

Journal of Advances in Engineering and Technology

Editor-in-Chief:

Dr. Ke Zhang

China Waterborne Transport Research Institute, Ministry of Transport, China

Copyright © 2026. ASIA PACIFIC SCIENCE PUBLICATIONS
COMPANY LIMITED. Complimentary Copy.

Journal of Advances in Engineering and Technology

Journal of Advances in Engineering and Technology (JAET) is an international, peer-reviewed and open access journal which publishes original articles, reviews, short communications, case studies and letters in the field of electronic research and application. It covers mainly but not limits to the following areas:

- Civil Engineering
- Mechanical Engineering
- Electrical Engineering
- Chemical Engineering
- Aerospace Engineering
- Computer Science and Engineering
- Environmental Engineering
- Materials Science and Engineering
- Biomedical Engineering
- Energy Engineering
- Robotics and Automation

About Publisher:

Asia Pacific Science Press (APSP) is a swiftly expanding publisher of peer-reviewed and open-access journals, strategically located in Hong Kong. As a reliable and esteemed corporation, APSP is dedicated to promoting and serving a wide array of subject areas, ultimately contributing to the betterment of humanity. By disseminating knowledge to a global community of scholars, practitioners, researchers, and students, we strive to establish ourselves as the world's leading independent academic and professional publisher.

Submission instructions: You can submit your manuscript through the official website (www.apspublisher.com) or email (editor.jaet@apspublisher.com), All manuscripts will go through a rapid peer review and production, making the process of publishing simpler and more efficient.

Publisher Headquarter

Room 03, 7th Floor, Block B, Tuen Mun Industrial Centre, 2 New Ping Street, Tuen Mun, Hong Kong, China
Website : www.apspublisher.com
Email : info@apspublisher.com

Fujian Province Office, China

603-1, 6th Floor, Building B20, Chengyi North Street, Software Park, Jimei District, Xiamen City, Fujian Province, China
Website : <https://ojs.apspublisher.com/index.php/amit>
Email : amit@apspublisher.com

Table of Contents

- 1 Comprehensive Assessment Model for Navigation Safety Risks in Area of Offshore Wind Farm**
Ke Zhang, Shiqi Li, Dongsheng Li, Guozhu Hao, Xiao Liu, Xingya Zhao

- 12 Stability of Geogrid-Reinforced High-Fill Slopes: A Review of Experimental and Numerical Advances**
Tiantian Xiong , Nurazim Ibrahim

- 22 Automatic Modulation Classification using Residual Connection and Bottle-neck Transformers with Trend-aware Self-Attention**
Geng Wang, Luming Li, Xin Chen

- 35 Progress on CNT/Al Composites**
Lian Xue, Kangdi Zhong, Bingbing Wang, Lei Sha, Yuanxiao Dang

- 42 GASP-D3QN: Geometry-Aware Safety-Prioritized Dueling Double Q-Network for Online UAV Path Planning**
Kexiao Wu, Bingyu Yang, Mingshen Xu

- 53 A Review of Lightweight Research on 3D Printed Prosthetic Structure Based on Topology Optimization**
Kangdi Zhong, Lian Xue, Wen Sun, Haoyu Wang

- 61 Plastic and Concrete Construction Products: A Comparative Review of Material Performance, Cost, Sustainability, and Market Substitution Pathways**
Guangtao Zhang, Jiahang Ni, Junlong Zou, Bo Chen, Guangkai Zhang

- 72 Technology for Preparing Nanocrystalline Diamond Coatings by Hot-Filament CVD Method**
Zhihang Xue, Jianfeng Wang, Xing Hu

- 87 Optimization of Key Process Parameters for Air-Jet Vortex Spun Yarn**
Kexiang Yang

- 97 **Analysis of Cross-Border Enterprise Cooperation Cases under the Belt and Road Initiative——Take the Padma Bridge of China Railway Group as an Example**
Luyao Liang
- 103 **Design and Implementation of a Cross-Cultural Pattern Classification System Based on ResNet18**
Alina Andreeva, Zhenning Li, Wenjie Tang, Chaofan Yue, Xinlei Cao, Xiang Zhang, Xiangyu Li
- 114 **From Data Possession to Operations Reconfiguration: Transforming Firm Operations under Big Data - An Integrative Literature Review**
Mengfei Xiao, Weixiang Gan
- 127 **Development of an AI-Augmented AR Furniture Exhibition System: A Case Study of Industry-Academia Collaborative Engineering**
Haoyun Ma, Wangshuang Zang, Congrong Xiao
- 135 **Advances in Nitroreductase-Activated Fluorescent Probes Based on Photophysical Response Mechanisms**
Chenye Jin, Jinmao You
- 143 **Practical Pathways for Enhancing Artificial Intelligence Literacy among University Teachers in Ningbo under the Background of Digital-Intelligent Transformation**
Kebiao Yuan, Yun Ren
- 159 **Heavy Metal Passivation and Microbial Community Succession during Anoxic-Aerobic Digestion of Municipal Sludge**
Ruihua Mu, Kaige Zhang, Wenzhuo Chen, Xiaoqian Shi, Fengqiu An, Ye Liu

Comprehensive Assessment Model for Navigation Safety Risks in Area of Offshore Wind Farm

Ke Zhang¹, Shiqi Li², Dongsheng Li¹, Guozhu Hao³, Xiao Liu³, Xingya Zhao^{3*}

1. China Waterborne Transport Research Institute, Beijing 100088, China

2. School of Navigation, Jimei University, Xiamen 361021, China

3. School of Navigation, Wuhan University of Technology, Wuhan 430063, China.

*Corresponding author: Xingya Zhao, zhaoxingya@whut.edu.cn

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: To address the challenge of assessing regional system safety risks in the context of the large-scale, clustered development of offshore wind farm regions, this study focuses on China's coastal offshore wind farm regions and constructs a comprehensive safety risk assessment model for offshore wind farm regions based on multi-spatio-temporal coupling. In this model, we first identify risk factors in offshore wind farm regions from a four-dimensional perspective encompassing people, machinery, environment, and management, and introduce a multi-spatio-temporal coupling factor to optimize the indicator system. Next, we use the Analytic Hierarchy Process (AHP) to determine the weighting of each assessment indicator, and then combine this with the Fuzzy Comprehensive Evaluation Method to establish a comprehensive safety risk assessment model for offshore wind farm regions. Finally, we apply this model to the risk assessment of a real coastal wind power region. The results indicate that, at the present time, the overall risk level of a specific offshore wind farm regions area is classified as "moderate risk." This assessment model can effectively quantify the dynamic coupled risks in offshore wind farm regions areas, providing a theoretical basis and technical support for the planning, layout, and safety supervision of offshore wind farms.

Keywords: Offshore Wind Farm; Navigation Safety; Fuzzy Comprehensive Evaluation Method; Risk Evaluation

Published: Mar 29, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1216>

1. Introduction

Against the backdrop of the ongoing advancement of the "dual carbon" goals and the strategy to build a maritime power, coastal wind power generation has become a key focus of clean energy development in China's coastal regions. Currently, as the siting of offshore wind farms shifts from single-site development toward large-scale, clustered development and full-lifecycle operations, safety issues have evolved from risks associated with individual facility failures or single operational phases to regional systemic risks involving the combined effects of multiple factors, including navigation environments, offshore construction, operation and maintenance activities, facility and equipment conditions, and organizational management. Particularly in scenarios where multiple wind farms coexist and construction and operational phases overlap, the risks in offshore wind energy areas may exhibit greater dynamism, coupling, and diffusion^[1]. Therefore, developing comprehensive assessment models that reflect the overall characteristics of the region and the dynamic evolution of risks has

become a critical issue in offshore wind energy safety management research.

Scholars both domestically and internationally have conducted research on safety risks associated with offshore wind farm regions from various perspectives, primarily covering navigation risks during construction and operation and maintenance (O&M) phases, risks related to turbine structures and facilities, specific disaster risks in wind farms, and wind farm planning and decision-making. Regarding navigation safety, Yu et al. systematically reviewed the current state of research on the quantification of vessel navigation risks in offshore wind farm regions waters^[2]. Li et al. proposed a path planning method for O&M vessels in offshore wind farms based on an improved A* algorithm^[3]. This method overcomes the limitations of the traditional A* algorithm, which fails to account for dynamic obstacles, water currents, and crossing channels during path planning for O&M vessels in offshore wind farms, resulting in significant improvements in path smoothness and the overall risk value of the planned path. Wang et al. established an evaluation index system for navigation safety in offshore wind farm construction waters^[4]. The results indicate that offshore wind farm construction has a significant impact on the navigation environment and traffic order in adjacent waters. Teng et al. proposed a comprehensive assessment method for the safety distance between vessels and offshore wind farms based on AIS data to reduce navigation risks for vessels near wind farm waters^[5]. Wang et al. combined fuzzy AHP-DEMATEL with a cloud model to study the navigation safety assessment of offshore wind farms; this method effectively enhances the systematic nature of indicator weight identification and risk determination^[6]. Zhang et al. constructed a multi-objective optimization model to assess vessel navigation risks in offshore wind farm areas, expanding the application of risk assessment in front-end planning^[7]. Some international scholars have studied the impact of offshore wind farm regions development on maritime traffic safety. Rutkowski et al. utilized a spatial vessel domain model to investigate navigation risks near proposed wind farms^[8]. Son et al. conducted research on the design of navigation channel widths adjacent to offshore wind farms based on AIS data and collision probability models, arguing that wind farm development requires a comprehensive consideration of collision risks between vessels as well as between vessels and wind turbines^[9]. Milin et al. pointed out through extensive literature analysis that offshore wind farms may increase the complexity of maritime traffic, which may lead to navigation safety and collision accidents^[10].

In terms of wind farm construction, operation and maintenance, and facility safety. Huang Hui et al. used the AHP fuzzy comprehensive evaluation method to quantitatively analyze the safety risks of construction and hoisting operations in offshore wind farms, and verified the applicability of the fuzzy evaluation method in high-risk operation scenarios. Dong et al. constructed a safety risk model for offshore wind turbine installation operations based on the AHP-fuzzy comprehensive evaluation method, covering five dimensions—personnel, management, equipment, vessels, and electrical systems—to achieve a comprehensive assessment of safety risks during installation^[11]. Zhu et al. proposed an FAHP-based evaluation method to address lightning disaster risks during the operational phase of offshore wind farms^[12]. In addition, some researchers have conducted studies from the perspectives of the full life cycle and systems engineering. Huang et al. identified multi-level risk factors and employed AHP and the risk matrix method to conduct safety evaluation research during the construction of offshore wind farm regions subsea cables^[13]. Kim et al. developed an HSE risk assessment module for offshore wind farm regions based on systems engineering methods, using a combination of risk assessment methods to address risks under conditions of uncertainty^[14].

Based on this, this paper identifies risk factors from four dimensions—human, machine, environment, and management—across the entire lifecycle of offshore wind farm regions areas. It further introduces time-coupled and space-coupled factors under multi-spatio-temporal coupling. Finally, by combining the Analytic Hierarchy Process (AHP) with the Fuzzy Comprehensive Evaluation Method, the paper constructs a comprehensive dynamic assessment model for offshore wind farm regions safety risks. Through a risk assessment of a specific offshore wind farm regions area along China's coastline, the research findings provide a theoretical basis for the planning and layout of offshore wind farms, the assurance of navigation safety, and safety supervision.

