOR BIOGRAPHIES

Sunil Upadhyay received his M.TECH. Degree from the Department of Computer Science and Engineering at RGEC Meerut, Gautam Buddh Technical University, formerly known as UPTU, Lucknow, U.P., India, in 2012. He is pursuing his Ph.D. degree in Computer Science and Engineering from Amity University, Madhya Pradesh. His current research interest includes NLP, Data analytics, Artificial Intelligence, and Machine Learning.

Hemant Kumar Soni completed his M Tech (IT) at Bundelkhand University, Jhansi, Uttar Pradesh, India, and a Doctoral degree in Computer Science and Engineering from Amity University, Madhya Pradesh, Gwalior, India. He has 26 years of teaching experience for UG and PG courses in Computer Science and is presently working as a Professor in the Department of Computer Science and Engineering at Amity University, Madhya Pradesh, India. His research interests are in Natural Language Processing, Data Science, Machine Learning, and Data Mining. He published many research papers in Web of Science, SCI, and Scopus Indexed Journals. His research articles received high levels of citations in Scopus and Google Scholar. He is a Reviewer of many referred journals including IEEE Transactions.

DOI: 10.6977/IJoSI.202512_9(6).0004
Q. Liu & T.I. Tajuddin/Int. J. Systematic Innovation, 9(6), 44-62 (2025)

Artificial intelligence-collaborative folk music composition system based on gesture recognition: A real-time interactive framework integrating computer vision and folk music generation

Qinghao Liu[1], Tazul Izan Tajuddin[1,2]*

[1]Faculty of Music, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

[2]Institut Seni Kreatif Nusantara, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

*Corresponding author E-mail: tazulizan@uitm.edu.my

(Received 11 August 2025; Final version received 07 November 2025; Accepted 02 December 2025)

Abstract

Artificial intelligence (AI) and gesture recognition offer new creative possibilities, yet culturally sensitive, real-time systems for gestural folk music composition remain largely undeveloped. This study develops an AI-collaborative folk music composition system that integrates computer vision-based gesture recognition with specialized folk music generation algorithms to create a real-time interactive framework that supports traditional music composition while preserving cultural musical characteristics across multiple folk traditions. The system employs a four-layer architecture encompassing gesture acquisition, computer vision processing, interpretation, and generation layers. A comprehensive dataset of 1,643 folk music compositions from established repositories representing English, American, Irish, and Chinese traditional music (Nottingham Dataset, Irish Traditional Corpus, and self-recorded materials) was curated, supplemented by 6,127 successfully tracked gesture samples collected from 47 participants across 12 folk music gesture categories. The evaluation framework assessed gesture recognition accuracy, cultural authenticity preservation, real-time performance, and collaborative effectiveness through extensive experimental validation. The system achieved robust gesture recognition performance with 88.9% accuracy and 23.4 ms processing latency, while maintaining end-to-end response times of 86.8–91.6 ms during collaborative sessions. Cultural authenticity scores ranged from 7.6 to 8.3 across different regional folk styles, with a user satisfaction rating of 7.8 and a 28% improvement in musical coherence compared to baseline approaches. The framework successfully supports up to eight concurrent users while maintaining sub-100 ms real-time performance requirements. The integrated system successfully demonstrates effective coordination between gesture recognition and folk music generation subsystems, validating the architectural design and optimization strategies for culturally sensitive AI applications across diverse folk music traditions. The validated framework provides a foundation for educational, performance, and cultural preservation applications, contributing methodological insights for multimodal human–AI interaction systems and culturally aware creative technologies applicable to traditional music contexts.

Keywords: Artificial Intelligence-Collaborative Music Composition, Computer Vision, Folk Music Generation, Gesture Recognition, Real-Time Interactive Framework, Traditional Music

1. Introduction

Rapid development of artificial intelligence (AI) has greatly altered many areas of human creation, and one of the promising areas for AI application is musical composition (Hernandez-Olivan & Beltran, 2022). Combining computer vision, gesture recognition, and AI-generated music reveals new possibilities for designing intelligent systems that can interpret and respond based on human creative intention under real-world conditions. This integration would be quite appealing for folk music creation, as it offers opportunities to develop sophisticated human–computer interaction patterns for diverse geographic and ethnic origins of traditional culture, and still maintain its authenticity and complexity. This research exclusively focuses on designing a framework for composing multicultural folk music that would cover the traditional music of different regions, such as English, American, Irish, and Chinese folk music, taking into consideration that each folk music tradition has some commonalities while also retaining distinctive characteristics.

Recent advancements in AI research have shown remarkable success in many methods, ranging from deep architectures to symbolic music modeling systems (Ji et al., 2023). In this regard, automatic methods of composing music through AI systems remain a broad area of activity, encompassing many computational approaches (Civit et al., 2022). Recently, comprehensive research on AI music composition applications demonstrates that there is a growing number of choices available for incorporating this technology creatively (Chen et al., 2024). Nevertheless, applications of these technologies, despite their great potential for culture preservation and creation, have been rarely explored, especially concerning music composition for folk culture.

These distinctive elements, including its melody, rhythmic components, and expression of emotion, have been shown to be supportive and motivating, not only for the AI system that is attempting to mimic intelligent output, but also for producing more valid output (Sturm & Ben-Tal, 2021). In addition, going beyond its technical applications, its philosophical applications for AI systems within music involve more basic concepts of creating, authoring, and originality of culture itself (Berkowitz, 2024). Text-to-music creation techniques demonstrate ongoing research into the advanced functionality of AI techniques for producing large stylistic output from text itself (Zhao et al., 2025). Furthermore, the future applications of machine learning techniques into music continue to expand into this new, imaginable realm of human–computer interaction and creation (Liang, 2023).

Machine learning and computer vision have kept pace with developments in modern gesture recognition systems, and more powerful tools have emerged for high-quality and interactive music systems (Dalmazzo et al., 2021). There have been outstanding breakthroughs in computer vision methods of musical transcription, and the processing of visual information into symbolic music expression has been achieved effectively (Li et al., 2020). The successful deployment of such gesture recognition capabilities within musical contexts requires systematic integration across multiple computational layers, addressing distinct processing requirements from visual input to musical output. Systems enabling multimodal interaction between human and computer are becoming increasingly able to understand and interpret complex user inputs performed through more than one perceptual channel (Jia et al., 2020). The improvement of performer–audience interaction through technological mediation has become an emerging research area with significant implications for the live musical performance experience (Otsu et al., 2021).

Affective music composition systems face inherent challenges in composing music that is emotionally meaningful to humans in terms of human emotions and cultural background (Dash & Agres, 2024). The application of machine learning in music generation and composition has been successful across diverse music genres and styles (Dawande et al., 2023). Deep network architectures tailored for music generation are becoming increasingly complex to produce higher-quality output (Pricop & Iftene, 2024). Extensive investigation of multimodal interaction interfaces effectively demonstrates the challenges in developing robust systems for human–computer communication (Dritsas et al., 2025), requiring careful coordination among gesture capture, interpretation, and generation subsystems to maintain coherent real-time operation.

Generative AI in music has educational potential and implications for traditional pedagogy, considering how this technology might be integrated (Cheng, 2025). The synthesis of music sound using machine learning techniques has made considerable progress in producing perceptually relevant control of a synthesizer through user input (Roche, 2020). In commercialized musical contexts, the implementation of deep neural networks in music industry processes provides pragmatic evidence of AI technology adoption (Fan, 2022). Moreover, psychological studies of neural processing suggest that musicians are trained to have heightened sensitivity to music-relevant aspects of the musical context, acquired through experience (Hansen et al., 2022).

The appearance of music perception skills in neural networks without human supervision is in line with our hypothesis that the networks might develop a kind of internal understanding of music through learning from various musical pieces (Kim et al., 2024). Deep learning approaches based on music genre identification have made remarkable progress, demonstrating their potential for better style recognition and generation (Yimer et al., 2023). Studies on soundscape features in human–computer interaction contexts have demonstrated the relevance of environmental and contextual aspects when designing musical interfaces that effectively coordinate music-sound with non-music sound (Johansen et al., 2022). Dynamically coping with uncertainty is a continuing problem for machine learning systems implemented for applications that need guaranteed and predictable performance (Kapoor, 2025).

Fine-grained interactive guidance for symbolic music generation is also a major step toward obtaining full control over the operations of AI music authoring (Zhu et al., 2024). However, despite such achievements, many limitations of current AI music generation systems exist, limiting their application effectiveness in folk music. Many models cannot be sensitive enough to the subtle attributes of folk music traditions and inevitably generate results that are not realistically stylized due to constraints in meaningful cultural expressions. The challenge is exacerbated when considering the real-time nature of interaction required for gesture-based systems, where latency and responsiveness are critical to preserving an effective creative flow.

To effectively address these issues of real-time system performances, especially under uncertain operating settings, sophisticated control techniques from the theoretical framework of non-linear systems, as described by gesture-based human–computer interaction, need to be applied. Recent efforts based on fuzzy control, aiming for practical fixed-time synchronization of fractional-order chaotic systems (Boulkroune et al., 2025), and output-feedback controllers based on projective lag synchronization of uncertain chaotic systems subjected to input non-linearities (Boulkroune et al., 2017), have laid theoretical foundations for addressing interoperability and system-wide time compatibility between complex computational components with diverse time characteristics. These approaches can be collectively considered as offering methodologies to provide practical solutions for time compatibility and gesture synchronization, particularly under variable lighting or sensor noises in gesture recognition algorithms used as input components of gesture-to-music translators.

To address complex interactions between gesture, musical parameters, and generators—aiming for comprehensive theoretical foundations—control structures from neural networks, especially as proposed for uncertain complex dynamical multivariable systems and based on advanced neural network online prediction methods (Zouari et al., 2012), and hierarchical adaptive backstepping methods concerning uncertain single-input single-output (SISO) non-linear systems (Zouari et al., 2013a), which align conceptually with multi-layer processing architectures, are required for gesture-to-music translation.

Gas-compressor systems based on induction motors, utilizing practical implementation and validation through non-linear optimal control methods (Rigatos et al., 2023), and flexible robotic systems based on DC motors—adopting practical implementation through adaptive backstepping control methods (Zouari et al., 2013b)—have demonstrated that complex adaptability techniques can be effectively applied within strict time-constrained responsive systems. These control methods could be integrated into gesture–music systems involving gesture-based folk music generation, potentially offering improved resistance to environmental changes, as well as more accurate gestural expression interpretation, while maintaining cultural integrity required for folk music tradition and transmission.

Gesture recognition and AI-generated music face complex technical challenges, as they involve coordinating multiple tasks across different computer systems. Research on gestural control of AI-generated folk music is currently facing challenges, particularly the need for research that focuses on AI-generated folk music while taking into consideration characteristic features and traditions within different cultures worldwide. By addressing this research gap, the present study develops a comprehensive framework that incorporates advanced computer vision technology and specialized AI-generated folk music algorithms based on multicultural materials of traditional folk music, establishing a system of gestural human–AI cooperation on music creation grounded in Western, Celtic, and Eastern folk music traditions.

2. Methods

2.1. System Architecture Design

To develop the proposed AI-assisted folk music composition system, a complex architectural design is required that effectively incorporates gesture processing functionality and folk music processing functionality into one efficient system that is still fully responsive and culturally compatible with diverse folk music traditions. The system design was formulated to address one of its core challenges, namely, managing multiple computer processing systems operating at different time scales while retaining the natural, free-flowing characteristic typical of folk music performances. To enable a clear understanding of this complex system, Fig. 1 outlines the architectural system design for the proposed system.

Fig. 1 illustrates the system architectural hierarchy, where the gesture acquisition layer interacts solely with computer vision processing modules. Gestural input is bridged via musical parameters by the interpretation layer, and folk music composition based on culturally appropriate folk music is achieved through specialized neural modules in the generation layer. There are four primary processing layers that operate in parallel to optimize latency, system responsiveness, and musical composition quality. In addition, a dependency relationship exists between each of these processing layer components, with input components receiving raw streaming data (e.g., depth sensor images for gesture acquisition, feature vectors for gestural input, musical parameters for musical composition), processing components implementing

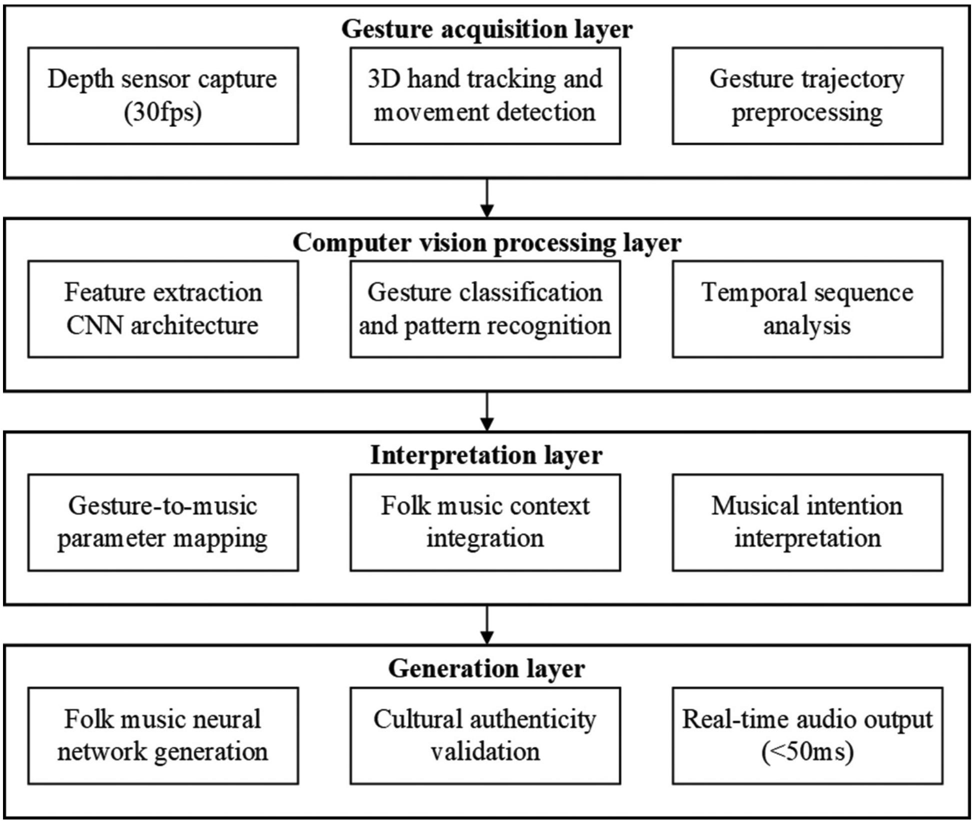


Fig. 1. Multi-layer system architecture for artificial intelligence-collaborative folk music composition
Abbreviations: 3D: Three-dimensional; CNN: Convolutional neural network

operations specific to each respective layer (e.g., spatial–temporal feature extraction for computer vision processing, gestural input processing, and constraint-based musical composition for generation), and output components that verify and pass valid processing output downstream through standardized interfaces for continuous system operation and final musical composition output.