2. Identification of Regional Risk Factors for offshore wind farm regions

Unlike the risks associated with a single wind farm, the regional risk assessment of offshore wind farms is a complex system comprising wind farms at various stages of development and in different water areas. It should be treated as a dynamic,

integrated whole, necessitating a comprehensive analysis; this also presents greater challenges for identifying and mitigating water-related risk factors.

This paper treats offshore wind farms within the entire lifecycle of an offshore wind farm regions area as a single entity, identifies their primary risk factors, and then conducts a risk assessment of the offshore wind farm regions area based on these identified risk factors. By analyzing risk factors from four perspectives—human, machinery, environment, and management—the risk factors for offshore wind farm regions areas are summarized as follows: environmental factors (natural environment and transportation environment), human factors (crew and operational personnel), management factors (corporate management and management by supervisory authorities), and machinery factors (vessel-related and equipment-related). The risk factors for offshore wind farm regions areas are presented in Table 1.

Table 1. Identification of key risk factors in offshore wind farm regions areas

Risk Type	Key risk factors
Environmental factors	Wind, waves, and currents (typhoon)
	Heavy fog
	Thunder and lightning
	Sensitive resources nearby
	Natural Conditions of the Project Site
	Insufficient water depth in the channel
	Intersection of customary routes and existing routes
	Transportation Environment
	There are many vessels operating in the wind farm area
	Traffic chaos
Human factors	Lack of navigation aids
	Insufficient crew training
	Crew members' violations
	Unfitness of crew members
	Lack of safety awareness
	Operators lack sufficient skill
Management factors	Operators
	Operators' non-compliance with operating procedures
	Other safety hazards
	Management negligence at the enterprise level
	Inadequate emergency response plans/safety operating procedures
	Unnecessary warning and monitoring equipment
Machine factors	Inadequate corporate communication systems
	Inadequate supervision or guidance from regulatory authorities
	Lack of clarity regarding the responsibilities of supervisory agencies
	Poor communication between regulatory agencies and businesses
	A collision caused by a loss of control of an individual vessel
	The ship's emergency lifesaving equipment is inadequate
Machine factors	Fire and Explosion Accidents
	Personal protective equipment is incomplete or in poor condition
	Inadequate safety equipment at wind farms
	Failure or malfunction of wind turbine structures and equipment components

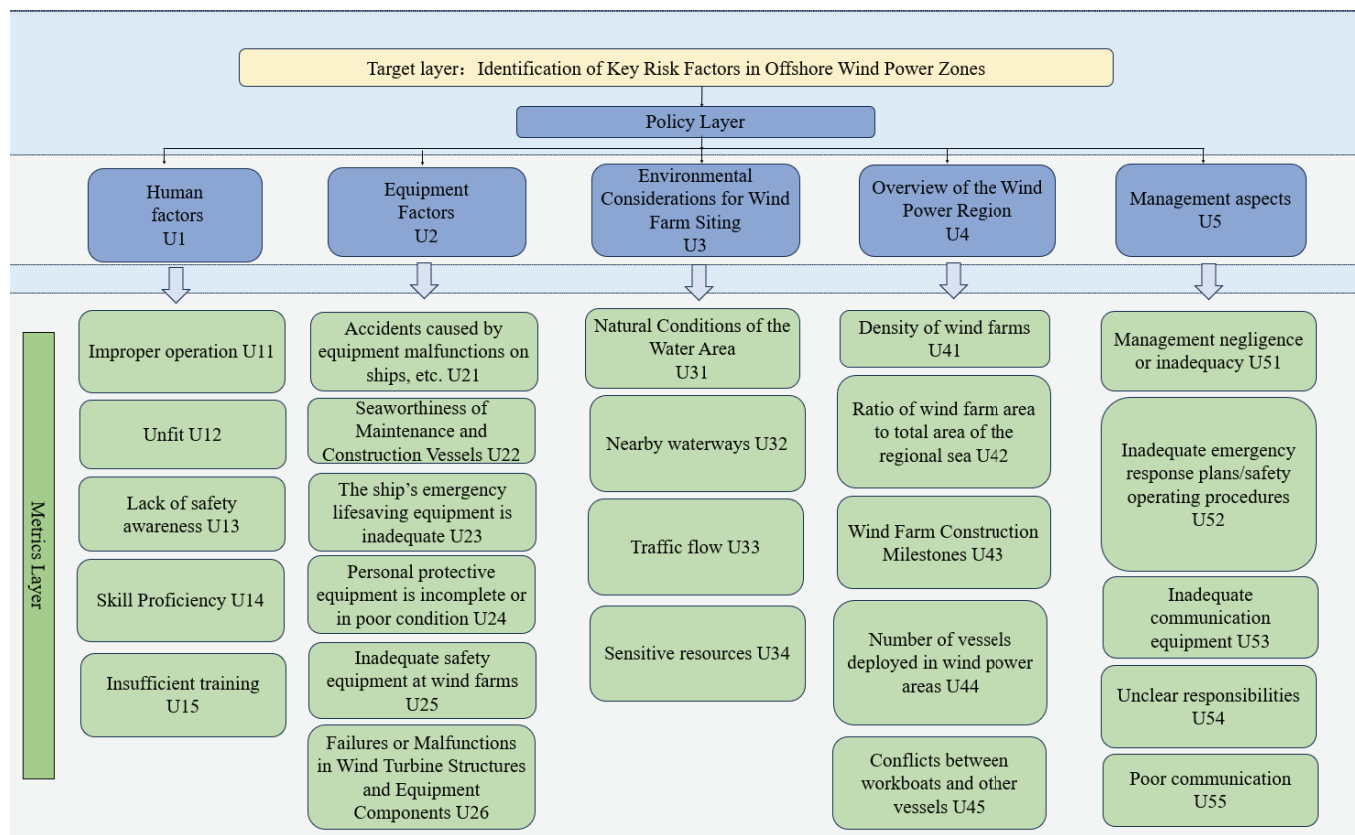
Given that offshore wind farm regions zones contain wind farms at various stages of development, and that these wind farms interact with one another both temporally and spatially, analysis cannot rely solely on initially identified indicators. Instead, it is necessary to conduct a more in-depth exploration of the inter-wind-farm indicator data. Therefore, risk factors under multi-spatio-temporal coupling must be identified to form coupled factors, and the risk factors listed in Table 1 must be adjusted accordingly.

3. Development of an Indicator System for Assessing Wind Power Safety Risks in Coastal Areas

A scientifically reasonable evaluation index system is an important prerequisite for conducting safety risk assessments in offshore wind power regions. The Analytic Hierarchy Process (AHP) is highly applicable to complex systems involving multiple objectives and criteria. It can decompose complex risk issues into several interrelated components and establish a hierarchical structure based on the dominance relationships among these factors, thereby laying the foundation for subsequent weight determination and comprehensive evaluation. For offshore wind farm regions, using the AHP to construct an indicator system helps achieve a systematic characterization of regional safety risks.

Offshore wind farm areas face a variety of uncertainties that vary over time and space; moreover, these uncertainties may interact with one another, posing greater challenges for the identification and mitigation of offshore risks. Therefore, it is necessary to apply coupling theory to integrate the spatiotemporal aspects of offshore wind farms throughout their entire lifecycle in order to identify risk factors. As shown in Figure 1, a three-tiered hierarchical evaluation index system has been established, comprising the objective layer, the criterion layer, and the indicator layer. The objective layer defines the ultimate purpose of the evaluation; the criterion layer classifies risk types across four dimensions; and the indicator layer breaks these down into multiple specific, measurable risk indicators, thereby forming a comprehensive risk evaluation index system. The criterion layer serves as a bridge connecting the objective layer and the indicator layer, playing a pivotal role in the index system.

Figure 1. Identification of Major Risk Factors in offshore wind farm regions Areas



Specifically, the human factor includes personnel involved in safety within the wind farm area, such as construction workers, operational staff, and managers. The probability of non-compliant operations refers to the likelihood of personnel in the wind

farm area engaging in non-compliant operations, quantified as a value between 0 and 1; a higher value indicates a greater probability of non-compliant operations and a higher risk.

With regard to equipment-related factors, accidents caused by vessel equipment failures refer to factors that may lead to risks due to malfunctions in the vessel's own facilities and equipment; these are quantified on a scale of 0 to 1, with higher values indicating greater risk. Failures or malfunctions in wind turbine structures and components reflect the reliability or failure rate of wind turbine facilities and equipment in the wind power area; these are quantified on a scale of 0 to 1, with higher values indicating higher risk.

Regarding environmental factors for wind farm siting, the natural conditions of the water area include natural factors such as typhoons, waves, and currents, quantified on a scale of 0 to 1; nearby shipping lanes refer to the conditions of shipping lanes near the wind farm, quantified on a scale of 0 to 1; and traffic flow refers to traffic conditions near the wind farm, quantified on a scale of 0 to 1.

Regarding the general characteristics of wind power areas, the distribution density of wind farms primarily reflects the degree of concentration of wind farms. It is generally believed that the higher the distribution density, the greater the risk. This is quantified on a scale of 0 to 1, with higher values indicating greater risk; Conflicts involving construction and O&M vessel traffic primarily consider the intersection of navigation routes among operational vessels. It is generally believed that the more intersections there are in the wind power area, the greater the potential risk. This is quantified as a value between 0 and 1, with higher values indicating greater risk.

With regard to management factors, management negligence or inadequacy refers to the failure of managers to adequately oversee wind farms and personnel in the wind power area. It is generally believed that the greater the negligence, the higher the risk of accidents or emergencies. This is quantified on a scale of 0 to 1, with higher values indicating greater risk. Poor communication refers to the level of information flow within the wind power area and is quantified on a scale of 0 to 1, with higher values indicating greater risk.

4. Comprehensive Dynamic Assessment Model for Wind Power Safety Risks in Coastal Areas

4.1 Determining Indicator Weights Using the Analytic Hierarchy Process

The weights for the regional risk assessment indicators of offshore wind farms were calculated using the Analytic Hierarchy Process (AHP). The specific calculation process is as follows:

(1) Construction of the Evaluation Matrix

Invite experts to assign importance scores to indicators, thereby constructing an evaluation matrix $X = \{X_1, X_2, \dots, X_n\}$, as shown in Equation (1):

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}$$

Where: x_{mn} represents the score assigned by expert n to indicator m .

(2) Calculation of combined weighted vector values

By combining the effects of different indicators, we obtain the combined weighted vector λ_{j+1} , as shown in the following equation:

$$\lambda_{j+1} = \frac{C_{n-1}^j}{\sum_{k=0}^{n-1} C_{n-1}^k} = \frac{C_{n-1}^j}{2^{n-1}}$$

Where: $\sum_{j=0}^{n-1} \lambda_{j+1} = 1$; C_{n-1}^j is the number of permutations of j elements selected from $n-1$ elements.

(3) Calculation of absolute weights.

Based on the weighted vectors, calculate the absolute weight w_i of indicator i , as shown in the following equation:

$$w_i = \sum_{j=0}^{n-1} \lambda_{j+1} \cdot X_j (j = 0, 1, \dots, n-1; i = 1, 2, \dots, m)$$

Where: X_j denotes the set of all indicator score vectors for expert j .

(4) Determination of relative weights

Based on the absolute weight of indicator i , calculate the relative weight σ_i of indicator i , as shown in the following equation:

$$\sigma_i = \frac{w_i}{\sum_{i=1}^m w_i}$$

4.2 A Dynamic Comprehensive Risk Assessment Method Based on Fuzzy Comprehensive Evaluation

Using the indicator weights determined by the Analytic Hierarchy Process (AHP), we developed a comprehensive regional risk assessment model for offshore wind farms based on the Fuzzy Comprehensive Evaluation Method.

(1) Determine the set of factors

The set of factors influencing the safety assessment of offshore wind farm regions under multi-temporal and spatial coupling can be represented by primary and secondary indicators as follows:

Level 1 indicator : $U = \{U_1, U_2, U_3, U_4, U_5\}$

Level 2 indicator : $U1 = \{U_{11}, U_{12}, U_{13}, U_{14}, U_{15}\}$, $U2 = \{U_{21}, U_{22}, U_{23}, U_{24}, U_{25}, U_{26}\}$, $U3 = \{U_{31}, U_{32}, U_{33}\}$, $U4 = \{U_{41}, U_{42}, U_{43}, U_{44}, U_{45}\}$, $U5 = \{U_{51}, U_{52}, U_{53}, U_{54}, U_{55}\}$

(2) Determine the weight set

The weights of the indicators were calculated and determined using the Analytic Hierarchy Process described in Section 3.1.

(3) Create a collection of comments

The safety risks of offshore wind farm regions areas are classified into four levels: I, II, III, and IV, and are grouped into Rating Category V. The safety risk assessment ratings and their descriptions for offshore wind farm regions areas are shown in Table 2, and the classification criteria for safety risks in offshore wind farm regions areas are shown in Table 3.

Table 2. Safety Risk Classification Comments and Descriptions for offshore wind farm regions Zones

Comments	Risk Profile	Risk Description
I	Low risk	The risk is acceptable; no action is required.
II	General risks	Where risks are tolerable and cost-effective, efforts should be made to further mitigate them.
III	Significant risk	Where technically and economically feasible, risks must be mitigated
IV	High risk	This is extremely dangerous; immediate corrective action is required.

Table 3. Safety Risk Classification Standards for offshore wind farm regions Areas

Evaluation Criteria	[50,70]	[70,80]	[80,90]	[90,100]
Safety Risk Classification for offshore wind farm regions Zones	I	II	III	IV

(4) Determine the membership function

This method, which derives indicator scores based on actual scores or expert ratings, does not account for the ambiguity of boundaries between grades. When applying the fuzzy comprehensive evaluation method, membership degrees are primarily determined using membership functions. This allows for the consideration of transitional states between grades, thereby resolving the issue of fuzzy boundaries between grades. Therefore, selecting the trigonometric method to calculate the membership degree of evaluation indicators within evaluation grades can eliminate the unreasonable phenomena caused by abrupt changes between adjacent grade regions. The membership function formula is as follows:

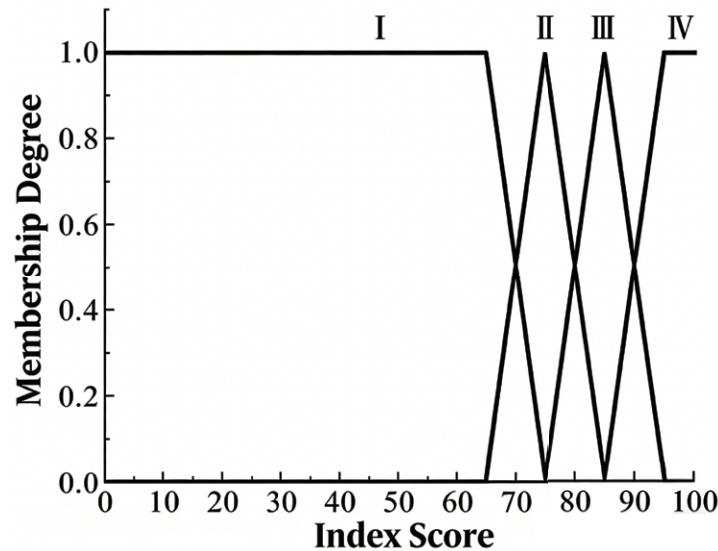
$$Y_I(x) = \begin{cases} 1, & 50 \leq x < 65 \\ \frac{75-x}{10}, & 65 \leq x < 75 \end{cases}$$

$$Y_{II}(x) = \begin{cases} \frac{x-65}{10}, & 65 \leq x < 75 \\ \frac{85-x}{10}, & 75 \leq x < 85 \end{cases}$$

$$Y_{III}(x) = \begin{cases} \frac{x-75}{10}, & 75 \leq x < 85 \\ \frac{95-x}{10}, & 85 \leq x < 95 \end{cases}$$

$$Y_{IV}(x) = \begin{cases} \frac{x-85}{10}, & 85 \leq x < 95 \\ 1, & 95 \leq x \leq 100 \end{cases}$$

Figure 2 Triangular membership function



(5) Determine the fuzzy evaluation matrix

Construct the fuzzy evaluation matrix R_1, R_2, R_3, R_4 . R represents a mapping from U to V defined by the fuzzy evaluation matrix, which can be expressed in matrix form as: $R = \{r_{ij}\}$, where $i=1,2,\dots,n; j=1,2,\dots,n$. r_{ij} represents the membership degree of each single factor U_i for evaluation level V_j .

(6) Determine the comprehensive evaluation score and risk level

First, the fuzzy comprehensive evaluation set B for the factor set U is calculated based on the $B=W*R$ fuzzy comprehensive evaluation criteria. Next, the fuzzy comprehensive evaluation vectors for the first-level indicators are calculated sequentially to obtain the fuzzy comprehensive evaluation vector for the evaluation target. Finally, the fuzzy comprehensive evaluation score is calculated using the $S=B*G$ formula to determine the comprehensive evaluation grade for offshore wind farm regions.

5. Case Study

At present, there are multiple existing and under construction wind farms in a certain offshore wind farm area along the coast of China, based on the distribution map of wind farms and research data of wind farm areas. Adopting the constructed fuzzy comprehensive evaluation method to evaluate wind farms in offshore wind farm regions.

(1) Indicator weights derived from the Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) was used to determine the weights for the fuzzy comprehensive evaluation of offshore wind farms under multi-spatiotemporal coupling; the weight values are shown in Table 4.

Table 4. Weighting Values for offshore wind farm regions Regional Indicators

Evaluation criteria	Weight allocation	Evaluation Criteria	Weight allocation
Human factors U1	0.2360	Improper operation U11	0.0557
		Unfit U12	0.0655
		Lack of safety awareness U13	0.0379
		Skill Proficiency U14	0.0464
		Insufficient training U15	0.0304

Evaluation criteria	Weight allocation	Evaluation Criteria	Weight allocation
Equipment Factors U2	0.1288	Accidents caused by equipment malfunctions on ships, etc. U21	0.0276
		Seaworthiness of Maintenance and Construction Vessels U22	0.0232
		The ship's emergency lifesaving equipment is inadequate U23	0.0112
		Personal protective equipment is incomplete or in poor condition U24	0.0195
		Inadequate safety equipment at wind farms U25	0.0311
		Failures or Malfunctions in Wind Turbine Structures and Equipment Components U26	0.0161
Environmental Considerations for Wind Farm Siting U3	0.1968	Natural Conditions of the Water Area U31	0.0534
		Nearby waterways U32	0.0444
		Traffic flow U33	0.0628
		Sensitive resources U34	0.0362
		Density of wind farms U41	0.0358
Overview of the Wind Power Region U4	0.2776	Ratio of wind farm area to total area of the regional sea U42	0.0446
		Wind Farm Construction Milestones U43	0.0771
		Number of vessels deployed in wind power areas U44	0.0546
		Conflicts between workboats and other vessels U45	0.0655
		Management negligence or inadequacy U51	0.0446
		Inadequate emergency response plans/safety operating procedures U52	0.0379
Management aspects U5	0.1608	Inadequate communication equipment U53	0.0207
		Unclear responsibilities U54	0.0316
		Poor communication U55	0.0259

(2) Comprehensive Risk Assessment Based on the Fuzzy Comprehensive Evaluation Method

Taking Wind Farm No. 19 as an example, the fuzzy comprehensive evaluation vector for the evaluation objective is calculated based on the $B = W * R$ fuzzy comprehensive evaluation criteria:

$$\begin{aligned}
 B = W * R &= (0.2360, 0.1288, 0.1968, 0.2776, 0.1608) * \begin{bmatrix} 0 & 0.2892857 & 0.455426 & 0.2553412 \\ 0 & 0.2511250 & 0.414914 & 0.3333951 \\ 0.027134 & 0.1361788 & 0.508943 & 0.3277439 \\ 0.019668 & 0.2503604 & 0.500576 & 0.2293951 \\ 0.019664 & 0.1464842 & 0.4263845 & 0.3913503 \end{bmatrix} \\
 &= (0.01396178, 0.22047102, 0.46860397, 0.29431102)
 \end{aligned}$$

Calculate the $S = B * G$ fuzzy comprehensive evaluation score using the formula:

$$S = B * G = \begin{pmatrix} 0.01396178 \\ 0.22047102 \\ 0.46860397 \\ 0.29431102 \end{pmatrix} * \begin{vmatrix} 100 \\ 85 \\ 75 \\ 60 \end{vmatrix} = 72.9$$

The final risk assessment result was 72.94. According to Tables 2 and 3, the comprehensive risk rating for Wind Farm No. 19 is Level II, which corresponds to “general risk”—meaning the risk is tolerable. However, if it is cost-effective to do so,

further risk mitigation measures should be implemented.

(3) Risk Assessment of Offshore Wind Farms

Based on collected data, field surveys, and consultations with wind power professionals, and taking into account risk factors, the wind farms in the study area were evaluated using assessment criteria across four dimensions: people, vessels, the environment, and management. The individual wind farms are shown in Figure 3.