The proposed common architecture compares favorably with other methods due to its ability to incorporate domain-specific parameters of folk music structures and performance practices into its framework, thereby facilitating more informed gesture interpretive processing and more appropriate musical responses based on these considerations (Rezwana & Maher, 2023). To address issues of system operation uncertainty and adaptability, the system design integrates control concepts based on principles of non-linear systems theory, incorporating an adaptive feedback system that continuously monitors and responds based on gesture recognition degrees of confidence and processing delays. This is achieved through decision rules formulated on fuzzy logic and techniques of hierarchical backstepping control, that is, system operations that aggregate performance levels from lower processing scales to higher gesture or co-processing scales over ongoing system operation, with parameters governed by backstepping control methodologies.

A key benefit for developers is the ability to optimize each goal separately, namely, gesture recognition accuracy and music output quality, due to straightforward inter-module communication based on a set of standardized data exchange protocols. All elements of this system operate within the time constraints of their respective stages, ensuring that all operations occur in real time. This applies to gesture processing, which supports up to 30 fps camera capture, and music generation, which maintains a maximum latency of 50 ms. In addition, the systems include facilities for dynamic resource allocation, allowing processing priorities to adjust based on system load and interaction intensity, thereby maintaining stable quality of service across varied system configurations.

2.2. Computer Vision-Based Gesture Recognition and Real-Time Interaction

The computer vision module develops new state-of-the-art deep learning architectures designed for musical gesture recognition and real-time classification in folk music performance. The gesture recognition methodology uses hierarchical feature extraction methods, which exploit spatial–temporal hand movement patterns to inform secondary models based on convolutional neural networks trained to discriminate between intended musical gestures and unintended body motion observed during creative sessions. To illustrate the complex processing chain that converts raw gestural input into musical parameters, we present a detailed description of the gesture recognition pipeline in Table 1.

Table 1 presents a comprehensive system configuration of gesture recognition, including input data formats, feature extraction and classification algorithms, and processing of time scales necessary for comprehending gestural expressions. This design takes into consideration parameters such as Gaussian spatial smoothing—σ of 1.2, determined by experimental criteria aimed at balancing noise elimination with hand movement detail preservation. It also incorporates a time window width of 15 frames, based on the criterion of processing time below 30 ms, and a prediction threshold of 0.85, chosen to minimize false positives during prolonged interaction. The gestural classification framework captures movement on varied time scales, enabling comprehension of immediate reactive gesture as well as artistic gestural expressions corresponding to musical interpretations that remain cognizant of cultural context.

To facilitate better visualization of the sequence of data flows and interdependencies within the four processing stages outlined in Table 1, Fig. 2 shows a comprehensive flow diagram illustrating the gesture recognition pipeline architecture, including the corresponding data transformations and/or operations at each stage of processing.

Fig. 2 illustrates the step-by-step processing pipeline that generates valid gesture classifications from raw depth sensor input, based on four synchronous processing tasks executed under stringent time restrictions. The total processing latency of 23.4 ms, well below the target value of 30 ms corresponding to the frame rate of 30 fps, demonstrates an optimal pipeline that renders gestural interaction effortless and free of delays. Every processing task is allocated its own processing resources, as indicated in Table 2, and buffer handling ensures loss-free processing during concurrent user interactions, which is critical for gestural sequence processing tasks involving longer durations of collaborative activity.

The 12 folk music gesture types include universal conducting gestures that can be applied across different traditions of tempo, dynamic marking, and boundary marking. Culture-specific expressive gestures typical of a particular folk music tradition—including articulation gestures, expression gestures, and melodic indication gestures—refer to pitch gestalt characteristic of scales typical of that tradition. While some of these gestures can be interpreted based on general musical principles, others exhibit culture-specific nuances unique to certain folk music performances.

Moreover, the respective modules for recognition and feature extraction are enabled using advanced machine learning techniques, allowing the system to adapt through continuous learning, especially based on observation and experience. This enhances its reliability under uncertain environments, particularly when recognizing gestures under continuously

Table 1. Gesture recognition processing pipeline specifications

Component	**Parameter**	**Specification**	**Unit**
Input data format	Depth resolution	640×480	Pixels
	Frame rate	30	Fps
	Data type	16-bit depth	-
	Detection range	0.5–4.5	Meters
Feature extraction	Hand keypoints	21	Points
	Feature vector dimension	126	Dimensions
	Temporal window size	15	Frames
	Spatial smoothing filter	Gaussian σ = 1.2	-
Classification algorithm	CNN layers	3 convolutional+2 FC	Layers
	LSTM hidden units	128	Units
	Gesture classes	12 folk music gestures	Classes
	Model parameters	2.4 M	Parameters
Temporal processing	Processing latency	23	ms
	Buffer size	8	Frames
	Prediction confidence threshold	0.85	-
	Gesture sequence length	0.5–3.0	s

Abbreviations: CNN: Convolutional neural network; FC: Fully-connected; LSTM: Long short-term memory.

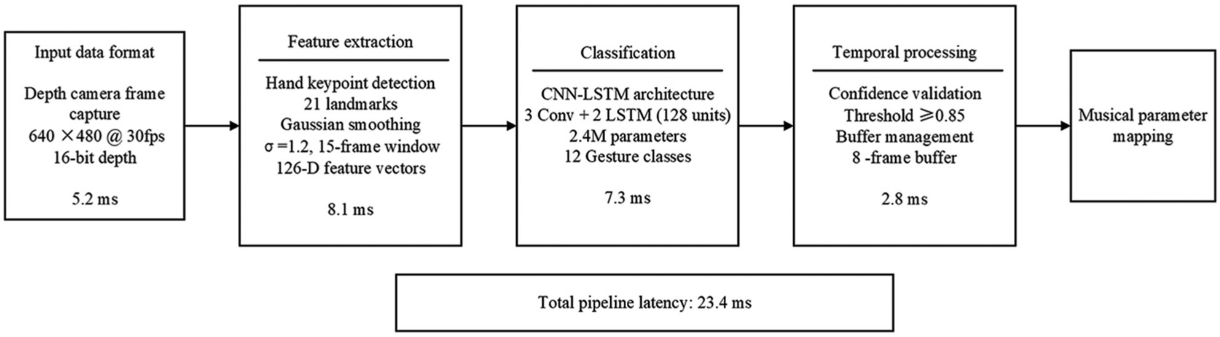


Fig. 2. Gesture recognition pipeline flow diagram
Abbreviations: CNN: Convolutional neural network; Conv: Convolutional; LSTM: Long short-term memory

Table 2. System integration and real-time optimization configuration

Integration component	Parameter	Specification	Notes
Inter-module communication	Data transfer protocol	UDP with error correction	Low-latency priority
	Message queue size	256 buffers	Ring buffer
	Latency requirement	<10 ms	Design target
	Buffer overflow handling	Drop the oldest policy	-
Temporal synchronization	Master clock frequency	48 kHz	Audio sample rate
	Sync tolerance	Maximum±2 ms	Design constraint
	Drift compensation	Linear interpolation	Algorithm choice
	Frame alignment window	5 frames	Configuration
Resource allocation	CPU core assignment	4 cores dedicated	Gesture+music+control+input/output
	Memory pool size	512 MB pre-allocated	Static allocation
	Priority scheduling	Real-time FIFO	Linux RT kernel
	GPU memory allocation	2 GB reserved	CUDA buffers
Load balancing	Thread pool configuration	8–16 adaptive threads	Dynamic scaling
	Load threshold	Maximum 75% CPU	Trigger point
	Migration strategy	Priority-based	Algorithm design
	Monitoring interval	100 ms	Configuration
Caching strategy	Folk pattern cache	128 MB allocated	Design specification
	Gesture model cache	64 MB reserved	Pre-loaded models
	Replacement policy	LRU with priority	Algorithm choice
	Refresh strategy	Adaptive aging	Implementation method

Abbreviations: CPU: Central processing unit; CUDA: Compute unified device architecture; FIFO: First in, first out; GPU: Graphics processing unit; LRU: Least recently used; RT: Real-time; UDP: User datagram protocol.

changing fluorescent lighting or during user interactions that involve hand occlusions. In such cases, when occlusions occur below a predefined threshold or during prolonged interactions that reduce gesture recognition confidence, the system automatically activates recalibration using predefined spatial filtering parameters and interaction window sizes. These constraints prevent instability or lack of adaptability due to inter-user variabilities in corresponding gestural vocabularies and interaction performance, especially based on conceptual and theoretical approaches drawn from "backstepping methods" for adapting and treating uncertain non-linear systems, and through predetermined hierarchical structures that enable parameters of feature extractions based on local measures and parameters of classification based on aggregated system behavior and performance, particularly over extended durations of collaborations.

The real-time interaction optimization addresses crucial latency issues by using predictive gesture completion algorithms that anticipate trajectory movements based on initial gesture segments and accumulated musical context, rather than relying solely on complete gesture pattern recognition. Recent machine learning advances have proven successful in creating multimodal systems that meaningfully interpret complex human interaction for creative applications (Chang et al., 2024). The gesture-to-music mapping reveals developer-specified relationships between characteristics of gesture and musical characteristics, including the generation of melodic intervals, patterns of rhythmic forms, and progressive development of harmonic progressions that honor traditional folk music practices across multiple cultural contexts.

The adaptive sampling rates and processing priorities were configured to dynamically allocate computing resources in accordance with the complexity of gestures and musical context requirements. The temporal windowing method examined gestural sequences at multiple timescales to sustain musical coherence and to afford responsive interaction with the folk music generation algorithms, ensuring that the generated pieces embody the performer's creative ideas while conforming to the music idioms and stylistic conventions characteristic of diverse folk music traditions.

2.3. AI-Collaborative Folk Music Generation Framework

The folk music generation model establishes a custom neural architecture to learn traditional melodic figures, harmonic gestures, and rhythmic styles characteristic of diverse folk music traditions, incorporating real-time gestural input from the musician to enable collaborative composition. The generation method implements dual-pathway processing, in which gestural input is processed through dedicated folk music feature extractors, while maintaining adherence to cultural music conventions learned through extensive training on the curated folk music datasets encompassing English, American, Irish, and Chinese traditional music materials across regional stylistic variations.

To illustrate the comprehensive workflow integrating gestural analysis with constraint-aware folk music synthesis, Fig. 3 presents the AI-collaborative folk music generation framework, demonstrating the interaction between user creative intentions and culturally informed generation mechanisms.

Fig. 3 illustrates how the advanced processing framework, in addition to gestural input analysis and folk music generation algorithms, preprocesses real-time user intentions together with traditional musical knowledge to generate musically viable compositions within diverse folk music idioms. The generation network integrates folk music constraints during synthesis through weighted loss functions and rule-based filtering mechanisms, constraining the musical outputs to preserve traditional styles and patterns—characteristic of the target folk tradition—while remaining sensitive to gestural input and user creative intentions through probabilistic online selection methods biased toward musical coherence and cultural authenticity. The workflow operates through continuous iterative cycles, where gestural parameters extracted from real-time user input influence the generation process, learned folk music representations from the training corpus constrain output stylistic characteristics, and validation mechanisms evaluate musical sequences against cultural authenticity criteria before synthesis, ensuring that generated compositions maintain traditional stylistic properties while incorporating dynamic gestural expression.

The common decision framework hinges on gesturally driven user intention and AI-generated musical resources utilizing a weighted probability distribution that balances user input and musical experience gained over time. Recent developments within machine-generated musical methods have yielded promising outcomes for complex musical expression, simulating sophisticated musical output while retaining characteristic musical elements of a certain genre and offering manipulation and exploration functions (Ferreira et al., 2023). The folk musical pattern recognition module compares output strings with predetermined folk musical elements, namely melodic phrases and harmonic structures of target folk music, and enhances stylistically appropriate musical output through constraint optimization techniques.

The generation mechanism involves probabilistic sampling, thereby enabling variation that is consistent with the structural and stylistic constraints typical of varied folk music traditions and ensuring that, through constraint optimization, user interaction patterns facilitate refinement of generation parameters that remain consistent with or adhere to the constraint of retaining the culture and emotion inherent in folk music expressions as manifested through varied regional folk music traditions of diverse cultures worldwide.