Figure 3. Map of Wind Farm Locations and Risk Assessment

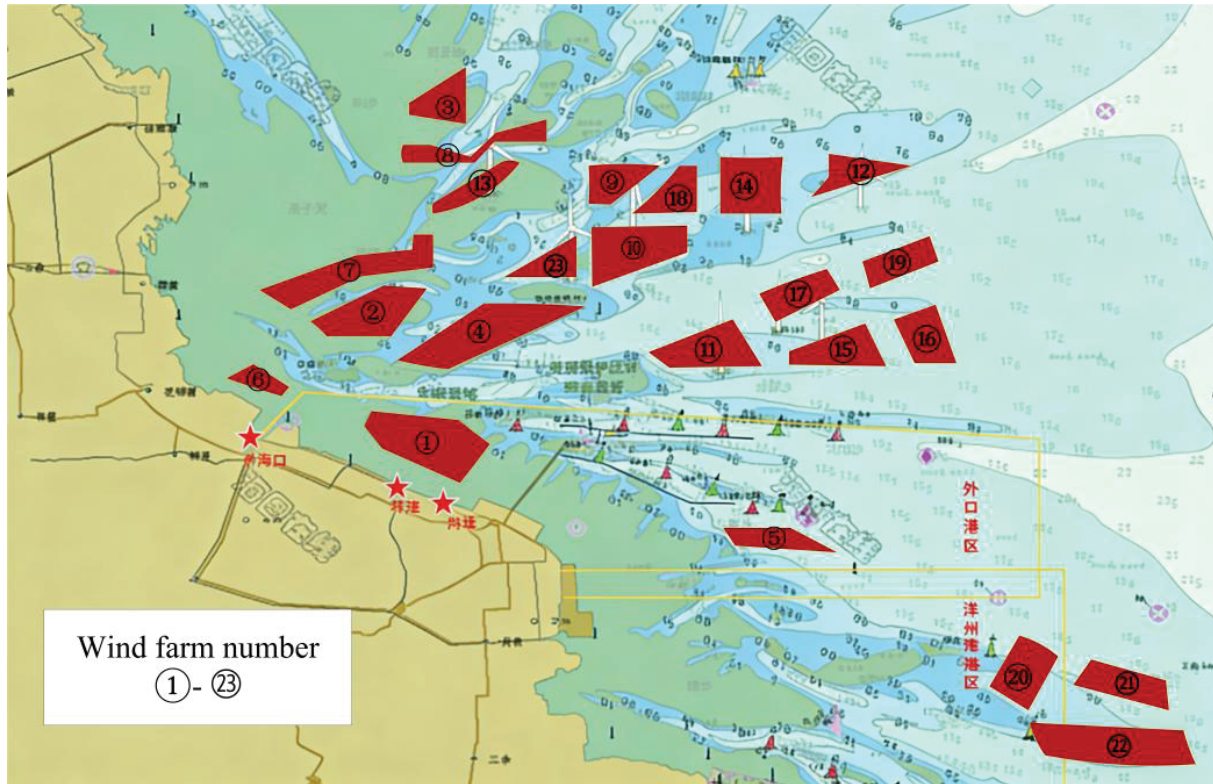
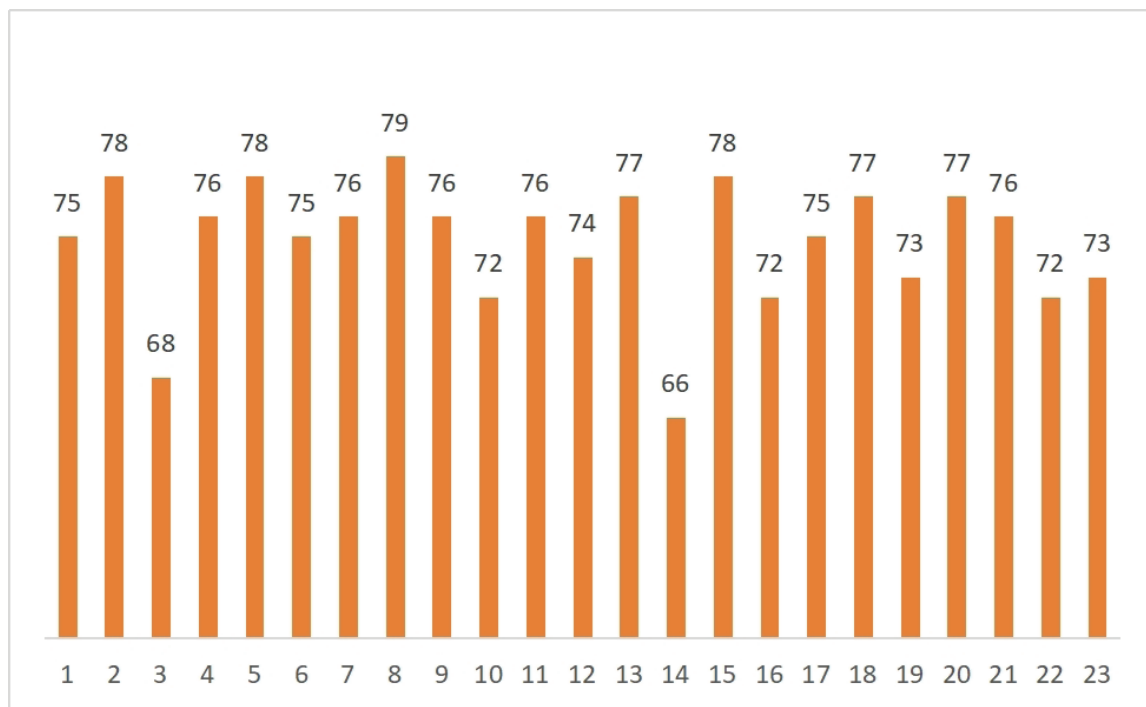


Figure 4. Distribution of Risk Assessment Values for Each Wind Farm



Based on the evaluation method described above, the final risk assessment ratings for each wind farm were determined, as shown in Table 3.

Table 5. Risk Assessment Results for Wind Farms in the Offshore Wind Farm Area

Number	Rating Level	Number	Rating Level
1	General risks	13	General risks
2	General risks	14	Low risk
3	Low risk	15	General risks
4	General risks	16	General risks
5	General risks	17	General risks
6	General risks	18	General risks
7	General risks	19	General risks
8	General risks	20	General risks
9	General risks	21	General risks
10	General risks	22	General risks
11	General risks	23	General risks
12	General risks		

An assessment of safety risks at 23 wind farms in the jurisdiction—covering both operational and construction phases—revealed that wind farms numbered 3 and 14 were classified as “low risk,” while the remaining 21 wind farms were classified as “moderate risk.” Although the overall risk level for the wind farm area is “moderate,” several wind farms are currently in the construction phase and may be significantly affected by construction vessels. Therefore, for wind farms with higher risk assessment values and those in the construction phase, wind power companies and relevant departments should pay special attention to risk safety management and prevention measures for these facilities, further reducing risks during this stage and preventing the risk level from escalating further.

6. Conclusion

This paper conducts a systematic risk assessment analysis of a specific offshore wind farm in a coastal region of China. Risk factors were identified based on multi-spatio-temporal interactions, and an indicator system for evaluating wind power safety risks in coastal areas was established. Finally, a comprehensive risk assessment model for coastal wind power was constructed using the Analytic Hierarchy Process (AHP) and the Fuzzy Comprehensive Evaluation Method. Case analysis indicates that the overall risk level of wind farms under construction and in operation in the region during this period is “moderate.” The application of the Fuzzy Comprehensive Evaluation Method to the risk assessment of a specific offshore wind farm in China’s coastal region provides valuable theoretical references and practical applications for regional offshore wind farm regions safety management.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Li, Y. N., Zhang, X. S., Wu, C. Y., & Others. (2026). A study on the optimization of condition-based maintenance for offshore wind turbines under uncertainty in lifespan prediction. *Journal of Safety and Environment*, 1–13. <https://link.cnki.net/urlid/11.4537.x.20260108.1712.002>
- [2] Yu, Q., Liu, K. Z., Yuan, Z. T., & Others. (2022). A review of the current status of research on the quantification of vessel navigation risks in offshore wind farm waters. *China Navigation*, 45(3), 21–25.

- [3] Li, Z. Y., Wan, X., Lü, J. Y., & Others. (2025). Route planning for offshore wind farm maintenance vessels based on an improved A* algorithm. *China Navigation*, 48(1), 132–140.
- [4] Wang, Z. H., Fan, Z. Z., & Feng, Y. (2022). Assessment of navigation safety in waters adjacent to offshore wind farm projects. *Shanghai Journal of Maritime Studies*, 43(2), 54–59. <https://doi.org/10.13340/j.jsmu.2022.02.009>
- [5] Teng, J. P., Xiong, Z. N., Yao, Z., & Others. (2023). A comprehensive method for assessing safety distances between ships and offshore wind farms. *Journal of Jimei University (Natural Science Edition)*, 28(6), 521–526. <https://doi.org/10.19715/j.jmuzr.2023.06.05>
- [6] Wang, F. W., Wang, K. S., & Zhang, Z. R. (2024). Evaluation of navigation safety in offshore wind farms based on fuzzy AHP-DEMATEL. *China Navigation*, 47(4), 19–27.
- [7] Zhang, J. F., Xiong, M. L., Wu, X. H., & Others. (2024). Multi-objective optimization of offshore wind farm siting considering navigation safety. *Chinese Journal of Navigation*, 47(4), 66–72.
- [8] Rutkowski, G., & Kubacka, M. (2025). Navigational risk assessment in offshore wind farms using spatial ship domain models. *Applied Sciences*, 15(12), 6943. <https://doi.org/10.3390/app15126943>
- [9] Son, J. W., & Cho, S. I. (2024). Optimal maritime traffic width for passing offshore wind farms based on ship collision probability. *Ocean Engineering*, 313, 119498. <https://doi.org/10.1016/j.oceaneng.2024.119498>
- [10] Milin, V., Skoko, I., Lekšić, Ž., & Others. (2025). Navigational safety hazards posed by offshore wind farms: A comprehensive literature review and bibliometric analysis. *Journal of Marine Science and Engineering*, 13(7), 1330. <https://doi.org/10.3390/jmse13071330>
- [11] Dong, T., Xu, H., Huang, H., & Others. (2026). Safety risk analysis and countermeasures for offshore wind turbine installation operations. *Hydropower and New Energy*, 40(1), 67–71. <https://doi.org/10.13622/j.cnki.cn42-1800/tv.1671-3354.2026.01.014>
- [12] Zhu, W. S., Huang, Y. D., Lu, W., & Others. (2024). Risk assessment of lightning hazards in offshore wind farms based on FAHP. *Guangdong Meteorology*, 46(1), 83–86+91.
- [13] Huang, H., Zhang, Q., Xu, H., & Others. (2024). Risk identification and safety evaluation of offshore wind power submarine cable construction. *Journal of Marine Science and Engineering*, 12(10), 1718. <https://doi.org/10.3390/jmse12101718>
- [14] Kim, S., Nam, K., Lee, T., & Others. (2024). Development of a Korean offshore wind power HSE risk assessment module based on systems engineering approach. *International Journal of Naval Architecture and Ocean Engineering*, 16, 100612. <https://doi.org/10.1016/j.ijnaoe.2024.100612>

Stability of Geogrid-Reinforced High-Fill Slopes: A Review of Experimental and Numerical Advances

Tiantian Xiong^{1,2,3*}, Nurazim Ibrahim¹

1. Faculty of Engineering Science and Technology, Kuala Lumpur University of Science and Technology, Selangor 43000, Malaysia

2. College of Urban Construction, Xi'an Siyuan University, Xi'an, 710038, China

3. School of Human Settlements and Civil Engineering, Xi'an Jiao Tong University, Xi'an, 710049, China

*Corresponding author: Tiantian Xiong, 233924311@s.klust.edu.my

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Geogrid Reinforcement Technology is one of the most widely used slope stabilization techniques in modern geotechnical engineering. In the following paper, the recent research achievements and development trends of geogrid reinforcement technology will be systematically introduced from the perspective of four aspects: interface interaction characteristics, strength properties of the reinforcement, numerical simulation and analysis, and long-term service performance. Based on the existing 40 core publications in the related field, the following conclusions can be drawn: Firstly, the study of interface characteristics has evolved from the analysis of the macro level to the exploration of anisotropy characteristics, cyclic responses, and environmental degradation effects; however, there is still a big gap between the micro-mechanisms and the macro-analysis; Secondly, the results of the triaxial test show that the main function of the geogrid is to improve the cohesion of the soil; Thirdly, the problem of particle fragmentation in special soils, such as calcareous sand and coral sand, has become a key point in the current study. (3) Numerical simulation methods have developed from traditional macroscopic finite element analysis to FEM-DEM coupled multi-scale methods. The application of green reinforcement materials and new structures has expanded the engineering application range of numerical simulations; (4) the study of long-term performance has changed from empirical judgment to full-scale monitoring and data assimilation, while long-term prediction models under coupled conditions of multiple fields still need improvement. This paper aims to identify the gaps of current research and forecast the direction of future study based on a review of the current status of related research, which can serve as a comprehensive reference for related studies.

Keywords: Geogrid; Reinforced Soil; Slope Stability; Interface Characteristics; Triaxial Test; Numerical Simulation

Published: Mar 29, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1214>

1. Introduction

1.1 Research Background and Importance

The geogrid reinforcement technology was invented in the late 1970s in accordance with the theory of reinforced soil developed by French engineer Henri Vidal. It has been over four decades since the geogrid reinforcement technology was invented and applied in the field. The geogrid reinforcement technology is now a very important technology in the field of

geotechnical engineering. The geogrid reinforcement technology is very simple in principle. The geogrid reinforcement technology is developed in accordance with the idea of placing geogrids in layers in the soil mass. By virtue of the frictional force between the geogrid and the soil mass and the passive resistance of the transverse ribs, it is possible to improve the overall tensile properties of the soil mass so that the soil mass and geogrids are combined in terms of their force-bearing properties. The geogrid reinforcement technology has been widely used in various projects because it is convenient in use, has low costs, and is environmentally friendly. With the continuous advancement of relevant research and the accumulation of extensive engineering experience, the understanding of the reinforcement mechanism of geogrids has evolved from the description of the macroscopic characteristics to the establishment of a systematic theory involving the coupling of multiple scales and multiple fields. The current research on the reinforcement mechanism of geogrids can be generally divided into three aspects. The first is the interface characteristics between the soil and the geogrids. This began with the dual mechanism interface model proposed by Jewell and Milligan in 1989^[1], which has since been extended to incorporate anisotropy^[6], responses to cyclic loading^[8,10], and environmental degradation^[11,12]. The second strand is the mechanical properties and testing methods in reinforced soil. This has mainly been achieved through direct shear tests^[3-5] and triaxial tests^[15-17,21] that have analyzed and determined the constitutive behavior of the mechanical properties of reinforced soil. The third strand is the analytical methods used in the stability analysis of reinforced slopes. This has progressed from the limit equilibrium methods that were first employed to the popular and recent finite element methods^[29-30,32], and then to the recent multi-field coupling and multi-source assimilations^[35-36], with the overall analytical methods being refined and coupled with each other.

1.2 Research Objectives

A review of the core literature in this field has identified three critical scientific questions and research gaps. Firstly, there is a lack of integration between the microscopic and macroscopic responses in interface property research. Although the anisotropy of interface strength and the effects of factors such as normal stress and grid type on interface strength have been identified in research, and even generalized equations for interface strength have been proposed^[6], the results have not been effectively integrated into the overall analysis of slope stability. Numerical simulation results relevant to this field also show that the interface constitutive models commonly used in the finite element method are too simple and are not effective in reflecting the spatial and directional variations in interface strength.

Secondly, the research on the reinforcement of special soils is still in the phenomenological description stage and lacks sufficient modeling. With the development of engineering projects into complex geological zones, the reinforcement of special soils such as calcareous sand and coral sand has received increasing attention. Through the research carried out so far, it has been found that particle fragmentation plays a very negative role in the reinforcement of special soils^[16-18]. Some scholars have studied the influence of the grid's aperture and tensile strength from the perspectives of macro and micro. However, there is still no consensus on the method for quantifying the influence of particle fragmentation and incorporating it into the actual design. Although the research carried out using 3D printing technology has investigated the corresponding relationship between the aperture and the particle size, providing a new direction for the refined design^[26].

Lastly, the study on the coupling of multiple fields and long-term performance has a deficiency in that there is "abundant short-term testing but weak long-term predictive capability." Various studies have focused on the degradative influence of factors like wet-dry cycles, freeze-thaw cycles^[11], moisture content, and compaction degree^[12] on the interface properties. In addition, some studies have drawn long-term performance laws based on full-scale monitoring results, while other studies have carried out long-term behavior studies through numerical simulations based on creep and environmental factors. However, there is still a lack of theoretical models for extrapolating short-term environmental impacts.

These research gaps have resulted in three engineering practice challenges related to the design of geogrid-reinforced slopes. Firstly, the difficulty in utilizing the results of interface research lies in the fact that the results are difficult to transform into engineering parameters. Secondly, the design of special soils is based on experience rather than theory. Thirdly, the lack of long-term performance prediction capability makes it impossible to assess the safety reserves of the slopes during the service period. Therefore, it is of great importance to organize the existing research results systematically, clarify the development strategies, and identify the key scientific issues in order to promote the theory of reinforced soils and push the innovation of

engineering applications.

Based on the above background, this paper reviews and synthesizes 40 typical studies on the stability of geogrid soil slopes through the following structure: In Section II, the research on the characteristics of the soil-geogrid interface is discussed, including the mechanisms, influencing factors, and patterns of special states; In Section III, the test results on the mechanical properties of reinforced soil are presented, including direct shear tests, triaxial tests, and long-term tests; In Section IV, the development and recent achievements in the stability analysis methods of reinforced soil slopes are analyzed; In Section V, the research trends and suggestions on the future research direction are presented.

2. Study on Soil-Geogrid Interface Characteristics

2.1 Theoretical Basis and Testing Methods

The interface interaction of the soil with the geogrids represents the heart of the entire reinforcement technology, as well as the most critical scientific problem associated with this technology. Already as early as the 1980s, fundamental work was carried out by Jewell & Milligan ^[1]. In this context, the study suggested a new approach to understanding the properties of the geogrids as continuous structures consisting of an infinite number of units capable of carrying loads, with the reinforcement effect of the geogrids being based on the properties of two types of resistance: friction resistance at the interface of the grid with the soil, as well as passive resistance of the grid's transverse ribs against the soil, a so-called "dual mechanism model" of understanding the principles of the interaction of the soil with the geogrids, which remains a fundamental theory until now.

Regarding the materials used, a systematic review was performed by Al-Barqawi and Shanableh in 2021 ^[2]. In their study, they focused on the analysis of the correlation between the properties of the materials used, design concepts, and structural configurations of the geogrids made from different types of polymers, including PET, PP, HDPE, etc. It was clearly stated that the type of used polymer, the method of production, and the form of the structure directly affect the mechanical properties and lifespan of the geogrids. For example, biaxial geogrids can be used in bidirectional loading conditions, while uniaxial geogrids are mainly used in reinforced earth retaining walls. Moreover, triaxial geogrids have the ability to provide isotropic bearing capacity. These findings provide significant references for the selection and design of materials in engineering applications. To explore the problems related to the interface, standardized test methods must be employed. Stacho et al. carried out tests using the large-scale direct shear test method to obtain the characteristics of the soil-grid interface's shear strength, as well as the correlation between the test conditions and the test results ^[3]. In the case of "asymmetric interfaces," i.e., when the soil properties on the two sides of the grid are different, Liu et al. carried out direct shear tests using the large-scale direct shear test method ^[4]. The test results show that the moisture content plays a significant role in the interface's strength when dealing with silty clay soils; the maximum value is achieved when the moisture content is optimal, and the interface's strength is increased with the compaction.

For the gravel-soil mixtures, Wang et al. found the following more specific quantitative patterns through the experiments they carried out: when the normal stress is not higher than 30 kPa, the peak interfacial stress increases linearly with the compaction degree; when the shear rate is 1.5 mm/min, the interfacial stress can reach its maximum value ^[5]. In addition, they found the following interesting phenomenon: with the increase in compaction, the "pseudo-cohesion" at the interface increases, and the "pseudo-friction angle" decreases.

2.2 Interface Anisotropy and Generalized Strength Equation

Past interface strength tests were carried out along a single direction, while engineering loads are normally not aligned along a single direction. To deal with this, a unique testing device was created, as presented by Zhang et al, which allows the direction of shear to be changed during a test. Shear tests were carried out on bidirectional and tridirectional geogrids at different angles: 0°, 45°, and 90° ^[6].

The results show a significant finding in the following way: under the same normal stress conditions, the interface's shear strength depends on the direction of the shear and thus directly confirms the high level of anisotropy in the interface between the soil and the grid. Further analysis has shown that the anisotropy in the interface between bidirectional grids and the soil is more significant than the anisotropy in the interface between the soil and the triaxial grids. Moreover, the size of the soil's particles plays an important role in this anisotropy.

Udomchai et al. changed the focus of their research to recycled materials and carried out large-scale direct shear tests on the interface characteristics between recycled asphalt pavement (RAP) and jute geogrids ^[7]. They identified the parameter D/FD (where FD is the content of recycled particles less than the opening of the geogrid) as the main parameter governing the interface strength and established the generalized interface strength equation as $\alpha = a(D/FD) + b$. The constants in the equation were found to be dependent only on the type of recycled material and not the type of geogrid material. The equation can be used as a valuable tool in the design calculation for the use of recycled materials in pavement engineering.

2.3 Effects of Cyclic Loads and Environmental Factors

During the service life of the structure, the structure will inevitably be subjected to cyclic loading such as traffic loading and seismic loading. The changes in the properties of the interface subject to cyclic loading have a direct influence on the long-term stability of the structure. In the study by Jiang et al. the influence of the gradation of the soil on the dynamic shearing properties at the interface between the coarse-grained soil and the geogrid was investigated through cyclic direct shearing tests ^[8]. It was concluded that if the content of the coarse-grained soil is high, the shearing stiffness K increases with the increase in the maximum particle diameter d_{max} , while the damping ratio D decreases. Optimal results were obtained at a ratio of 0.073.

In addition, Zhang et al. studied four-directional geogrids with regard to their cyclic shear behavior at the interface with sandy soils ^[9]. It is evident from the findings of this study that the peak interface strength is directly proportional to the increasing normal stress and the displacement amplitude of the shear. Furthermore, it is evident that increasing the normal stress and the displacement amplitude reduces the number of cycles to peak interface strength. In addition, Ferreira et al. studied the interface between HDPE geogrids and residual soils, revealing different evolutionary patterns of the interface under cyclic loading ^[10].

Besides mechanical loading, environmental factors are also important considerations in the engineering design of cold regions and areas with alternating wet and dry conditions. The effects of wet-dry cycles and freeze-thaw cycles were systematically studied by Zhou et al. ^[11]. The results indicated that these two types of cyclic loading can cause the mechanical properties of the geogrid material to deteriorate, which in turn affects the synergy between the soil and the geogrid material. In addition, the deterioration becomes more severe with the number of cycles.

In the case of loess, which is widely distributed in China, Zhang conducted research on this material ^[12]. In this regard, it was revealed from the experiments conducted by Zhang that when the moisture content does not exceed the plastic limit, a higher moisture content results in a shift in the curve from a softening type to a hardening type in the shear stress-shear displacement curve. At the same time, a significant reduction in cohesion and friction angle at the interface occurs. In regard to compaction degree, when this degree reaches 90%, a significant increase in thickness in the zone of shearing occurs. Therefore, in reinforced loess engineering, this degree preferably should not be lower than 90%.

2.4 Particle Morphology and Micromechanism

In recent years, the influence of the morphological properties of soil particles on interfacial shear properties has become a popular topic in the field of academia. In the study carried out by Li et al., a comprehensive method of direct shear tests and discrete element method simulations was used to investigate the influence of particle size and roundness of the gravel particles on the interface of the gravel grid ^[13].

Their results were conclusive; the reinforced soil interface displayed its characteristic shear softening behavior. Peak and residual shear stresses decreased by 15% to 16% as the particle roundness increased from 0.35 to 0.75. On the other hand, as the particle size increased from 5 mm to 13 mm, the internal friction angle increased by 10.8%, and the apparent cohesion increased significantly by 144%. At the micron-scale, they observed a unique "peanut-shaped" distribution of shear contact forces, and the tangential contact forces displayed a "petal-shaped" distribution.