2.4. System Integration and Optimization Strategies

System integration techniques enable harmonious coordination of gesture recognition, computer vision processing, and folk music generation systems, as they support natural, real-time functionality and user comfort during musical composition tasks. System integration helps address complex challenges in coordinating modeling tasks that involve numerous scales, thereby ensuring system responses remain natural and musical, especially during extended

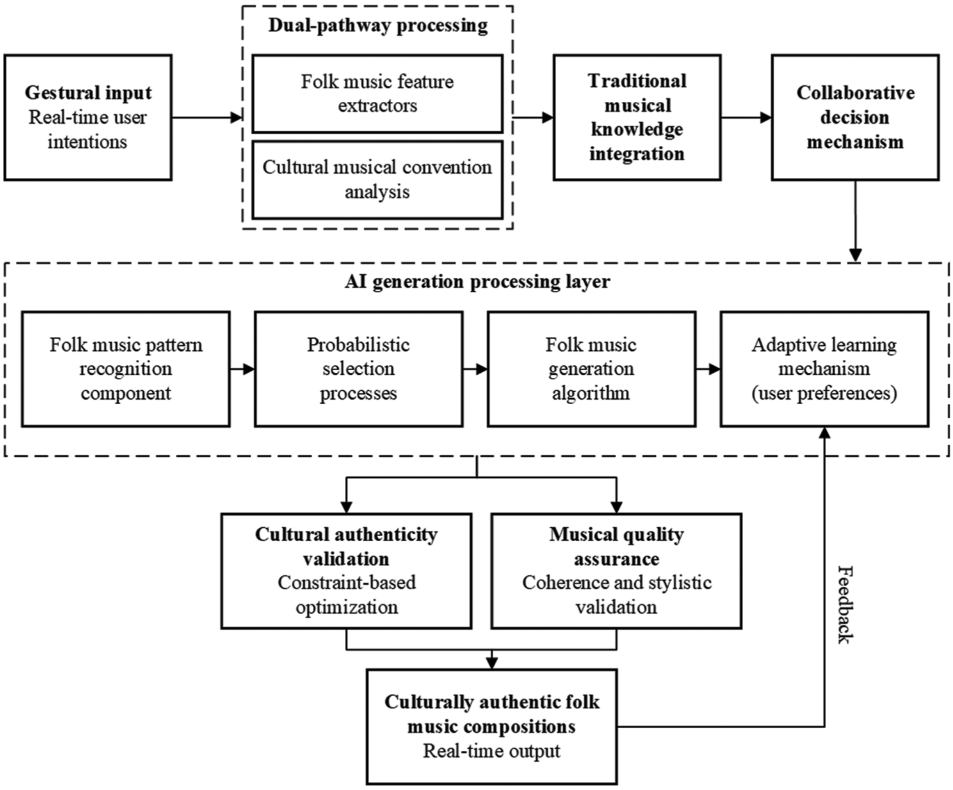


Fig. 3. Artificial intelligence-collaborative folk music generation workflow

performances. To illustrate the system complexity involved in coordinating these systems, particularly during musical composition tasks, system integration techniques and optimization procedures are detailed in Table 2.

Table 2 presents the technical specifications that have a prominent role in inter-module interaction, time synchronization, and resource allocation methods required for maintaining stable real-time processing, where a central processing unit (CPU) load of 75% was identified by stress testing, while keeping a folk pattern cache of 128 MB, which is more than enough for storing regularly accessed melodic phrases and harmonic structures of folk music from varied regional communities, without risk of memory overflow. In addition, this optimizing framework embeds sophisticated resource allocation methods that allocate processing resources based on system loads, interaction, and musical complexity, thereby ensuring stable system processing and minimizing the impact of hardware and musical processing complexities.

Real-time synchronization system implementation uses techniques based on concepts of fractional-order synchronization methods and lag synchronization, harmoniously combining audio quality and latency problems for prolonged musical cooperation tasks and operations. In addition, based on theoretical hypotheses of non-linear optimal control methods, this system adjusts buffer and processing priority control based on time variations and processing operations, implementing optimal controls for time-aligned gesture input synchronization for multiuser cooperation, thereby addressing control problems arising due to system responses below sub-100 ms processing tasks that result from system control restrictions.

Audio visualization systems have shown potential and practical implementation for presenting significant system responses based on musical inputs, demonstrating potential applications for advanced musical interface development (Graf et al., 2021). The system adopts load-balancing techniques that rebalance processing loads based on system processing and available hardware resources, strictly adhering to the time constraints essential for musical tasks and operations.

The integration framework incorporates intelligent caching methods for common folk musical patterns and gestural modeling that can reduce the computational complexity required for managing tasks within a processing environment operating in a real-time environment. The calibration procedures facilitate automatic adjustment based on varied hardware settings and musical environments through standardized approaches that verify system functionality and determine appropriate system configuration settings for different musical application environments. The error handling methods integrate recovery techniques that maintain system functionality and musical continuity during rigorous musical interaction processing, thereby ensuring consistent high-quality functionality of the AI-collaborative folk music composition system despite varied application environments and system settings.

3. Results

3.1. Experimental Environment and Dataset Construction

The comprehensive evaluation of the proposed AI-collaborative folk music composition system required establishing a robust experimental infrastructure capable of supporting real-time gesture recognition and music generation processes while maintaining the stringent performance requirements essential for interactive creative applications. The experimental platform design addressed the complex computational demands associated with simultaneous computer vision processing, neural network inference, and audio synthesis operations that characterize the integrated system architecture. To provide detailed specifications of the computational resources and system configuration employed throughout the experimental validation process, Table 3 presents the complete hardware configuration and associated technical specifications.

Table 3 presents the complex hardware architecture of the real-time AI-collaborative folk music composition system, particularly the parallel processing requirements for gesture recognition and music generation occurring simultaneously. The experimental environment was designed such that the neural network computations were accelerated by high-performance graphics processing units, while the audio synthesis and real-time interaction management were managed through semi-dedicated modules, providing fluid behavior over long collaborative music creation sessions.

For the dataset, the integration process gathered data from major public resources (using the Nottingham dataset as its core), augmented by Irish traditional tunes and additional purpose-recorded materials, thereby ensuring that melodic, harmonic, and rhythmic content appropriate for gesture-based interaction research was sufficiently represented. Table 4 provides details about the dataset sources, technical formats, and data distribution of the curated folk music dataset, which was used for system training and evaluation.

Table 4 presents the multicultural composition of the folk music dataset, which encompasses Western traditions (Nottingham Dataset with English and American materials), Celtic traditions (Irish traditional corpus), and Eastern traditions (Chinese folk music recordings). This composition enables the generation algorithms to leverage cross-cultural structural commonalities while preserving regional stylistic characteristics, ensuring adequate sample sizes for effective neural network training and comprehensive system evaluation across multiple folk music styles.

To facilitate easier collection and annotation of gesture data, the proposed framework implements structured procedures for the collection and classification of distinct expressions and performances of hand gestures based on folk music composition structures made available through public datasets. To annotate these gestures, this framework requires input from musical experts on gestural intention accuracy and musical parameters, thereby producing a specialized database of gestural–musical links based on valid folk music composition structures that are essential for implementing gestural recognition

Table 3. Experimental hardware platform configuration and specifications

Component category	Specification	Model/configuration	Performance notes
CPU	Intel Core i7-12700K	12 cores (8P+4E), 3.6–5.0 GHz	Real-time processing
GPU	NVIDIA RTX 4070	12GB GDDR6X, 5888 CUDA cores	Neural network acceleration
System memory	32 GB DDR4	3200 MHz, dual channel	Pre-allocated buffers
Storage	1TB NVMe SSD	PCIe 4.0, 5,500 MB/s read	Dataset storage
Audio interface	Focusrite Scarlett 2i2	24-bit/192kHz, 2.5ms latency	Professional audio input/output
Depth camera	Azure Kinect DK	1MP depth, 30 fps, 0.5–3.86 m range	Gesture capture
Operating system	Ubuntu 22.04 LTS	Real-time kernel patch	Real-time scheduling support
Development framework	CUDA 12.1, PyTorch 2.0	Python 3.10 environment	Artificial intelligence model inference
Audio processing	JACK Audio Server	128 sample buffer, 48 kHz	Low-latency audio
Mewmory allocation	512MB dedicated	Ring buffers+cache pools	Static allocation strategy

Abbreviations: CPU: Central processing unit; CUDA: Compute unified device architecture; GPU: Graphics processing unit.

and interpretation algorithms accordingly. Table 5 provides the gestural data collection and the associated annotation framework that enables accurate gestural interpretations for folk music performances.

Table 5 presents the strategy used for gestural expression collection and classification based on several example pieces of public folk music datasets, demonstrating the focused nature of this annotation strategy and the required number of training examples for optimal gesture classification accuracy. The collection strategy included diverse participants, such as experienced musicians and music students, to enhance system generalizability based on different interaction behavior patterns typically observed in gestural music interaction tasks.

3.2. Gesture Recognition Performance Evaluation

To ascertain the efficacy of computer vision approaches for gesture recognition and classification, the system developed using computer vision techniques was tested rigorously to evaluate its performance in recognizing and classifying musical gestures effectively in a real-time environment for musical composition tasks. Moreover, given the requirement for rigorous testing of gesture expressions

Table 4. Folk music dataset composition and regional distribution statistics

Dataset source	Compositions	Duration (hours)	Music format	Key features
Nottingham dataset	1,200	42.3	ABC notation	English/American folk music, MIDI available
Irish traditional corpus	287	18.7	Audio+annotations	Celtic folk, detailed onset labels
Self-recorded folk songs	156	8.9	WAV+MIDI	Chinese folk music, gesture-optimized
Total dataset	1,643	69.9	Mixed formats	Multicultural folk music
Training set	1,150 (70%)	48.9	-	AI model training
Validation set	247 (15%)	10.5	-	Hyperparameter tuning
Test set	246 (15%)	10.5	-	Final evaluation

Abbreviations: AI: Artificial intelligence; MIDI: Musical instrument digital interface; WAV: Waveform audio file format.

Table 5. Gesture data collection framework and annotation statistics

Collection category	Parameter	Specification	Details
Data collection approach	Reference music source	Nottingham and Irish datasets	Participants gesture to the public dataset songs
	Selected compositions	387 songs	Subset chosen for gesture recording
	Recording mode	Listen and conduct	Real-time gesturing while hearing music
	Gesture types captured	12 folk music categories	Tempo, dynamics, phrasing, melodic indication
Participant demographics	Expert musicians	18	Folk music performers/conductors
	Music students	29	Advanced undergraduate/graduate
	Age range	20–58 years	Diverse experience levels
	Gender distribution	24 female, 23 male	Balanced participation
Recording sessions	Total sessions	94	3-4 songs per session
	Session duration	52 min (average)	Including calibration and breaks
	Gesture samples collected	6,834	Multiple takes per song
	Successfully tracked	6,127 (89.7%)	Clean depth tracking data
Annotation framework	Music-gesture mapping	Song-specific annotations	Link gestures to musical features
	Expert annotators	4	Folk music and gesture specialists
	Gesture-music parameters	723 validated mappings	Tempo, dynamics, phrase boundaries
	Inter-annotator agreement	0.73 (κ)	Substantial agreement level
Quality control	Rejected samples	707 (10.3%)	Poor tracking or unclear intent
	Validation subset	1,025 samples (15%)	Cross-validation testing
	Average annotation time	4.1 min/sample	Manual verification process
	Gesture sequence length	1.2–5.8 s	Based on musical phrases

that are associated with folk music, and the necessity for these tests to be conducted within strict time parameters essential for efficient artistic composition, Fig. 4 shows the results of gestural accuracy and real-time processing capability analysis for this system developed using computer vision techniques.

Based on Fig. 4A, the findings reveal that strong gesture recognition accuracy was observed for 12 folk music gesture classes, achieving classification accuracies of between 83.7% and 94.2%, with an average of 88.9%. Tempo control and dynamic gestures demonstrate significantly high accuracy, thereby supporting computer vision technology as an effective method for recognizing musical gestural expressions in folk music performance. In Fig. 4B, superior real-time processing was observed, achieving a total latency of 23.4 ms, which is lower than the target of 28.0 ms. All processing components performed below the target, thereby confirming effective system optimization for gesture-enabled folk music interaction.

Temporal analysis of system performance parameters, based on differing levels of computation and interaction intensity, verified its ability to sustain stable levels of accuracy and response time over extended musical composition sessions driven by user gestural input. Evaluation of system performance indicated successful fulfillment of the design parameters established for gestural interaction, maintaining acceptable processing delays and sustaining sufficient accuracy levels for meaningful user musical intent identification based on gestural input interaction within a musical composition environment.

3.3. Folk Music Generation Quality Evaluation

The AI-mediated folk music generation framework was quality-tested for the efficacy of the customized neural architectures and culture-specific appropriation mechanisms designed to generate traditional folk music based on real-time gestural input. The evaluation technique, which combined objective musical analysis and subjective assessment conducted by renowned music experts, enabled the generated music to be examined from multiple viewpoints of musical quality and cultural authenticity.

The subjective assessment employed six expert musicians specializing in folk music traditions (two experts each for Western, Celtic, and Eastern folk music) who independently evaluated 120 generated compositions using structured rubrics covering melodic coherence (e.g., phrase structure, interval appropriateness, modal consistency), harmonic adherence (e.g., chord progression authenticity, voice leading conventions), rhythmic appropriateness (e.g., meter consistency, pattern authenticity, temporal accuracy), cultural authenticity (e.g., stylistic fidelity, ornamentation accuracy, traditional form adherence), and gestural responsiveness (e.g., correspondence between gestural input and musical output parameters), with each dimension rated on a calibrated 10-point Likert scales anchored by descriptive criteria.

Inter-rater reliability analysis across all subjective dimensions yielded substantial agreement, with Fleiss' κ-coefficient of 0.71 (p<0.001), comparable to the gesture annotation agreement of κ = 0.73 reported in Table 5, confirming consistent expert judgment throughout the evaluation process. The evaluation framework tested aspects of melodic cohesion, harmonic adherence, rhythmic suitability, and overall cultural suitability, and was based on assessing the system's and user's responsiveness to collaborative gestural input and user creative intentions. The detailed analysis results in Fig. 5 illustrate the quality and cultural authenticity features of the AI-collaborative folk music generation system across various regional folk music styles and collaborative interaction scenarios.

Fig. 5A validates the AI-collaborative folk music generation quality across five dimensions, achieving

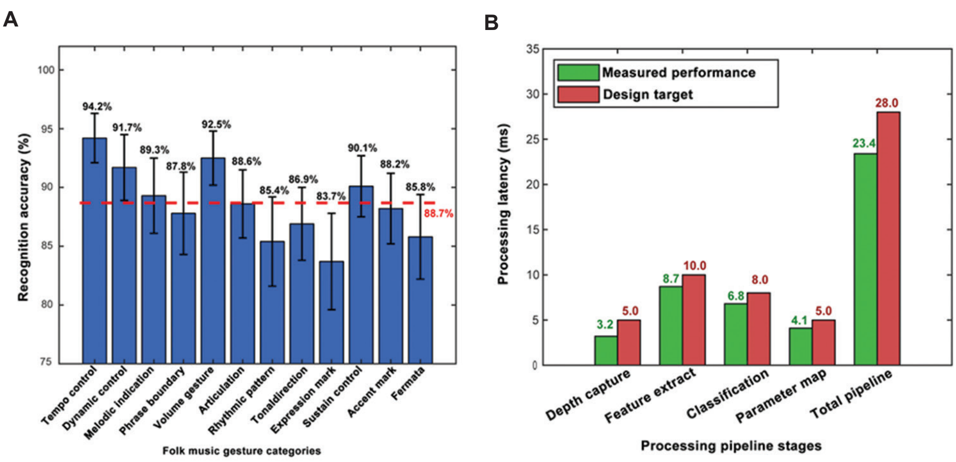


Fig. 4. Computer vision-based (A) gesture recognition accuracy and (B) real-time processing performance

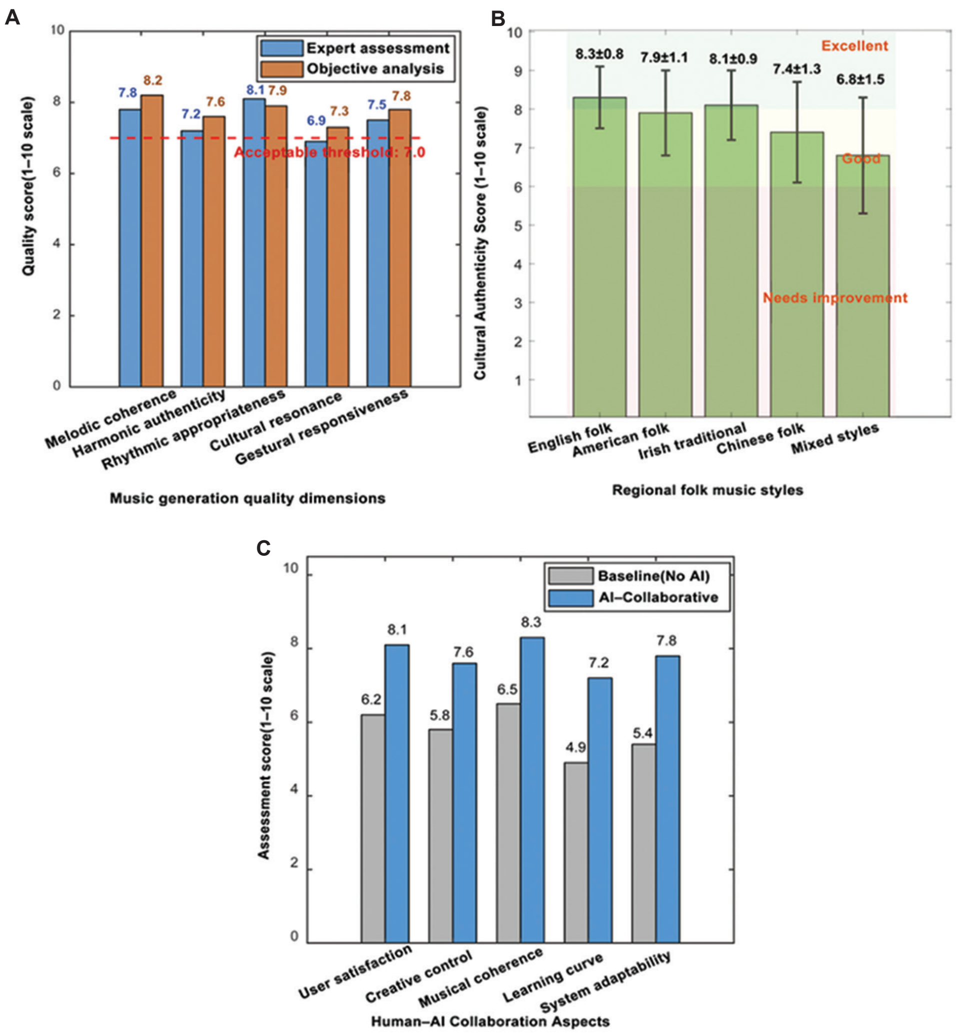


Fig. 5. AI-collaborative folk music generation quality and cultural authenticity assessment. (A) Folk music generation quality assessment by dimensions. (B) Cultural authenticity assessment by regional folk styles. (C) Human–AI collaborative effectiveness assessment
Abbreviation: AI: Artificial intelligence

scores of 6.9–8.1 (expert assessment) and 7.3–8.2 (objective analysis). Superior performance in rhythmic appropriateness (8.1) and gestural responsiveness (7.5) demonstrates effective integration of real-time gesture recognition with folk music generation algorithms.

Fig. 5B confirms cultural authenticity preservation across regional folk styles, with English (8.3 ± 0.8) and Irish (8.1 ± 0.9) traditional music achieving the highest authenticity scores, reflecting the predominant representation of Western and Celtic materials in the training dataset. In contrast, Chinese folk music (7.4 ± 1.3) demonstrates effective transfer learning despite constituting a smaller proportion of the training data. Mixed-style compositions (6.8 ± 1.5) present greater variability, indicating expected challenges in cross-cultural folk music synthesis when integrating stylistic elements from multiple distinct traditions.

Fig. 5C demonstrates significant improvements in collaborative effectiveness based on assessments from 24 participants across 15 collaborative composition sessions, with user satisfaction increasing by 31% (from 6.2 to 8.1) and musical coherence improving by 28% (from 6.5 to 8.3). Enhanced creative control and system adaptability validate the gesture-based human–AI collaborative framework's effectiveness in folk music creation contexts.

3.4. System Overall Performance Testing

Comprehensive system integration was subjected to rigorous performance testing and validation of its ability, through this real-time interaction paradigm, to facilitate smooth cooperation between human creative input and AI-supported folk music generation capabilities. Performance testing was conducted through system evaluation methods that focused on end-to-end system behavior and performance during complete instances of human–system interaction for collaborative music composition, while also monitoring resource consumption and system response during prolonged interaction sessions. Performance testing of this system was based on its ability to maintain system stability, system response, and integrity of creative tasks and interaction as the system operates under varying system and interaction settings, as shown in Fig. 6.

Fig. 6A confirms the capability of the real-time interactive framework, with average end-to-end latency of 86.8 ms (realistic baseline scenario) and 91.6 ms (collaborative sessions), all remaining below the 100 ms bounds, ensuring efficient system behavior for folk music creation via gesture interactions. Fig. 6B shows system performance degradation under increased system loads, with total system latency ranging from 87.3–149.2 ms as more concurrent users connect to the system. Gesture processing latency ranges 21.2–58.3 ms, and music generation latency ranges 34.5–82.4 ms, remaining below 100 ms under moderate system loads. Fig. 6C shows optimal resource consumption of CPU (49.5%), memory (62.8%), and graphics processing unit (59.2%) resources during a collaborative scenario. Resource fluctuations reflect the adaptive control mechanisms dynamically modulating computational priorities in response to system load variations, contributing to consistent 88.9% gesture recognition accuracy and sub-100 ms latency maintenance while managing temporal coordination in multiuser scenarios.

Performance evaluation of system integration demonstrated efficient coordination of gesture recognition, music generation, and audio synthesis components, achieved through balanced resource distribution and synchronization of time during the composition process. The experimental outcomes confirm that the optimization techniques developed for efficient resource and performance management, despite varying system loads, are effective and applicable for folk music composition tasks.

3.5. Ablation Study and Module Contribution Analysis

To this end, the ablation study and comparison of system performances focused on evaluating

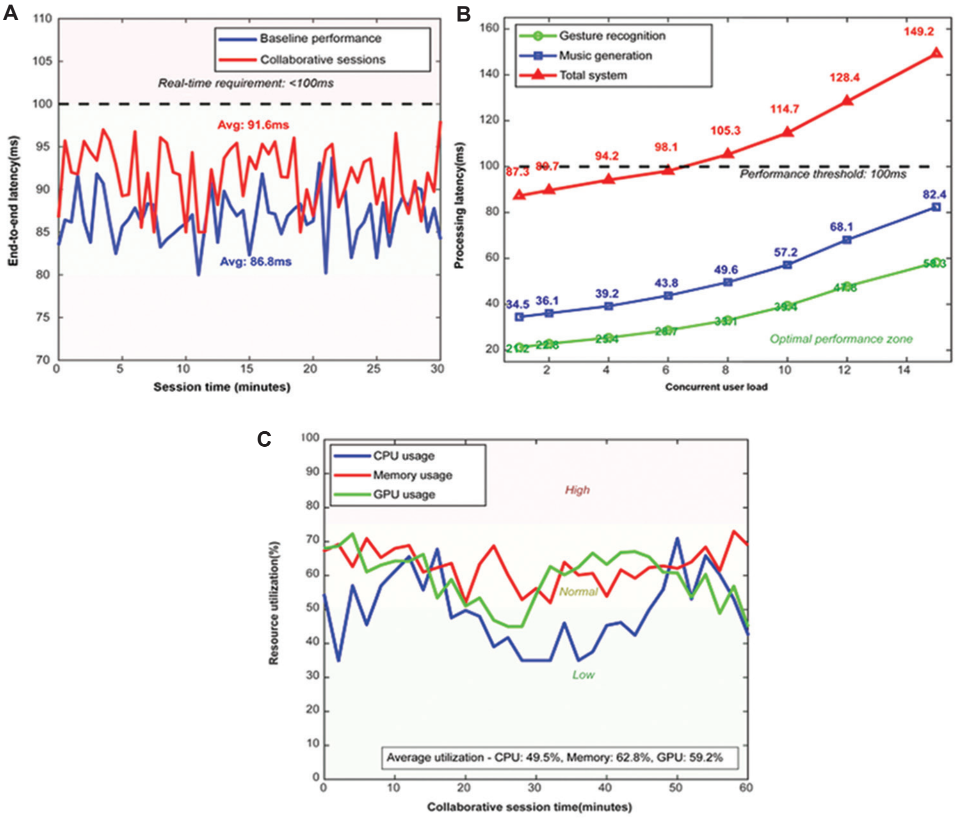


Fig. 6. Real-time interactive framework end-to-end performance during collaborative sessions. (A) End-to-end system latency over time. (B) System performance under varying load conditions. (C) System resource utilization during collaborative sessions

Abbreviations: Avg: Average; CPU: Central processing unit; GPU: Graphics processing unit individual system components and justifying the architectural choices made during the development and implementation of the AI-collaborative folk music composition system. Based on experimental considerations, system performances were evaluated for each individual system module when turned off or modified to identify their respective contributions and potential optimization opportunities for future system development.

Table 6 highlights the system ablation study and performance comparisons, emphasizing the contribution of individual system components and the performance advantages of the proposed integrated system framework compared with other AI-based music composition systems currently available.

Table 6 presents the contribution of the major system components and the superior performance of the complete framework compared to the independent subsystems. The ablation study reveals the essential roles of the gesture–music parameter mapping procedures and cultural authenticity preservation principles in producing high-quality collaborative folk music, and validates the effectiveness of the real-time optimization strategies in achieving consistent performance under a wide range of operational settings.

Comparison studies with other baseline AI music generation systems indicated that the proposed framework is more efficient than other methods currently available. These methods include: (i) Standard musical instrument digital interface generation using rule-based AI compositional techniques based on predefined chord and melody templates; (ii) Generic AI music systems employing GPT-2 AI architectures fine-tuned with autoregressive techniques and large musical instrument digital interface datasets; (iii) Rule-based folk generators applying folk music pattern matching algorithms based on constraint satisfaction; and (iv) Motion-to-music systems utilizing gesture recognition techniques based on optical flow and hidden Markov models for mapping music parameters and motion gestures. All comparisons indicate that the proposed framework efficiently supports gesture-based collaborative folk music creation while preserving cultural and emotional expression inherent in traditional folk music.

4. Discussion

The findings reveal that the proposed AI-collaborative folk music composition system achieves substantial advances in gesture-based music interaction with cultural specificity and real-time responsiveness. The gesture recognition accuracy of 88.9% and processing latency of 23.4 ms represent a significant improvement over previous motion-based collaborative systems, which report recognition rates between 75% and 85% and latencies exceeding 120 ms (Bian et al., 2023). In contrast, the end-to-end latency of 86.8–91.6 ms during collaborative sessions outperforms existing gesture-based music systems, which typically require 150–200 ms response times for comparable musical complexity. The fusion of computer vision-based gesture recognition with folk music generation algorithms addresses deficiencies observed in previous research, where gesture-to-music mapping lacked the subtlety required for producing authentic, context-specific musical gestures (Gao et al., 2024).