In a study by Wang et al., the researchers emphasized the content of fine grain materials by evaluating its impact on the shearing behavior of geogrid-reinforced gravelly soil ^[14]. The findings showed that with increasing content of fine grain materials in the soil, its peak strength and shearing parameters initially increased and then decreased. At the same time, a significant relationship was established in which a lower content of fine grain materials coupled with a normal stress led to

higher shearing dilatancy at the interface.

3. Reinforced Soil Mechanics and Experimental Studies

3.1 Triaxial Tests and Strength Characteristics

Triaxial compression tests still represent the most fundamental and widely used approach for the study of the mechanical properties of reinforced soils. In the study by Zakarka et al., the influence of density and the non-uniformity coefficient was investigated on the mechanical properties of single-layer geogrid-reinforced sand and gravel soils through consolidated drained triaxial tests ^[15]. From the results obtained in the study, it is evident that the pressure has a critical influence on the effectiveness of the reinforcement process, with higher pressure resulting in the reduction of the effectiveness of the grid. Moreover, the deviatoric stress values for the gravel soils are higher compared to the sandy soils. The strength parameters of the soils have shown that the grid has the greatest influence on cohesion but results in a reduction in the friction angle in most test conditions.

For the special fill material of calcareous gravelly sand, the results of triaxial tests conducted by Liu et al. are as follows: the effectiveness of reinforcement in strengthening the soil is obvious, and the fragmentation of particles has a significant negative influence on the strengthening effect of reinforcement ^[16]. In the study of layered reinforcement, Cui et al. tested composite materials with the following components: geogrid, sand, and clay ^[17]. It is shown in the study that the layered arrangement of reinforcement has a significant influence on the performance of composite materials. The advantages of different materials can be fully utilized through the design of layered composite materials in a reasonable way.

With the improvement of experimental methods, 3D printing technology has also been used in the construction of geogrids. Shi et al. used 3D printing technology to prepare triaxial geogrids, which were used to study the mechanical properties of reinforced marine coral sand-clay mixture materials ^[18]. It was found that smaller amounts of clay, more reinforcement layers, and higher confining pressures significantly improved the mixture's strength and cohesion. In addition, a linear relationship was found between the rate of particle breakage and fractal dimension.

From the numerical simulation point of view, Liu used the three-dimensional discrete element method to investigate the macro-micro response of the reinforced coral sand subjected to high confining pressure ^[19]. It was observed that the mechanical responses of the reinforced and unreinforced specimens showed significant differences at high confining pressure, with the reinforced specimens displaying obvious hardening characteristics and suppressing the phenomenon of strain softening and swelling deformation. In addition, Zakarkaa and Skuodis put forward a parameter calibration method for the grid-reinforced soil model based on the triaxial test results and identified the sensitive parameters that provided parameter support for the subsequent numerical simulation ^[20].

3.2 Large-Scale Triaxial Tests and Coarse-Grained Soils

For engineering practice, such as highway and railway embankments, reinforcement of coarse-grained soils is a common practice. In this area, small-scale triaxial tests are no longer suitable, and the only method used in this area of research is the large-scale triaxial test.

Liu et al. took three typical gravel soils from Xinjiang and carried out large-scale triaxial tests to study their reinforcing effect ^[21]. The results showed that the internal friction of the soil did not change significantly before and after reinforcement, which means it remained constant. The increase in strength mainly resulted from the increase in cohesion between the soil particles. Regarding volumetric strain, it decreased significantly with the increase in confining pressure for reinforced gravelly soil. Moreover, with the increase in confining pressure, the difference in volumetric strain between reinforced and unreinforced soil became greater.

An in-depth study on the behavior of the shear response of calcareous gravelly sand was carried out by Chen et al. ^[22]. The study compared the results obtained through large-scale triaxial tests and discrete element simulations, which indicated that the soil has stronger resistance to strain softening compared to calcareous coarse sand; however, it is also found that the soil is prone to fragmentation. Regarding the reinforcement mechanism, it is found that the main mechanism of reinforcement using geogrid is that it restricts the lateral movement of the soil, which is equivalent to the suppression of volumetric shear swelling behavior of the soil, which in turn increases the peak value of the shear stress.

3.3 Reinforcement of Clayey Soils and Special Soils

In recent years, studies on reinforcement using clayey soils as fill materials have gradually become a focus of academic attention. In the study conducted by Wang et al., the authors carried out a systematic investigation of the mechanical properties of clay soils reinforced with geogrids using triaxial tests and discrete element simulations ^[23]. The results showed that the peak strength and cohesion of the reinforced soil can be significantly enhanced with an increasing number of reinforcement layers, whereas the effect on the internal friction angle is limited. Meanwhile, it is verified that the stress-strain relationship of reinforced soil follows the classical Duncan-Zhang model. Through discrete element micromechanical analysis, it is found that the reinforcement can reduce the fluctuation of soil porosity, significantly improve the interparticle contacts, and densify the soil structure.

With regard to the reinforcement layers' influence, minor differences were noted depending on the soil type used. In their study on microgrid-reinforced sandy soil, Teles Júnior et al. noted that an increase in reinforcement layers significantly increases the soil's strength parameters, even causing a change in the soil's stress-strain curve from strain softening to strain hardening ^[24]. On the other hand, in their study on granitic residual soil, Deng et al. noted a "marginal effect" in the reinforcement influence, whereby the reinforcement effect of the two layers was not in a simple linear relationship, and the reinforcement effect of the second layer showed a diminishing trend ^[25]. Nevertheless, the studies by both authors validate an important fact: the fundamental principle governing geogrid reinforcement in soil is an increase in cohesion, thereby justifying the quasi-cohesion increment formula.

3.4 Advanced Manufacturing Technologies and Novel Materials

Significant developments have been observed in the manufacturing domain, providing new means of conducting advanced studies on geogrids through innovative means such as 3D printing. Oliveira and Falorca avoided the usage of existing pre-fabricated grids by designing test specimens with different apertures and stiffness values through CAD and 3D printing ^[26]. In the study, the mechanical properties of two-layered soil consisting of a base layer of granite aggregate and a subgrade layer of sandy soil were examined.

The results of their research produced a very pragmatic conclusion: the reinforcement effect is often optimal when the ratio of the grid aperture and the median soil particle size is close to 1. Moreover, they found nonlinear characteristics of interfacial load transfer, indicating that the optimal interfacial behavior is not necessarily dependent on the grid's stiffness.

With regard to numerical model application, Zakarka et al. calibrated the hardened soil model parameters based on experimental results obtained during a triaxial test ^[27]. The study showed that specific values of the power index 'm' and the unloading-reloading modulus 'Eur' are required to improve the model's accuracy in matching experimental results, which has important engineering implications.

3.5 Macro-Micro Correlation and Particle Fragmentation

To clarify the reinforcement mechanism, it is crucial to establish the link between the macro-mechanical response and micro-particle behavior. Luo et al. specifically focused on this, using the discrete element method to extensively study the macro-micro mechanical behavior of geogrid-reinforced calcareous sand under three-dimensional loading ^[28].

The findings indicate that reducing the aperture size of the geogrid and increasing its tensile strength both enhance the macro-mechanical properties of the reinforced body. At the microniveau, particle failure is primarily governed by shear strain and confining pressure, while the confining effect of the grid partially suppresses soil swelling deformation. By examining the interaction between granular particles and the grid, this study provides a clear explanation for the intrinsic source of strength enhancement in reinforced soil, offering valuable microniveau insights for a deeper understanding of the reinforcement mechanism.

4. Numerical Simulation and Engineering Applications

4.1 Macro-Scale Finite Element Analysis

Finite element analysis has become a crucial numerical tool in the design of reinforced soil structures. Sundaravel and Dodagoudar employed the finite element method to analyze the performance of reinforced soil retaining walls, comparing and evaluating the applicability of different constitutive models such as Mohr-Coulomb and hardening soil models ^[29]. This

provides a reference basis for subsequent numerical simulations of reinforced soil structures.

Samal et al. performed static stability analysis on bamboo grid-reinforced slopes using MIDAS GTS NX finite element software, exploring the feasibility of applying such green reinforcement materials in slope stabilization projects ^[30]. Onur et al. also employed the finite element method to investigate the enhancement of overall slope stability through geogrid reinforcement, conducting a quantitative analysis of the relationship between reinforcement parameters and safety factors ^[31]. Systematic research on the reinforcement mechanism and design methods for the tension type of reinforced earth retaining walls has been carried out by Sun et al. ^[32]. The structural responses of the geotextile strips and the geogrids as the reinforcement materials were compared in the research carried out by Sun et al. The results show that the tension type of reinforced retaining walls can reduce the vertical stresses in the walls and the earth pressures on the wall panels.

In the control of the horizontal deformation of the walls, the wrapped geogrids are found to be more effective compared to the geotextiles. However, in terms of safety, deformation, cost, and difficulty in the construction, the economic efficiency and simplicity in the construction process of the walls using the geotextiles are more effective. Xu et al. carried out the analysis on the displacement characteristics of reinforced earth retaining walls with super-high fill slopes subjected to the loads from the structures on the crest using the FLAC3D software ^[33]. The analysis resulted in the findings on the significant differential settlement in the wall and the slope behind the wall. This would be highly useful in the layout and designing of the structures on the crest. Yoo et al. carried out the research on the reinforcement efficiency of the geogrid-reinforced slopes using the finite element analysis method ^[34]. They validated the reliability of the model using the field measurement data.

4.2 Multi-Scale and Multi-Field Coupled Analysis

Computational advancements have allowed for multi-scale simulations, which can perform cross-scale studies from microscopic interfaces to macroscopic structures. Zhao et al. carried out multi-scale numerical simulations using a coupled finite element and discrete element approach for the analysis of reinforced soils using geogrids, and the intrinsic mechanisms of reinforcement were studied at the mesoscopic scale ^[35]. It can thus be shown that multi-scale models can simulate the macro-micro coupling in reinforced soils quite well. Reinforcement of soils can modify the internal force chains and anisotropy of particle contacts in soils, and these are the microscopic mechanisms for the macroscopic strength of reinforced soils.

Multiphysics coupling analysis is also a research hotspot in this area. Ng et al. carried out a research on the response characteristics of unsaturated soil-reinforced slopes under hydraulic-mechanical coupling conditions through numerical analysis and centrifuge tests ^[36]. It is found that the hydraulic coupling has a great effect on the stability of the unsaturated reinforced slope under rainfall conditions; the reinforcement measures can well control the problems of deformation under the condition of suction loss; and the results of the centrifuge tests verify the rationality and reliability of the numerical model.

4.3 Long-Term Performance and Field Monitoring

Long-term performance studies are of great importance in the development of a solid base for the achievement of full-life-cycle design of reinforced soil structures. Bathurst and Allen discussed the long-term performance characteristics of geosynthetic-reinforced retaining walls and slopes based on field monitoring results of full-scale structures, with a focus on the importance of rational design and construction quality for long-term structural stability of reinforced soil structures ^[37].

Abdi and Zandieh studied the long-term mechanical performance of reinforced soil structures subjected to combined material creep and environmental effects through numerical simulations and proposed relevant long-term performance prediction methods ^[38].

Gaur et al. carried out a comparison of the characteristics of the pull-out resistance of various geogrids through the integration of the results of the field pull-out tests and numerical simulation ^[39]. It has been found that the M65 geotextile strips show 20% higher initial pull-out strength in sandy soils compared to the biaxial geogrids. However, the ultimate pull-out strength is found to be 8% less in the M65 geotextile strips. It has been found that the geotextile strips are more appropriate for use in the construction of low-grade roads and other backfill works, while the geogrids show more advantages in the construction of high-grade roads.