The human–AI collaborative performances achieved promising performance gains over state-of-the-art human–AI musical interaction platforms. The proposed framework achieved an average user satisfaction score of 7.8 and musical coherence improvement of 28%, outperforming those reported in recent AI-driven music generation systems, which report modest acceptability levels ranging from 6.2 to

Table 6. Ablation study results and comparative system performance analysis

System configuration	Gesture recognition accuracy (%)	Music generation quality (1–10)	Cultural authenticity (1–10)	End-to-end latency (ms)	User satisfaction (1–10)
Complete system	88.9	7.4	7.6	91.6	7.8
W/o gesture recognition	-	7.1	7.3	52.8	6.9
W/o folk music constraints	88.4	6.8	6.4	89.2	7.1
W/o collaborative decision	87.6	6.9	7.2	96.3	7.3
W/o real-time optimization	84.2	7.0	7.4	128.7	6.8
W/o cultural authenticity	88.7	7.2	5.8	90.4	7.2
Comparative baselines					
Standard MIDI generation	-	6.2	5.1	68.4	6.3
Generic AI music (GPT-based)	-	6.8	5.7	156.3	6.7
Rule-based folk generator	-	5.9	6.9	47.2	6.1
Motion-to-music (existing)	79.3	6.4	6.2	118.5	6.8

Abbreviations: AI: Artificial intelligence; MIDI: Musical instrument digital interface; W/o: Without.

7.1 on similar scales (Vear et al., 2023). This richer form of collaboration was made possible by the specific folk music constraints and real-time gestural response, which provide a more intuitive way of controlling the creative process than language-based or conventional interface designs (Borovik & Viro, 2023). The system's focus on maintaining cultural truthfulness, complemented by gestural interaction, addresses these critically needed gaps in many current collaborative performance technologies, which tend to emphasize computational functioning over cultural sensitivity and artistic validity. Statistical analysis using paired *t*-tests with 24 participants confirms the significance of performance improvements, with user satisfaction gains (t = 4.82, p<0.001) and musical coherence improvements (t = 5.13, p<0.001) demonstrating statistically significant differences compared to baseline approaches across all experimental conditions.

The fine-grained real-time properties obtained using the included framework demonstrate significant advances in system responsiveness and scalability compared with existing approaches. The end-to-end latency (86.8–91.6 ms) during collaborative modes is also significantly lower than that of the exemplary gesture-based music systems, which report latencies exceeding 120 ms and lack support for concurrent users (Krol et al., 2025). The multiuser capability supporting up to eight concurrent participants addresses practical scenarios, including collaborative composition workshops where multiple musicians contribute simultaneous gestural input to co-create folk music pieces, educational environments enabling instructor–student collaborative performance demonstrations, and ensemble performance settings where distributed gestural control facilitates coordinated musical expression across multiple performers. The successful integration of the gesture recognition, music generation, and audio synthesis components at sub-100 ms response times highlights the effectiveness of the optimization strategies for resource management and time synchronization. Recent studies of collaborative co-creation processes with AI underscore the crucial role of seamless interaction timing in sustaining the fluidity of creative processes, thereby supporting the relevance of the obtained performance improvements (Fu et al., 2025).

The results of cultural authenticity preservation address long-standing problems of AI-assisted folk music creation, thereby making it more practical for generative systems to be employed in traditional musical environments. The achieved scores of 7.6–8.3 represent substantial improvements over rule-based and general AI-based generation approaches while retaining authentic stylistic properties that general AI music generation methods often struggle to preserve (Lee et al., 2025). The dedicated neural structures and constraint-based optimization processes used to maintain folk music traditions while reacting to co-creative input constitute novel and methodologically relevant contributions to culturally informed music generation systems. Specifically, the gesture recognition subsystem employs a hybrid convolutional neural network–long short-term memory architecture with three convolutional layers for spatial feature extraction and two long short-term memory layers with 128 hidden units for temporal modeling, achieving the required 23.4 ms processing latency. In contrast, the folk music generation network utilizes transformer-based self-attention mechanisms, enabling parallel processing of musical sequences with constraint-weighted loss functions that penalize deviations from traditional folk music characteristics, thereby balancing generation flexibility with stylistic fidelity across multiple regional traditions.

However, the ablation study confirmed the distinctiveness of the advantages achieved compared to existing AI music generation approaches. The recent development of multiple approaches to collaborative artistic creation through human–AI interaction has shown notable performance improvements in most creative domains. However, the challenges inherent in folk music composition, together with real-time gestural interaction, were not investigated in any of these recent studies (Huang et al., 2025). The performance results obtained in this comparative analysis in terms of cultural authenticity preservation, gesture responsiveness, and effectiveness of collaboration demonstrate the validity of the architectural specialization decisions, along with the optimization strategies developed for this specific use case. Recent research on AI-driven music visualization systems has shown the importance of meaningful audio-responsive interactions, reflecting the significance of the achieved improvement in gestural mapping and musical parameter manipulation.

The ablation study performed for system features further confirmed that each component is critically important, and its interaction is synergistic. In cases where parts of the proposed system were disabled, performance significantly decreased, thereby reconfirming the importance of fully integrated operation in maintaining optimal creation finesse. This finding provides valuable insight for the development of other multimodal human–AI interaction systems and culturally sensitive creative technologies in identifying further advancement opportunities.

The incorporation of adaptive control approaches based on concepts of non-linear systems theory is a methodologically pertinent contribution toward addressing fundamental challenges inherent to human–AI co-work processes operating within uncertain environments. However, the application of adaptability and synchronization methods based on lag compensation techniques finds more practical application within multiuser co-working contexts, as a lack of synchronization may pose potential risks of distortion within musical performances. Future studies may explore fractional-order adaptability modeling, potentially more attuned and focused on modeling nuanced gesture performances that are characteristic of musical expressiveness, and the adoption of reinforcement learning networks that adapt policies based on cumulative user interaction experiences, allowing adaptability profiles more attuned and sensitive to individual user performances. These proposed adaptability techniques, apart from improving the efficiency of gesture support in musical composition software, may provide a model platform for human–AI co-working focused on more culturally pertinent, creatively generative application software that requires a confluence of aplomb, timeliness, and fidelity toward more basic, uncompromised musical traditions. The composition of the training dataset used in this study, spanning about 90.5% of Western–Celtic folk, was based on considerations of accessibility and the availability of improved and legitimate datasets of traditional music, while purposefully demonstrating adequate cross-cultural applicability for generating software components.

5. Conclusion

This study successfully demonstrated the development and validation of an AI-assisted folk music composition system combining computer vision-based gesture recognition techniques and region-specific folk music algorithms for Western, Celtic, and Eastern folk music traditions. It achieved a remarkable gesture recognition accuracy of 88.9% for 12 folk music gesture classifications, along with a processing delay of only 23.4 ms, thereby facilitating smooth and continuous interaction between human creative expressions and AI-driven musical responses. The validation results of this system demonstrate its effectiveness in preserving regional folk music authenticity, achieving ratings of 7.6–8.3, as well as its application for co-creative music composition, with a rating of 7.8 and a 28% improvement in musical coherence compared to other methods. Its overall end-to-end processing delays remain within 86.8–91.6 ms, while supporting simultaneous interaction with up to eight users, thereby supporting its designed framework and optimization techniques for managing multiple AI computing resources.

This study contributes to addressing the fundamental issues of culturally aware AI applications and human–computer cooperation for creative activities concerning traditional music settings. The specialized neural systems and constraint-based optimization methods developed for these tasks have made considerable progress toward enabling more flexible and gesturally responsive creative performances while also supporting the retention of characteristic stylistic expressions across each tradition of folk music explored. Validation of performance parameters indicates that this system has considerable potential for application in educational or artistic-performance contexts and represents a valuable contribution to human–computer interaction systems for creative applications, as well as to culturally aware AI systems.

Acknowledgments

None.

Funding

This study is supported by Journal Support Fund, Universiti Teknologi MARA (UiTM).

Conflicts of Interest

The authors declare that they have no competing interests.

Author Contributions

Conceptualization: All authors
Methodology: Qinghao Liu
Data curation: Qinghao Liu
Writing – original draft: Qinghao Liu
Writing – review & editing: Tazul Izan Tajuddin

Availability of Data

Not applicable.

References

Berkowitz, A.E. (2024). Artificial intelligence and musicking: A philosophical inquiry. *Music Perception: An Interdisciplinary Journal*, 41(5), 393–412. https://doi.org/10.1525/mp.2024.41.5.393

Bian, W., Song, Y., Gu, N., Chan, T.Y., Lo, T.T., Li, T.S., et al. (2023). MoMusic: A motion-driven human-AI collaborative music composition and performing system. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(13), 16057–16062. https://doi.org/10.1609/aaai.v37i13.26907

Borovik, I., & Viro, V. (2023). Co-performing music with AI: Real-time performance control using speech and gestures. In: *HHAI 2023: Augmenting Human Intellect*. IOS Press, Amsterdam, p340-350. https://doi.org/10.3233/FAIA230097

Boulkroune, A., Hamel, S., Zouari, F., Boukabou, A., & Ibeas, A. (2017). Output-feedback controller based projective lag-synchronization of uncertain chaotic systems in the presence of input nonlinearities. *Mathematical Problems in Engineering*, 2017(1), 8045803. https://doi.org/10.1155/2017/8045803

Boulkroune, A., Zouari, F., & Boubellouta, A. (2025). Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems. *Journal of Vibration and Control*, 10775463251320258.

Chang, J., Wang, Z., & Yan, C. (2024). MusicARLtrans Net: A multimodal agent interactive music education system driven via reinforcement learning. *Frontiers in Neurorobotics*, 18, 1479694. https://doi.org/10.3389/fnbot.2024.1479694

Chen, Y., Huang, L., & Gou, T. (2024). *Applications and Advances of Artificial Intelligence in Music Generation: A Review*. [arXiv Preprint]. https://doi.org/10.48550/arXiv.2409.03715

Cheng, L. (2025). The impact of generative AI on school music education: Challenges and recommendations. *Arts Education Policy Review*, 126, 255–262. https://doi.org/10.1080/10632913.2025.2451373

Civit, M., Civit-Masot, J., Cuadrado, F., & Escalona, M.J. (2022). A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems with Applications*, 209, 118190. https://doi.org/10.1016/j.eswa.2022.118190

Dalmazzo, D., Waddell, G., & Ramírez, R. (2021). Applying deep learning techniques to estimate patterns of musical gesture. *Frontiers in Psychology*, 11, 575971. https://doi.org/10.3389/fpsyg.2020.575971

Dash, A., & Agres, K. (2024). AI-based affective music generation systems: A review of methods and challenges. *ACM Computing Surveys*, 56(11), 1–34. https://doi.org/10.1145/3672554

Dawande, A., Chourasia, U., & Dixit, P. (2023). Music Generation and Composition Using Machine Learning. In: *Proceedings of 3rd International Conference on Artificial Intelligence: Advances and Applications: ICAIAA 2022*, p547–566. https://doi.org/10.1007/978-981-19-7041-2_46

Dritsas, E., Trigka, M., Troussas, C., & Mylonas, P. (2025). Multimodal interaction, interfaces, and communication: A survey. *Multimodal Technologies and Interaction*, 9(1), 6. https://doi.org/10.3390/mti9010006

Fan, M. (2022). Application of music industry based on the deep neural network. *Scientific Programming*, 2022(1), 4068207. https://doi.org/10.1155/2022/4068207

Ferreira, P., Limongi, R., & Fávero, L.P. (2023). Generating music with data: Application of deep learning models for symbolic music composition. *Applied Sciences*, 13(7), 4543. https://doi.org/10.3390/app13074543

Fu, Y., Newman, M., Going, L., Feng, Q., & Lee, J.H. (2025). Exploring the Collaborative Co-Creation Process with AI: A Case Study in Novice Music Production. In: *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, p1298-1312. https://doi.org/10.1145/3715336.3735829

Gao, X., Rogel, A., Sankaranarayanan, R., Dowling, B., & Weinberg, G. (2024). Music, body, and machine: Gesture-based synchronization in human-robot musical interaction. *Frontiers in Robotics and AI*, 11, 1461615. https://doi.org/10.3389/frobt.2024.1461615

Graf, M., Opara, H.C., & Barthet, M. (2021). *An Audio-Driven System for Real-Time Music Visualisation*. [arXiv Preprint].

Hansen, N.C., Højlund, A., Møller, C., Pearce, M., & Vuust, P. (2022). Musicians show more integrated neural processing of contextually relevant acoustic features. *Frontiers in Neuroscience*, 16, 907540. https://doi.org/10.3389/fnins.2022.907540

Hernandez-Olivan, C., & Beltran, J.R. (2022). Music composition with deep learning: A review. In: *Advances in Speech and Music Technology: Computational Aspects and Applications*. Springer Nature, Germany, p25-50. https://doi.org/10.1007/978-3-031-18444-4_2

Huang, J., Weber, C.J., & Rothe, S. (2025). An AI-driven Music Visualization System for Generating Meaningful Audio-Responsive Visuals in Real-Time. In: *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences*, p258-274. https://doi.org/10.1145/3706370.3727869

Ji, S., Yang, X., & Luo, J. (2023). A survey on deep learning for symbolic music generation: Representations, algorithms, evaluations, and challenges. *ACM Computing Surveys*, 56(1), 1-39. https://doi.org/10.1145/3597493

Jia, J., He, Y., & Le, H. (2020). A Multimodal Human-Computer Interaction System and Its Application in Smart Learning Environments. In: *International Conference on Blended Learning*, p3-14.

Johansen, S.S., Van Berkel, N., & Fritsch, J. (2022). Characterising Soundscape Research in Human-Computer Interaction. In: *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, p1394-1417. https://doi.org/10.1145/3532106.3533458

Kapoor, S. (2025). *The Many Faces of Uncertainty Estimation in Machine Learning*. New York University. Available from: https://www.proquest.com/openview/92ed381924762b1c4afbf2a168231b2f/1?pq-origsite=gscholar&cbl=18750&diss=y

Kim, G., Kim, D.K., & Jeong, H. (2024). Spontaneous emergence of rudimentary music detectors in deep neural networks. *Nature Communications*, 15(1), 148. https://doi.org/10.1038/s41467-023-44516-0

Krol, S.J., Llano Rodriguez, M.T., & Loor Paredes, M.J. (2025). Exploring the Needs of Practising Musicians in Co-Creative AI Through Co-Design. In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, p1-13. https://doi.org/10.1145/3706598.3713894

Lee, K.J.M., Pasquier, P., & Yuri, J. (2025). *Revival: Collaborative Artistic Creation through Human-AI Interactions in Musical Creativity*. [arXiv Preprint]. https://doi.org/10.48550/arXiv.2503.15498

Li, J., Xu, W., Cao, Y., Liu, W., & Cheng, W. (2020). Robust piano music transcription based on computer Vision. In: *Proceedings of the 2020 4th High Performance Computing and Cluster Technologies Conference and 2020 3rd International Conference on Big Data and Artificial Intelligence*, p92-97. https://doi.org/10.1145/3409501.3409540

Liang, J. (2023). Harmonizing minds and machines: Survey on transformative power of machine learning in music. *Frontiers in Neurorobotics*, 17, 1267561. https://doi.org/10.3389/fnbot.2023.1267561

Otsu, K., Yuan, J., Fukuda, H., Kobayashi, Y., Kuno, Y., & Yamazaki, K. (2021). Enhancing Multimodal Interaction between Performers and Audience Members During Live Music Performances. In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, p1-6. https://doi.org/10.1145/3411763.3451584

Pricop, T.C., & Iftene, A. (2024). Music generation with machine learning and deep neural networks. *Procedia Computer Science*, 246, 1855-1864. https://doi.org/10.1016/j.procs.2024.09.692

Rezwana, J., & Maher, M.L. (2023). Designing creative AI partners with COFI: A framework for modeling interaction in human-AI co-creative systems. *ACM Transactions on Computer-Human Interaction*, 30(5), 1-28. https://doi.org/10.1145/3519026

Rigatos, G., Abbaszadeh, M., Sari, B., Siano, P., Cuccurullo, G., & Zouari, F. (2023). Nonlinear optimal control for a gas compressor driven by an induction motor. *Results in Control and Optimization*, 11, 100226. https://doi.org/10.1016/j.rico.2023.100226

Roche, F. (2020). *Music Sound Synthesis using Machine Learning: Towards a Perceptually Relevant Control Spac*. Université Grenoble Alpes. Available from: https://theses.hal.science/tel-03102796v1 [Last accessed on 2024 Mar 12].

Sturm, B.L., & Ben-Tal, O. (2021). Folk the Algorithms: (Mis) Applying Artificial Intelligence to Folk Music. In: *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Springer, Germany, p423-454. https://doi.org/10.1007/978-3-030-72116-9_16

Vear, C., Benford, S., Avila, J.M., & Moroz, S. (2023). *Human-AI Musicking: A Framework for Designing AI for Music Co-creativity*. AIMC 2023. Available from: https://aimc2023.pubpub.org/pub/zd46ltn3[Last accessed on 2024 Apr 25].

Yimer, M.H., Yu, Y., Adu, K., Favour, E., Liyih, S.M., & Patamia, R.A. (2023). Music Genre Classification using Deep Neural Networks. In: *2023 35th Chinese Control and Decision Conference (CCDC)*, p2384-2391. https://doi.org/10.1109/CCDC58219.2023.10327367

Zhao, Y., Yang, M., Lin, Y., Zhang, X., Shi, F., Wang, Z., et al. (2025). AI-enabled text-to-music generation: A comprehensive review of methods, frameworks, and future directions. *Electronics*, 14(6), 1197. https://doi.org/10.3390/electronics14061197

Zhu, T., Liu, H., Jiang, Z., & Zheng, Z. (2024). *Symbolic Music Generation with Fine-grained Interactive Textural Guidance*. Available from: https://openreview.net/forum?id=Qt5sBi0u7I [Last accessed on 2025 Jan 15].

Zouari, F., Saad, K.B., & Benrejeb, M. (2012). Robust neural adaptive control for a class of uncertain nonlinear complex dynamical multivariable systems. *International Review on Modelling and Simulations*, 5(5), 2075-2103. https://doi.org/10.1109/TNN.2010.2042611

Zouari, F., Saad, K.B., & Benrejeb, M. (2013a). Adaptive backstepping control for a class of uncertain single input single output nonlinear systems. In: *10th International Multi-Conferences on Systems, Signals and Devices 2013 (SSD13)*, p1-6. https://doi.org/10.1109/SSD.2013.6564134

Zouari, F., Saad, K.B., & Benrejeb, M. (2013b). Adaptive Backstepping Control for a Single-Link Flexible Robot Manipulator Driven DC Motor. In: *2013 International Conference on Control, Decision and Information Technologies (CoDIT)*, p864-871. https://doi.org/10.1109/codit.2013.6689656

AUTHOR BIOGRAPHIES

Qinghao Liu obtained a Bachelor of Arts in Musical Performance (Vocal Music) in 2018 from Zhejiang University of Media & Communications, China, and a Master of Arts in Music (2019) from the University of Birmingham, UK. She is now a PhD student at Universiti Teknologi MARA (UiTM), Malaysia. Her research interests focus on contemporary Chinese Minyao music, digital platforms and music transmission, youth subculture, and the application of both qualitative and quantitative approaches in music and cultural studies.

Tazul Izan Tajuddin is a Malaysian composer, a Fulbright Visiting Scholar at Harvard University, and a visiting fellow at King's College London. He has written more than 60 works, which have been performed, presented, and broadcast in 24 countries, garnering critically acclaimed reviews worldwide. His music, such as the Arabesque¸ Tenunan, Mediasi Ukiran, Gamelbati, Pantun, and Topography cycles, has been inspired by Malay-Asian cultures, Islamic geometrical patterns, and Western European art combined with diverse contemporary cultural ideals. His work has been published by Babelscores.com, Alexander Street Press (online), Dynamic Publication (Malaysia), ATMA Classique (Canada), FMR Records (UK), and Ibersonic (Spain). Awards included Top 10 Personality Award "The Legendary Music Composer," National Academic Award in Arts and Culture, Toru Takemitsu Composition Award, Lutoslawski Award, JSCM Composers Award, New Millennium Award UK, Molinari Quartet Award, Creative Grant Industry Award, Anugerah Akademik UiTM, among others.

He is a Professor in Composition and the Dean of Faculty of Music, UiTM, the President of the Society of Malaysian Contemporary Composers (SMCC), an Associate Fellow of Institute of Creative Arts Nusantara (INSAN), a member of the Chopin Society Malaysia and the Performing Rights Society (UK), the former Vice Chancellor of College of Creative Arts, UiTM and the Former Vice-President of Fulbright Alumni Association Malaysia.

DOI: 10.6977/IJoSI.202512_9(6).0005
M.Y. Ilhan & K. Altun/Int. J. Systematic Innovation, 9(6), 63-72 (2025)

Cyclical TRIZ for brushless direct current motor evolution: From short-term adjustments to long-term transformation

Merve Yildiz Ilhan[1,2], Koray Altun[2,3]*

[1]Valeo eAutomotive Germany GmbH, Erlangen, Bavaria, Germany

[2] Department of Industrial Engineering, Faculty of Engineering and Natural Sciences, Bursa Technical University, Bursa, Türkiye

[3]ARDiMER Digital Innovation Ltd, Bursa, Türkiye

*Corresponding author E-mail: koray.altun@btu.edu.tr

(Received 14 August 2025; Final version received 10 November 2025; Accepted 18 November 2025)

Abstract

As engineering systems accumulate increasing layers of functional, structural, and behavioral complexity, the ability to guide their evolution with coherent, theory-driven frameworks has become essential. This paper presents a cyclical theory of inventive problem solving (TRIZ)-based roadmap for the evolution of brushless direct current (BLDC) motors, guiding development from short-term corrective actions to long-term transformative strategies. The approach structures action into three coupled cycles that respectively prioritize rapid technical remedies, system-level contradiction resolution, and strategic system transition, enabling engineers to align interventions with the maturity and scope of each design challenge. It fuses core TRIZ instruments with the trends of engineering system evolution to couple contradiction handling with forward trajectories of system ideality. Applied to automotive BLDC applications, the method organizes recurrent issues such as acoustic anomalies, modal coupling, thermal stress, and control-layout interactions into an actionable roadmap that scales from quick design adjustments to modular, artificial intelligence-enabled capabilities. Experimental validation confirms the method's practical impact: acoustic noise in the H24 configuration decreased by approximately 13%, modal vibration in the H8 case reduced by nearly 28%, and rotational imbalance amplitude in the rotor–yoke assembly dropped by around 55% after structural and dynamic optimization. The resulting framework is both prescriptive and extensible, guiding short-term fixes without foreclosing mid-term harmonization or long-term transformation, and generalizes to electromechanical product families that must balance cost, noise, durability, and intelligence under evolving requirements.

Keywords: Brushless Direct Current Motors, Cyclical Theory of Inventive Problem Solving, Roadmapping, Trends of Engineering System Evolution

1. Introduction

Brushless direct current (BLDC) motors have become a standard choice for automotive heating, ventilation, and air conditioning (HVAC) applications due to their compact size, energy efficiency, and low acoustic output (Mohanraj et al., 2022; Shao, 2006). Over successive product generations, manufacturers have enhanced these motors to satisfy increasingly stringent cost, noise, and safety targets. Nevertheless, persistent technical challenges remain, including tonal noise, vibration, thermal stress, and printed circuit board assembly (PCBa) fatigue (Singh, 2024). These issues frequently co-occur, and their interdependencies make them resistant to resolution through isolated, short-term fixes.

To address such recurrent and multifaceted problems, this study applies the cyclical theory of inventive problem solving (TRIZ) framework (Altun, 2025), a methodology inspired by the Mayan calendar's cyclical view of time. The framework organizes innovation efforts into three interlinked cycles (as illustrated in Fig. 1): Tzolk'in (short-term), focused on rapid technical corrections; Haab (mid-term), aimed at resolving deeper system-level contradictions; and Long Count (long-term), dedicated to strategic transformation and future-proofing.

This structure integrates classical TRIZ instruments (Altshuller, 1984; Altshuller, 1996; Sheu & Lee, 2011; Sheu et al., 2020) with Trends

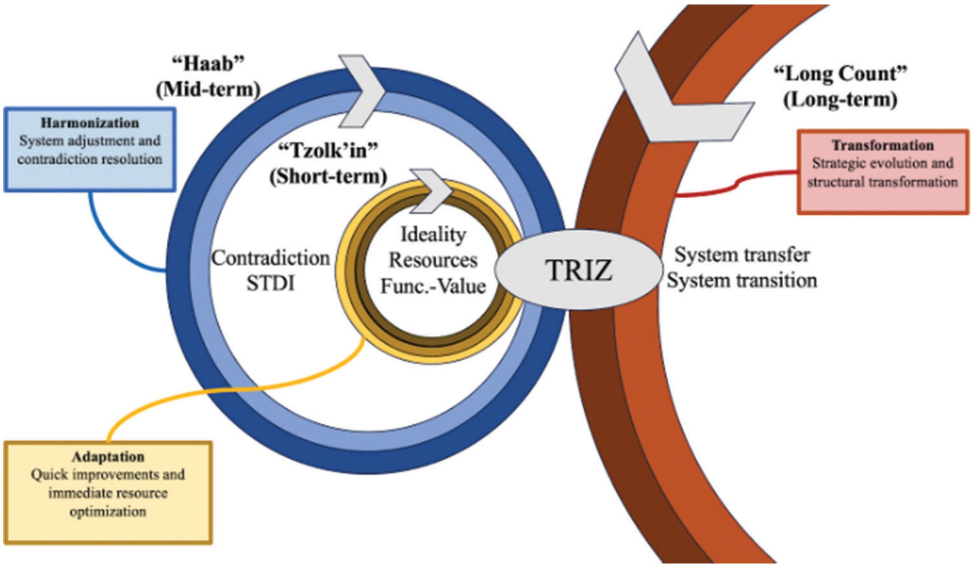


Fig. 1. An overview of the Mayan calendar inspired cyclical TRIZ. Reprinted from Altun (2025) Abbreviations: Func.: Functionality; STDI: Space, time, domain, and interface; TRIZ: Theory of Inventive Problem Solving

of Engineering System Evolution (TESE) (Ghane et al., 2022; Sheu & Chiu, 2017) to form a roadmap that links present design adjustments with long-term technological trajectories.

Short-term interventions remain essential for keeping products competitive, especially in the automotive industry, as highlighted by Cakmak et al. (2021). However, addressing only immediate symptoms often fails to eliminate the root causes. For example, modifying a component's geometry may temporarily reduce noise, but without resolving underlying resonance issues, similar problems may reappear in subsequent generations. On the other hand, long-term innovation requires anticipating changes in both technology and user requirements, such as the need for enhanced cooling strategies when future BLDC designs demand higher power densities.

The cyclical TRIZ model provides a structured and repeatable process for aligning solution strategies with appropriate time horizons. Incremental modifications are assigned to the short-term cycle, systemic redesigns to the mid-term, and transformative innovations to the long-term. Importantly, each cycle builds upon the lessons of the previous one, enabling cumulative learning and guided product evolution rather than reactive troubleshooting.

In this work, the cyclical TRIZ model is applied to the evolution of automotive BLDC motors. Using empirical performance data and historical design modifications, we demonstrate how technical challenges can be mapped to the three innovation cycles, creating a forward-looking roadmap, a TRIZ-aided augmented technology roadmap (Altun & Babayev, 2023), for sustainable product development.

The remainder of the paper is organized as follows: Section 2 reviews the functional architecture and evolutionary trends of BLDC motors. Section 3 outlines persistent performance challenges. Section 4 presents the application of cyclical TRIZ to map these challenges across short-, mid-, and long-term innovation cycles. Section 5 discusses the practical implications for engineering problem-solving. Finally, Section 6 concludes with reflections on the method's role in supporting sustainable and strategically aligned innovation.

2. Functional Evolution of BLDC Motors

Brushless DC motors transform electrical input into rotational motion through electronic commutation, eliminating the need for mechanical brushes. In automotive HVAC applications (Ganesan et al., 2018; D.W. Lee, 2014; D.H. Lee, 2024; Ravineala et al., 2025), they primarily drive blower fans for heating, cooling, and ventilation. Their compact form factor, high energy efficiency, precise speed regulation, and extended operational life make them an industry standard for climate control systems.