Gör et al. studied the improvement in the stability of the slope through the use of geogrid reinforcement by model tests and numerical analysis ^[40].

4.4 Engineering Application Cases

The applicable theoretical and experimental results have been used in various engineering practices. To illustrate, Reference ^[32] carried out optimization of essential design parameters of tension-reinforced retaining walls in soil, based on the UAE's heavy haul double track high-speed railway project. Reference ^[33] performed displacement analysis of super-high embankment slopes with building loads at the crest, which was used as a basis for choosing a building site. Reference ^[39] carried out in-situ pull-out tests of the NH148B National Highway project in India, verifying the applicability of reinforced solutions with geotextile.

Collectively, the engineering examples show that the theoretical results of laboratory testing, numerical simulation, and theoretical analysis are gradually transforming from theoretical study into engineering practice, offering strong backing for the advancement of reinforced soil technology.

5. Conclusions

This paper focuses on the stability of geogrid-reinforced earth slopes. By systematically summarizing and synthesizing 40 core papers, the recent research achievements, existing difficulties, and development directions in the future can be comprehensively outlined as follows:

- (1) Research on the characteristics of the interface between soils and geogrids has formed a comprehensive research framework. This framework includes the theoretical basis, the development of test methods, the exploration of characteristic laws, and the in-depth study of the micro-mechanisms. In addition, the academic community has gained a relatively clear understanding of the influence of factors such as interface anisotropy, cyclic response, environmental degradation effects, and particle morphology on the interface. However, the key difficulty lies in the fact that these advanced micro-mechanisms have not been effectively integrated into the analysis framework at the macro-level.
- (2) The research on the strength characteristics of reinforcing elements has shown a clear path of extension in the following aspects: expanding from ordinary soils to special soils while incorporating advanced technologies to build connections between the macro and micro scales. A large number of triaxial test results have confirmed the basic mechanism of the reinforcement effect on the strength characteristics of soils—increasing cohesion. At the same time, the particle fragmentation characteristics in special soils such as calcareous sand and coral sand are hot research fields. It is noteworthy that the application of 3D printing technology has provided a new means for improving the research on the parameters such as grid size and stiffness.
- (3) Numerical simulation techniques are progressing rapidly. Analytical techniques have developed from traditional macro-scale finite element analysis to FEM-DEM multiscale coupling and hydromechanical multiphysics coupling analysis. At the engineering application level, the emergence of green materials such as bamboo grids and new engineering structures such as tension-supported retaining walls has greatly increased the engineering applicability of reinforcement technology. The use of field monitoring combined with data assimilation techniques has greatly improved the reliability of numerical simulation models.
- (4) In the aspect of the study on the long-term performance, the industry has gradually shifted its focus from the traditional empirical estimations to the full-scale monitoring and data assimilation methods. Nevertheless, the prediction models that can take into account the effects of multi-field coupling in the long-term performance analysis have not yet been developed. Besides, the studies on the pull-out resistance and the long-term durability of new types of reinforcement materials, such as geotextile tapes, are still in the primary stage and need to be further explored.
- (5) Considering the present research status and existing gaps, the following four aspects are the direction for the future research: Firstly, the establishment of theoretical models with accurate micro-macro correlations; secondly, the promotion and establishment of the design theory for the reinforcement of certain soils; thirdly, the establishment of a science-based method for predicting the service performance; and fourthly, the promotion and development of the research and engineering applications of green reinforcement materials so as to provide theoretical guidance for the development of geogrid reinforcement technology.

Funding

This work was supported by the Xi'an Si yuan University President's Fund Research Project (Project No. XASYB24ZHD05). Investor: Xi'an Siyuan University, Investment Amount: 60,000 RMB.

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Jewell, R.A., & Milligan, G.W.E. (1989). Interaction between soil and geogrids. *Proceedings of the 12th International Conference on Soil Mechanics and Foundation Engineering*, Vol. 2, pp. 1419–1422.
- [2] Al-Barqawi, M., & Shanableh, A. (2021). Polymer geogrids: A review of material, design and structure (PET, PP, and HDPE geogrids). *Polymers*, 13(16), 8399111.
- [3] Stacho, J., Sulovska, M., & Slavik, I. (2020). Determining the shear strength properties of a soil-geogrid interface using a large-scale direct shear test apparatus. *Periodica Polytechnica Civil Engineering*, 64(4), 989–998.
- [4] Liu, K., Zhang, L., Wang, X., & Zhao, Y. (2019). Large-scale direct shear tests on soil-geogrid interface behavior. *Chinese Journal of Geotechnical Engineering*, 41(4), 185–196.
- [5] Wang, J., Qi, H., Huang, S., & Tang, Y. (2022). Large-scale direct shear test study of interface interaction between geogrid and gravelly soil mixture. *Hydrogeology and Engineering Geology*, 49(4), 81–90.
- [6] Zhang, J., Liu, Q., & Wang, Z. (2021). Anisotropic shear behavior of soil-geogrid interfaces. *Applied Sciences*, 11(23), 11387.
- [7] Udomchai, W., Arulrajah, A., Rahman, M.A., et al. (2021). Generalized interface shear strength equation for recycled materials reinforced with geogrids. *Sustainability*, 13(16), 9446.
- [8] Jiang, M.J., Shi, J.Y., Li, S.Y., Hu, R.F., & Mei, G.X. (2024). Experimental study on the influence of gradation on cyclic shear characteristics of coarse-grained soil-geogrid interface. *Water Resources and Hydropower Engineering*, 55(3), 162–172.
- [9] Zhang, M., Ruan, X., & Jiang, L. (2024). Experimental study on cyclic shear performance of the four-way geogrid reinforcement-soil interface. *Applied Sciences*, 14(4), 1373.
- [10] Ferreira, F.B.F., Vieira, C.S.D.L., & Queirós, P.G.M. (2023). Effects of cyclic loading on soil-geogrid interaction characteristics. In *Soil-Geogrid Interaction under Cyclic and Post-Cyclic Pullout Loading Conditions*. University of Porto Repository.
- [11] Zhou, Y., Guo, L., & Wang, Y. (2020). Effects of wetting-drying and freeze-thaw cycles on the tensile properties of geogrids and soil-geogrid interface behavior. *Geotextiles and Geomembranes*, 48(6), 823–835.
- [12] Zhang, W. (2025). Study on shear mechanical properties of loess-geogrid interface. *Chinese Journal of Applied Mechanics*. ChinaXiv:202512.00076.
- [13] Li, J., Liu, F., Gao, C., & Jin, Y. (2025). Test and DEM study on the influence of particle morphology on the shear characteristics of the reinforcement-soil interface. *Science Press*.
- [14] Wang, J., Lin, H., Tang, Y., & Tang, Y. (2025). Effect of fine particle content on shear characteristics of geogrid-reinforced gravelly soil. *Hydrogeology & Engineering Geology*, 52(4), 264–272.
- [15] Zakarka, M., Skuodis, Š., et al. (2023). Triaxial test of coarse-grained soils reinforced with geogrids. *Applied Sciences*, 13(22), 12480.
- [16] Liu, J., Zhang, H., & Wang, Y. (2024). Study on the shear and deformation characteristics of geogrid-reinforced calcareous gravelly sand via triaxial tests. *Frontiers in Earth Science*, 12, 1287718.
- [17] Cui, L., Ma, T., & Xue, J. (2021). Analysis of layered geogrids-sand-clay reinforced composites under triaxial compression tests. *Applied Sciences*, 11(21), 9952.
- [18] Shi, D., Xu, K., Chao, Z., Cui, P., & Zheng, J. (2025). Experimental study on mechanical properties of triaxial geogrid-reinforced marine coral sand-clay mixtures based on 3D printing technology. *SSRN*.
- [19] Liu, L. (2024). Effect of high stress levels on the shear behavior of geosynthetic-reinforced coral sand under triaxial

- conditions. *Journal of Marine Geomechanics*, 12(11), 2081.
- [20] Zakarkaa, M., & Skuodis, Š. (2024). Geogrid reinforced soil model calibration based on triaxial compression tests and numerical simulation. *ISSMGE Conference Paper*.
- [21] Liu, J., Pan, J., Wang, B., Hu, C., & Liu, Q. (2024). Study on the shear and deformation characteristics of geogrid-reinforced gravelly soils based on large-scale triaxial tests. *Frontiers in Earth Science*, 12, 1287718.
- [22] Chen, J.F., Bao, N., Ma, C., & Sun, R. (2024). Triaxial behavior of unreinforced and geogrid-reinforced calcareous gravelly sand: Experiment and discrete element modelling. *Construction and Building Materials*, 430, 136405.
- [23] Wang, X., Hu, Q., Liu, Y., & Tao, G. (2024). Triaxial test and discrete element numerical simulation of geogrid-reinforced clay soil. *Buildings*, 14(5), 1422.
- [24] Teles Júnior, R.C.B., Soares, A.B., & Aguiar, M.F.P. (2025). Drained tests on sandy soils reinforced with microgrids. *Soils & Rocks*, 48(1), e2025003222.
- [25] Deng, X.W., Duan, W.J., Yang, X.Q., & Zhu, Z.Y. (2024). Testing and mechanism analyses on the shear strength characteristics of reinforced granite residual soil. *Science Technology and Engineering*, 24(22), 9535–9546. <https://doi.org/10.12404/j.issn.1671-1815.2306170>.
- [26] Oliveira, G.M., & Falorca, I.M.C.F.G. (2025). Evaluation of two-layered soils reinforced with 3D printed geogrid models under axisymmetric loading conditions. *Geotextiles and Geomembranes*, 53(3), 798–810.
- [27] Zakarka, M., Skuodis, Š., & Kuhlmann, J. (2024). Geogrid reinforced soil model calibration based on laboratory testing. *Proceedings of the 28th European Young Geotechnical Engineers Conference (EYGEC 2024)*, pp. 343–350.
- [28] Luo, Z.G., Ding, X.M., Ou, Q., & Lu, Y.W. (2024). Macro-microscopic mechanical behavior of geogrid reinforced calcareous sand subjected to triaxial loads: Effects of aperture size and tensile resistance. *Geotextiles and Geomembranes*, 52(3), 456–472.
- [29] Sundaravel, V., & Dodagoudar, G.R. (2024). Finite element analysis of reinforced earth retaining structures: Material models and performance assessment. *Indian Geotechnical Journal*.
- [30] Samal, R.N.S.D., Sahoo, S., & Badavath, N. (2023). Static stability analysis of bamboo grid-reinforced slopes using finite element method (MIDAS GTS NX). *Engineering Research Express*, 5(4), 045073.
- [31] Onur, E., Keskin, M.E., & Laman, M. (2016). Effect of geogrid inclusion on slope stability using finite element analysis. *Proceedings of the International Conference on Interdisciplinary Research, Valencia, Spain*.
- [32] Sun, S.X., Yang, G.Q., Ma, T.M., Li, X.J., & Gao, P.L. (2025). Numerical simulation study on structural behavior of opposite-pull reinforced earth retaining walls. *Journal of Jilin Jianzhu University*, 42(1), 51–59.
- [33] Xu, J.C., Wang, Y.S., Zheng, J.H., Shen, M.L., & Jiang, T. (2025). Displacement characteristics of reinforced earth retaining wall slopes under building loads on slope top. *Hans Journal of Civil Engineering*, 14(1), 208–216.
- [34] Yoo, S.E., Kim, S.Y., Jung, Y.J., Jeon, H.Y., & Ko, E.H. (2024). Numerical analysis of geogrid effects on geogrid-reinforced slope. *Journal of the Korean Geosynthetics Society*, 23(2), 45–53.
- [35] Zhao, J., Zhang, F., & Liu, J. (2023). Multi-scale modeling of geogrid-reinforced soils using coupled FEM–DEM approaches. *Acta Geotechnica*, 18(6), 2451–2465.
- [36] Ng, C.W.W., Zhou, C., & Liu, H. (2022). Coupled hydro-mechanical behaviour of unsaturated soil slopes reinforced with geogrids: Numerical analysis and centrifuge validation. *Computers and Geotechnics*, 145, 104679.
- [37] Bathurst, R.J., & Allen, T.M. (2022). Long-term performance of geosynthetic-reinforced soil walls and slopes: Lessons from full-scale monitoring. *Geotextiles and Geomembranes*, 50(6), 1152–1170.
- [38] Abdi, M.R., & Zandieh, A.R. (2017). Numerical modelling of long-term behaviour of geosynthetic-reinforced soil structures considering creep and environmental effects. *Computers and Geotechnics*, 90, 125–137.
- [39] Gaur, K., et al. (2025). A comparative study of pullout strength of geostrips and implications for long-term durability. *Applied Sciences*, 15(14), 7715.
- [40] Gör, M., Taher, N.R., Aksoy, H.S., & Awlla, H.A. (2022). Effect of geogrid inclusion on slope stability. *Proceedings of the 5th International Conference on Interdisciplinary Scientific Research*.