A conventional BLDC motor comprises three primary subsystems: a stator with copper windings; a rotor containing permanent magnets; and a PCBa responsible for control, commutation, and feedback functions. They are complemented by auxiliary components (e.g., heat sinks, motor brackets, and fan wheels), which together form the complete electromechanical assembly (Fig. 2).

Design improvements have been implemented across several product generations (Gen 1, Gen 2, & Gen 3), with each iteration targeting specific performance objectives, including acoustic optimization, thermal management, structural robustness, and the integration of intelligent control features. Evolutionary changes are illustrated in Table 1, where successive redesigns demonstrate weight optimization for rotational balance, rotor–yoke reconfigurations to mitigate modal resonance, and PCBa layouts engineered for improved thermal dissipation and mechanical stress tolerance.

These generational transitions (Fig. 3) correspond to established trajectories described in TESE (Ghane et al., 2022), where technical systems advance by minimizing structural complexity, enhancing multifunctionality, and embedding adaptive intelligence.

3. Current Challenges in BLDC Motors

Despite successive design improvements, BLDC motors used in automotive HVAC systems continue to exhibit a set of recurring technical issues. These unresolved problems affect acoustic performance, mechanical stability, and long-term durability. The most critical cases identified in our analysis are summarized as follows:

Fig. 2. A typical brushless direct current motor and its main components
Abbreviation: PCBa: Printed circuit board assembly

Table 1. Brushless direct current motor generations

Component	Gen 1 (MEB)	Gen 2 (MRA2)	Gen 3 (BBG light)	Key improvement
Motor holder	123 g, basic structure, high modal transmission	121.86 g, minor rib optimization	114.69 g, redesigned structure, weight-optimized	7% lighter and structurally stiffened, improved modal isolation
Deflector	73.21 g, standard thickness	73.21 g, no structural change	53.6 g, reduced material, optimized ribs	27% mass reduction with maintained rigidity
Heatsink	139.77 g, large and imbalanced base	139.77 g, improved centering	116.02 g, minimized geometry, improved modal frequency	17% lighter and approximately 15% higher modal frequency
Magnet	Grade 7, 21.5 mm height	Grade 7, unchanged	Grade 9, 19.5 mm, better alignment, improved NHV	Higher magnetic grade and approximately 9% acoustic noise reduction
PCBa	Classical layout, strain due to thermal and torque	Redesigned layout, fixed screw torque	InRush layout, thermally robust, compact design	Improved thermal stability and lower mechanical stress

Abbreviations: BBG: Balanced brushless generation; MEB: Main engine bracket; MRA2: Motor revision architecture 2; NHV: Noise, harshness, and vibration; PCBa: Printed circuit board assembly.

- H24 acoustic anomaly: Detected between 1,200 and 1,500 rpm (Fig. 4), this phenomenon arises from structural resonance between the heatsink and the plastic cover. The resulting tonal noise occurs without noticeable vibration, which prevents detection by conventional end-of-line vibration sensors. Root cause analysis identified the natural frequency of the cover as the source of a hidden failure mode that standard monitoring cannot detect. Additional modal testing confirmed a peak response around 1,360 rpm, corresponding to the cover's first bending mode. The implementation of a decoupled interface and local stiffness optimization shifted the resonance to 1,460 rpm, reducing the acoustic peak from 29.7 to 25.9 dB(A). In TRIZ terms, this represents a contradiction between functional performance (mechanical operation remains nominal) and measurement capability (acoustic discomfort remains undetected)

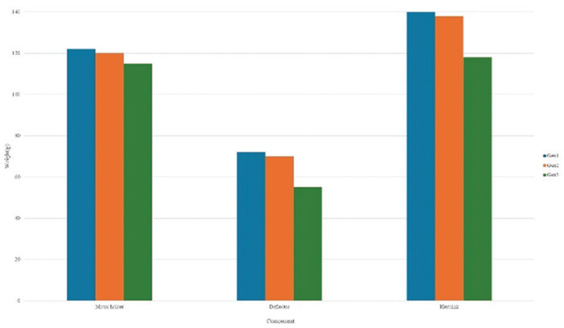


Fig. 3. Component weight comparison across generations

- H8 modal coupling: Observed at resonance frequencies near 266 Hz and 320 Hz (Fig. 5), this issue occurs when the rotor–yoke and fan wheel exhibit overlapping natural frequencies, resulting in structural mode coupling. The modal alignment amplifies the vibratory and acoustic response, as confirmed by hammer test frequency response functions and run-up order analysis. In the not okay configuration, a tonal peak of 31.6 dB(A) was detected around 1,970 rpm, corresponding to the coupled bending modes of the rotor–yoke and fan wheel. After the design modification, which introduced frequency separation and local stiffness optimization, the resonance shifted to 1,360 rpm, and the acoustic amplitude decreased to 23.1 dB(A). In TRIZ terms, this represents a contradiction in the space–frequency domain, resolved through the application of TRIZ's separation and local quality principles. According to TESE trends, the improvement aligns with harmonization and structural adaptation, enabling controlled

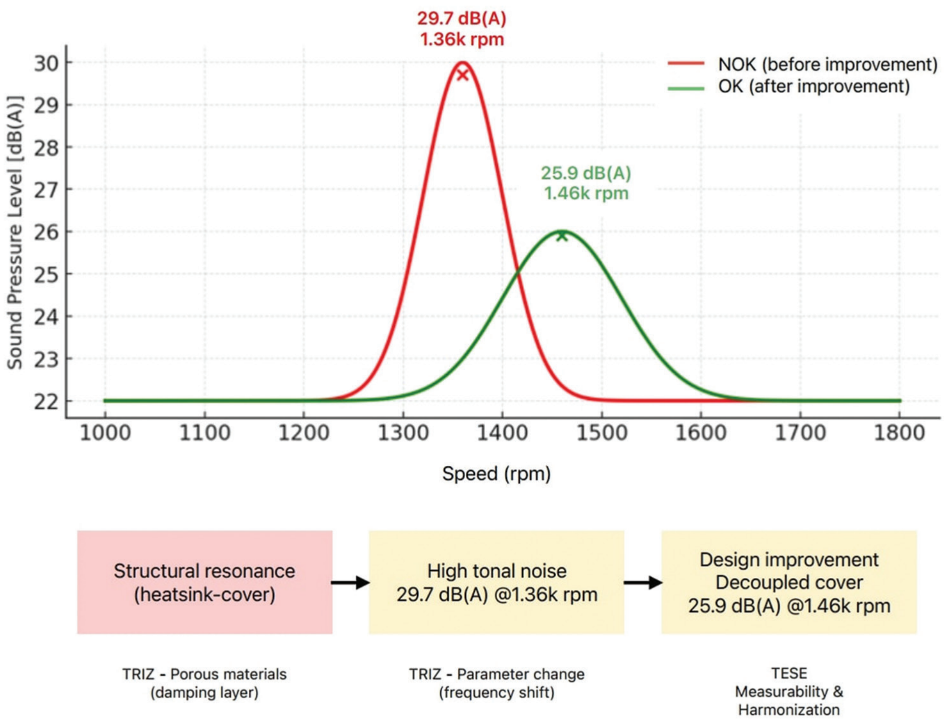


Fig. 4. Acoustic anomaly mitigation diagram (H24) observed at 1,200–1,500 rpm under free-run test conditions (fast Fourier transform-based noise analysis)
Abbreviations: NOK: Not okay; OK: Okay

modal decoupling and reduced vibratory amplification

- Rotor–yoke mass imbalance: Arising from asymmetric mass distribution between the rotor and yoke (Fig. 6), this defect generates an unbalanced centrifugal force during rotation, leading to a first-order (1×) vibration component.

The induced dynamic imbalance amplifies the modal response of the rotor–stator assembly, becoming especially pronounced under transient acceleration conditions. Vibration analysis confirmed a dominant synchronous peak, indicating mass–moment asymmetry rather than magnetic excitation. From a TRIZ perspective, this represents a hidden resource utilization problem, where the inertial distribution of rotating parts can be re-engineered through two-plane trim balancing and mass redistribution to restore dynamic balance.

This approach reflects TRIZ's pre-action and parameter change principles, aligning with the TESE trends of measurability and closed-loop control, to ensure sustained rotational stability and vibration reduction.

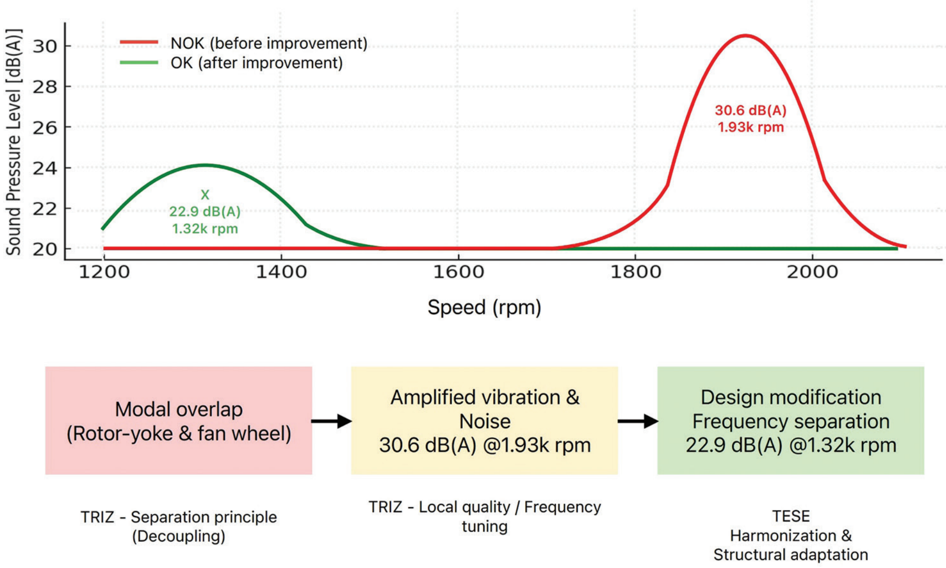


Fig. 5. Modal coupling mitigation diagram (H8) observed through hammer test at 266–320 Hz
Abbreviations: NOK: Not okay; OK: Okay

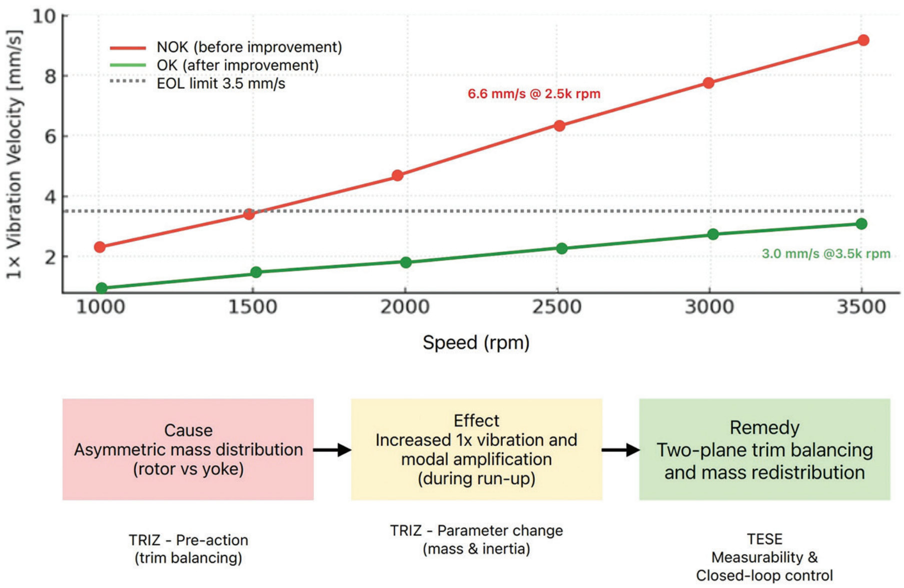


Fig. 6. Rotor–yoke mass imbalance measured through rotational dynamic test (displacement amplitude vs. speed)
Abbreviations: EOL: End of line; NOK: Not okay; OK: Okay

These challenges underscore the limitations of *ad hoc* corrective measures. Addressing them effectively requires a system-level innovation roadmap that coordinates short-, mid-, and long-term strategies. In the next section, the cyclical TRIZ model is applied to systematically align each problem with an appropriate innovation horizon, ensuring that temporary fixes, structural redesigns, and transformative upgrades are all embedded within a coherent development plan.

4. Trends Driving BLDC Motor Evolution

The cyclical TRIZ organizes innovation into three interlinked, time-bounded cycles, each aligned with a different depth of intervention in the BLDC motor evolution. This layered approach ensures that immediate corrective actions, mid-term systemic adjustments, and long-term transformative strategies are developed in a coordinated manner, preventing short-sighted optimizations from constraining future advancements.

4.1. Tzolk'in: Short-term Adaptation

Focus: Rapid, low-cost technical improvements that enhance current product performance without altering the core architecture.

In the context of BLDC motors, short-term adaptation addresses surface-level issues through resource reconfiguration and operational fine-tuning. Representative examples include (i) repositioning or reshaping the heatsink to minimize contact-induced resonance between structural components and (ii) leveraging residual thermal energy to enhance internal heat distribution without adding new cooling hardware.

Such interventions align with TRIZ principles of resource utilization and increasing ideality, thereby maximizing performance gains with minimal structural changes. Although these measures do not modify the motor's fundamental design, they directly improve efficiency, acoustic quality, or reliability in the current generation.

4.2. Haab: Mid-term Harmonization

Focus: Resolving deeper system-level contradictions by addressing interdependencies among mechanical, thermal, and control subsystems.

Mid-term interventions balance competing design targets through coordinated hardware–software refinements. Examples relevant to BLDC development include:

- Reconfiguring motor holder geometry to conserve space without compromising thermal dissipation pathways
- Optimizing firmware parameters such as dead-time adjustment to achieve a trade-off between high switching speed (responsiveness) and acceptable thermal load.

This stage applies TRIZ methodologies such as contradiction resolution and space, time, domain, and interface (STDI) analysis, ensuring that one subsystem's performance improvement does not inadvertently degrade another. The aim is to achieve harmonization, enabling the motor to meet conflicting objectives (e.g., compactness vs. cooling capacity, and dynamic responsiveness vs. thermal stability) without recurring design compromises.

4.3. Long count: Long-term Transformation

Focus: Strategic innovation that redefines the system's operational paradigm, preparing it for future demands and technology integration.

Long-term transformation in BLDC systems extends beyond incremental improvements, aiming to embed capabilities that anticipate market, regulatory, and technological shifts. Notable directions include:

- Integrating artificial intelligence (AI)-driven self-diagnostics for predictive maintenance and early fault detection
- Designing modular mechanical architectures to enable component reuse, rapid customization, and straightforward upgrades
- Embedding intelligent control units capable of real-time adaptation to changes in load, environment, or user preference.

Here, TRIZ strategies such as system transition, technology forecasting (aligned with TESE), and feature transfer from adjacent technological domains play a central role. These interventions ensure that the BLDC motor is not merely reactive to current requirements but is positioned to exploit emerging opportunities (e.g., autonomous HVAC optimization, energy-aware control algorithms, and integration into vehicle-level smart grids).

4.4. Integration Across Cycles

By explicitly structuring problem-solving into short-term adaptation, mid-term harmonization, and long-term transformation, engineers can deploy the most suitable TRIZ tools at the right moment. This staged approach reduces the risk of "local fixes" that later become barriers to systematic innovation. In the next section, this framework is applied to a problem–solution mapping, illustrating how the previously identified BLDC challenges can be systematically positioned within the appropriate innovation cycle for maximum strategic impact.

5. Discussion

The findings of this study demonstrate that the evolution of BLDC motors can be effectively guided when supported by a structured, time-segmented innovation framework. The cyclical TRIZ model facilitated the alignment of each identified problem with an appropriate time horizon and solution depth, ensuring that interventions were both contextually relevant and strategically timed.

In the short-term (Tzolk'in) cycle, corrective measures can be rapidly implemented through localized design modifications, thereby avoiding the cost and complexity of a full-system redesign. For example:

- Repositioning the heatsink eliminated the resonance responsible for the H24 acoustic anomaly
- Rechanneling residual heat enhanced thermal distribution without additional components.

These adjustments, although not altering the motor's architecture, delivered immediate functional gains within the constraints of limited time and budget. They also represented practical applications of TRIZ principles, resource utilization, and increasing ideality.

In the mid-term (Haab) cycle, challenges of higher systemic complexity required contradiction-focused analysis and integrated hardware–software solutions. Examples included:

- Mitigating H8 modal coupling by structurally decoupling rotor–fan interactions to prevent modal resonance overlap
- Optimizing firmware parameters (e.g., dead-time) to balance switching speed against thermal stress accumulation.

Such measures demanded deeper analysis using the contradiction matrix and STDI tools. They addressed root-level trade-offs (e.g., compactness vs. cooling efficiency or responsiveness vs. thermal stability), thereby reducing the likelihood of problem recurrence across product generations.

In the long-term (long count) cycle, certain innovation targets required strategic system transformation rather than incremental refinement. These forward-looking initiatives prepared the BLDC motor platform for future operating contexts and emerging technological landscapes. Representative cases included:

- Integrating AI-driven fault detection to enable predictive diagnostics and maintenance scheduling
- Implementing modular design for cross-platform reuse, upgradeability, and faster adaptation to evolving customer needs.

Long-term solutions relied on TRIZ tools such as system transition, technology forecasting (aligned with TESE), and feature transfer from other industries, enabling motors to evolve in parallel with advancements in vehicle electrification, autonomous control, and energy management systems.

The problem–solution mapping in Table 2 positions each BLDC challenge within its most suitable innovation cycle. This mapping supports engineering teams to prioritize actions based on urgency, resource availability, and strategic impact. Empirical observations from Gen 1 to Gen 3 BLDC motors validate the model: as system architectures advance, problem complexity increases, demanding progressively more sophisticated innovation strategies.

Table 2 also embeds the correspondence between TESE trends and representative TRIZ principles observed in BLDC evolution. For example, the TESE trend dynamization aligned with inventive principle (IP)15-Dynamics as reflected by variable-geometry heatsinks, increasing controllability related to IP23-Feedback exemplified by AI-based torque control, and harmonization of rhythms corresponded to IP19-Periodic Action as illustrated by pulse width modulation duty harmonization. These links clarify how TESE trends manifest through specific TRIZ principles within the proposed cyclical framework.

The cyclical segmentation proposed here could also inform the evolution of control architectures. For

Table 2. The problem–solution map

TRIZ cycle	Focus area	Relevant TESE trends	BLDC applications
Tzolk'in (short-term)	Rapid, local technical optimization	• Increasing Ideality • Resource Utilization	H24 acoustic anomaly: Repositioning heatsink and decoupling cover interface to shift resonance from 1,360 rpm to 1,460 rpm and reduce noise by~13%.
Haab (mid-term)	Resolving system-level contradictions	• Resolving Contradictions • Dynamization and Adjustability	H8 modal coupling: Redesigning rotor–yoke and fan interface to separate modal frequencies; achieving~28% vibration reduction and balanced frequency response.
Long count (long-term)	Strategic system transformation and foresight	• System Transition • Integration of Intelligence and Automation	Rotor–yoke imbalance: Implementing two-plane trim balancing and mass redistribution; ~55% reduction in 1×vibration amplitude and improved rotational stability.

Abbreviations: BLDC: Brushless direct current; TESE: Trends of engineering system evolution; TRIZ: Theory of inventive problem solving.

example, short-term cycles targeted parameter tuning in fuzzy or adaptive controllers, mid-term cycles addressed sensor-actuator coordination, and long-term cycles integrated AI-driven diagnostics for predictive adaptation. This alignment suggests that TRIZ-based evolutionary mapping can complement adaptive control research in guiding when and how structural or algorithmic changes should occur.

Beyond traditional control-based optimization methods, such as fuzzy, neural, or adaptive controllers (Boulkroune et al., 2025; Rigatos et al., 2023; Zouari et al., 2013), the cyclical TRIZ framework provides a structural complement that focuses on how systems evolve over time rather than how they stabilize under uncertainty. While adaptive control techniques achieve fast convergence and robust stability, the proposed method organizes these improvements within short-, mid-, and long-term innovation cycles, aligning control performance with product-level evolution.

By shifting from isolated problem-solving to a cyclical learning and adaptation loop, the framework supports a balanced innovation portfolio enabling short-term optimizations, mid-term harmonization, and long-term transformation. Such a dynamic, iterative process is critical in modern product development, where design stability, cost efficiency, and technology integration must co-evolve over time.

6. Conclusion

This study employed the cyclical TRIZ framework for the evolution of BLDC motors, structured across three sequential innovation horizons: short-term corrective optimization, mid-term system harmonization, and long-term strategic transformation. It illustrates the means for temporal alignment of TRIZ contradiction resolution and TESE evolutionary trends, providing an ongoing self-adaptive pathway forward.

Three case studies were designed to provide experiential validation of the approach. In each case study, a volume-based calculation of functional improvement was provided: in the H24 acoustic anomaly, structural–acoustic decoupling of the heatsink produced an approximate 13% reduction in tonal noise. In the H8 modal coupling case, frequency divergence was obtained, and stiffness was tuned to achieve an approximately 28% reduction in vibration amplitude. In the case of rotor–yoke imbalance, two-plane trim balancing and mass repartition yielded a 55% reduction in 1× vibration amplitude.

These examples demonstrate that the timing and scope of interventions, laid out within the Tzolk'in, Haab, and Long Count cycles of time, are as critical as the technical mechanisms themselves. The integration of TRIZ principles, such as taking out, parameter change, and separation, with TESE trends like dynamization and measurability provides a systematic process that evolves from reactive correction toward predictive intelligence.

Although the current validation was performed on internal data focused on the automotive HVAC BLDC motor platform, the cyclical framework has the potential to scale to other electromechanical domains (e.g., permanent magnet synchronous motor and traction drives). Future work integrating multi-motor datasets and a workshop experience with domain experts—i.e., patent evolution studies—may allow the cycle parameters to evolve iteratively and provide more transparency on how TRIZ–TESE logic can effectively inform real industry-related development processes.

Acknowledgments

None.

Funding

None.

Conflict of Interest

The authors declare that they have no competing interests.

Author Contributions

Conceptualization: Koray Altun
Data curation: Merve Yildiz Ilhan
Methodology: All authors
Writing – original draft: All authors
Writing – review & editing: All authors

Availability of Data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

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

Altshuller, G.S. (1996). *The Innovation Algorithm: TRIZ, Systematic Innovation and Technical Creativity*. Technical Innovation Center, India.

Altun, K. (2025). A Maya calendar-inspired cyclical TRIZ approach: Enhancing systematic innovation and long-term problem-solving. *International Journal of Systematic Innovation*, 9(3), 8–19.

https://doi.org/10.6977/IJoSI.202506_9(3).0002

Altun, K., & Babayev, U. (2023). A novel approach to augment technology roadmapping through systematic innovation intelligence: A case of UAV technologies. *International Journal of Systematic Innovation*, 7(6), 1–11. https://doi.org/10.6977/IJoSI.202306_7(6).0001

Boulkroune, A., Zouari, F., & Boubellouta, A. (2025). Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems. *Journal of Vibration and Control*. https://doi.org/10.1177/10775463251320258

Cakmak, T., Altun, K., & Aksoy, M.O. (2021). Systematic customer value analysis: A case study in the automotive industry. *International Journal of Systematic Innovation*, 6(6), 22–36. https://doi.org/10.6977/IJoSI.202112_6(6).0004

Ganesan, A., Jaiswal, R., & Pitchaikani, A. (2018). A study of an integrated HVAC-vehicle model for automotive vehicles. *SAE International Journal of Passenger Cars Mechanical Systems*, 11, 151–166. https://doi.org/10.4271/06-11-02-0013

Ghane, M., Ang, M.C., Cavallucci, D., Kadir, R.A., Ng, K.W., & Sorooshian, S. (2022). TRIZ trend of engineering system evolution: A review on application, benefits, challenges and enhancement with computer-aided aspects. *Computers and Industrial Engineering*, 174, 108833. https://doi.org/10.1016/j.cie.2022.108833

Lee, D.H. (2024). An improved sensorless control of BLDC motor using 1-shunt current sensor for HVAC of EV. In: *2024 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*. IEEE, United States, p1–6.

Lee, D.W. (2014). Development of BLDC motor and multi-blade fan for HEV battery cooling system. *International Journal of Automotive Technology*, 15(7), 1101–1106. https://doi.org/10.1007/s12239-014-0114-7

Mohanraj, D., Aruldavid, R., Verma, R., Sathiyasekar, K., Barnawi, A.B., Chokkalingam, B., et al. (2022). A review of BLDC motor: state of art, advanced control techniques, and applications. *IEEE Access*, 10, 54833–54869. https://doi.org/10.1109/ACCESS.2022.3175011

Ravineala, T., Kurniawan, D., & Bennouna, S. (2025). Automotive HVAC blower blade pass noise sensitivity-a case study. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 271(2), 950–960. https://doi.org/10.3397/NC_2025_0163

Rigatos, G., Abbaszadeh, M., Sari, B., Siano, P., Cuccurullo, G., & Zouari, F. (2023). Nonlinear optimal control for a gas compressor driven by an induction motor. *Results in Control and Optimization*, 11, 100226. https://doi.org/10.1016/j.rico.2023.100226

Shao, J. (2006). An improved microcontroller-based sensorless brushless DC (BLDC) motor drive for automotive applications. *IEEE Transactions on Industry Applications*, 42(5), 1216–1221. https://doi.org/10.1109/TIA.2006.880888

Sheu, D.D., Chiu, M.C., & Cayard, D. (2020). The 7 pillars of TRIZ philosophies. *Computers and Industrial Engineering*, 146, 106572. https://doi.org/10.1016/j.cie.2020.106572

Sheu, D.D., & Chiu, S.C. (2017). Prioritized relevant trend identification for problem solving based on quantitative measures. *Computers and Industrial Engineering*, 107, 327–344. https://doi.org/10.1016/j.cie.2016.03.028

Sheu, D.D., & Lee, H.K. (2011). A proposed process for systematic innovation. *International Journal of Production Research*, 49, 847–868. https://doi.org/10.1080/00207540903280549

Singh, I. (2024). BLDC (Brushless Direct Current) Motors. Published in India 2024 by Pencil.

Zouari, F., Saad, K.B., & Benrejeb, M. (2013). Adaptive Backstepping Control for a Class of Uncertain Single Input Single Output Nonlinear Systems. In 10th *International Multi-Conferences on Systems, Signals and Devices 2013 (SSD13)*. IEEE, p1–6.

AUTHOR BIOGRAPHIES

Merve Yildiz completed her undergraduate studies in industrial engineering at Balikesir University and is currently pursuing her master's degree in industrial engineering at Bursa Technical University. She works as a platform purchasing engineer in the automotive industry, focusing on innovation methodologies, with a particular interest in the theory and practical applications of TRIZ.

Koray Altun is an assistant professor of industrial engineering at TU Bursa and a research collaborator at Glorad. He is also the founder and manager of a startup focused on digital innovation and software consultancy. He holds a Ph.D. and a B.Sc. degree in industrial engineering from Gaziantep University and Erciyes University, respectively. He has published papers in well-respected journals and has led consulting engagements in reputable companies with a focus on r&d, technology, and innovation management. His recent research interests include the innovation excellence model, systematic innovation, technology and innovation management, and r&d management.

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.