Automatic Modulation Classification using Residual Connection and Bottle-neck Transformers with Trend-aware Self-Attention

Geng Wang^{1,3*}, Luming Li², Xin Chen^{1,3}

1. Beijing Institute of Satellite Information Engineering, Beijing, 100095, China

2. Beijing Institute of Spacecraft System Engineering, Beijing, 100094, China

3. State Key Laboratory of Space Information System and Integrated Application, Beijing, 100095, China

*Corresponding author: Geng Wang, wg312wg@gmail.com

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: With the increase of modern communication equipment, the demand for AMC (Automatic Modulation Classification) in communication systems is increasing. Since deep learning was introduced into AMC, people have been working on improving the recognition accuracy and robustness of AMC. To achieve this, the study proposes the design of a ResDBoTTA (Residual DSC Bottleneck Transformers with Trend-aware Attention) AMC Network for communication signal modulation pattern classification. The DSC (Depthwise Separable Convolution) used in the model can significantly reduce the model parameters. The introduced time trend-aware self-attention mechanism can eliminate the influence of abnormal noise. Finally, global deep convolution is applied to enhance recognition accuracy. To verify the classification performance of ResDBoTTA AMC Net, relevant simulation studies are carried out. Experimental results demonstrate that ResDBoTTA AMC Net achieves superior recognition performance compared to existing technologies.

Keywords: Automatic Modulation Classification; Deep Learning; Depthwise Separable Convolution; Bottleneck Transformers

Published: Apr 15, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1262>

1. Introduction

1.1 Research Background and Importance

In modern wireless communication systems, Automatic Modulation Classification (AMC) is one of the core technologies for signal demodulation and identification. Widely used in spectrum monitoring^[1], electronic countermeasures, MIMO systems^[2], Internet of Things (IoT)^[3], next-generation 6G networks^[4] and other fields. The goal is to quickly and accurately identify the modulation mode of the signal by analyzing the time-frequency characteristics of the received signal, thus laying the foundation for sub-sequent signal demodulation and information extraction. However, as the communication environment becomes increasingly complex (for example, due to multi-path fading and non-Gaussian noise interference^[5]), traditional AMC methods that are based on manual feature extraction (for example, high-order statistics^[6], constellation diagram analysis^[7] and wavelet transform^[8], etc.) are faced with problems relating to insufficient feature robustness and limited generalization capability. The performance is shown to be especially susceptible to significant degradation when the signal-to-noise ratio is at a low level^[9].

In the past few years, deep learning has become a novel approach to the challenges of AMC. O'Shea et al.^[10] pioneered the application of convolutional neural networks (CNN) to the AMC task, attaining an 89% accuracy rate for 12 modulation types. This study sub-stantiated the efficacy of CNN in facilitating local feature extraction. Subsequently, the re-searchers mitigated the gradient vanishing problem of deep networks and improved training stability by introducing residual connectivity (ResNet) for modulated signal iden-tification^[11]. However, conventional CNN face challenges in modeling the long-range temporal dependence of signals and exhibit sensitivity to local noise. To address these limitations, Zhou F et al.^[12] Zhang put forward a proposal for a CNN-LSTM network modulation identification scheme. This method combines the advantages of CNN and LSTM networks, using the periodic characteristics of modulated signals and employing a recurrent network to capture timing information. This approach significantly enhances the identification accuracy of signals in complex electromagnetic environments. Hamza et al.^[13] designs a bidirectional long short-term memory (BiLSTM) model to classify signals with different digital modulations.

This paper proposes a novel network architecture that fuses residual connected depth-separable convolution with a trend-aware self-attention mechanism. This architec-ture is referred to as ResDBoTTA (Residual DSC Bottleneck Transformers with Trend-aware Attention) AMC Network. It utilizes residual-connected deep separable con-volution (DSC)^[14] to enhance the practicality of deployment on industrial devices, and introduces a trend-aware self-attention mechanism and global deep convolutional feature reconstruction. The network gives us fresh perspectives on AMC systems that have excel-lent modulation recognition capabilities, strong stability and minimal computation. The major contributions of this work include the following:

A Depth separable convolution residual block (DSC residual block) is applied. This type of block addresses the gradient vanishing problem by employing jump connec-tion, and it reduces the computational complexity by integrating Depthwise Separa-ble Convolution (DSC) to facilitate efficient multi-scale feature extraction.

A time trend-aware multi-head self-attention mechanism (MHSA-TA) was intro-duced. The one-dimensional causal convolution was incorporated into the mul-ti-head self-attention (MHSA) module in BOT, thereby establishing a novel MHSA-TA module. This module employs local context information to discern the temporal trend of the signal, suppress noise interference on attention weights, and augment the model's noise-resistance.

Feature reconstruction using global deep convolution (GDConv) is proposed: instead of GPA, we use the GDConv module to spatially reconstruct the feature map through global deep convolution. This approach enables the adaptive learning of feature im-portance at varying locations, thereby enhancing the classification accuracy.

The proposed ResDBoTTA will be validated by evaluating its performance on the publicly available RadioML2016.10a and RadioML2018a dataset. The experimental results demonstrate that ResDBoTTA exhibits a high level of performance in these two datasets, achieving classification accuracies of 93.23% and 97.43% at a 12 dB signal-to-noise ratio. These results indicate that ResDBoTTA significantly surpasses the performance of existing methods. The ablation experiments further validate the advantages of the temporal trend-aware attention mechanism and GDConv.

2. Related Works

This section reviews early AMC methods and further reviews the recent research of AMC.

2.1 Traditional AMC methods

Most of the early AMR methods for electromagnetic signals are designed for Gaussi-an noise channels, and they are mainly based on the likelihood based (LB) modulation identification method. The LB algorithm optimizes the model identification accuracy ac-cording to the likelihood functions of different modulated signals (e.g., mixed likelihood ratio, average likelihood ratio^[15], generalized likelihood ratio^[16]), and it can be consid-ered that the LB algorithm ensures that the modulation identification results can get the optimal value under the Bayes' minimum misclassification cost criterion. It can be con-sidered that the LB algorithm ensures that the modulation identification result can get the optimal value under the Bayes minimum error cost criterion, but when there are more types of modulation signals, it directly leads to the increase of computational complexity of the algorithm, and it is sensitive to the mismatch of the model and the parameter devia-tion. With the advancement of technology, feature based (FB) modulation recognition methods have gradually become the

mainstream. FB algorithms have low computational complexity and can achieve better recognition results through reasonable design of feature extraction. Feature based recognition algorithms generally include feature extraction and classification network, and many researchers have studied various types of features for extraction, including high-order cumulative quantities (HOC)^[17], cyclic smoothness features^[18], approximate entropy^[19], etc., and the feature extraction algorithms can be used in a variety of ways. However, in the case of identification and classification of multiple types of modulated signals, the above methods generally suffer from the problems of cumbersome manual feature extraction and high computational complexity, and their recognition effect is poor in low signal-to-noise ratio environments.

2.2 Recent AMC Method Based on Deep Learning

After CNN and LSTM, the Transformer^[20] model has been shown to further enhance feature modeling capability by capturing long-range dependencies through the Self-Attention mechanism (Self-Attention). Srinivas A et al.^[21] initially proposed BoTNet (Bottleneck Transformer), a fusion network of CNN and Attention. This network replaces the 3x3 convolution operation in the middle of the Bottleneck layer, as implemented in ResNet50, with Multi-Head Self-Attention. This modification serves to reduce the amount of parameters by reducing the number of feature map channels in the middle layer. S. Lin et al.^[22] proposed a spectral CNN model built on the AMC Time-Frequency Attention mechanism to correlate the time and frequency domains of the signal. W. Zhang et al.^[23] proposed an AMC model based on CNN and Spatial Self-Attention mechanism to enhance the connection between channel and space. Liang Z et al.^[24] used multi-head attention in combination with CNNs to facilitate modulation recognition, thereby capturing the relationship between IQ signals. M Qi et al.^[25] increases the amplitude and phase characteristics of the signal while using the IQ signal. Y. Feng et al.^[26] added the Gramian angular field as a third feature input in addition to the two IQ signals used for AMC. Zhang et al.^[27] combine CNN with Transformer and performed a sliding window process to fuse the IQ information prior to signal input. R. Duan et al.^[28] used multi-modal inputs such as constellation diagrams to improve accuracy at high SNRs, and used only IQ signals at low SNRs to ensure accuracy. Z. Luo et al.^[29] utilized a combination of CNN, LSTM and transformer algorithms to achieve optimal outcomes by employing multi-modal input signals. W. Kong et al.^[30] used the Transformer for self-supervised pre-training, after fine-tuning using a small number of labeled signal samples to accomplish AMC.

In general, CNN Self-Attention Mechanism models have achieved good performance for AMC. Nonetheless, these methodologies continue to exhibit deficiencies. BoTNet solves the problem of gradient vanishing during feature extraction in deep networks, but the model parameter scale remains a potential area for enhancement. Moreover, the traditional multi-head self-attention mechanism is sensitive to noise, and is prone to erroneous allocation of attention weights under outliers or local disturbances, thus reducing model robustness.

3. Signal Model and Proposed Model

3.1 Signal Model

In this part, the signal model for automatic modulation classification is shown. In wireless communication systems, the received signal is usually affected by multipath fading, noise interference, and nonlinear distortion. Assume that the received signal passes through an additive Gaussian white noise (AWGN) channel, its baseband equivalent signal model can be described as:

$$\mathbf{y}(t) = \mathbf{h}(t) * \mathbf{s}(t) + \mathbf{n}(t) \quad (1)$$

where $\mathbf{s}(t)$ is the transmitted modulated signal, $\mathbf{h}(t)$ is the channel impulse response, $\mathbf{n}(t)$ is the noise obeying the complex Gaussian distribution $\mathcal{CN}(0, \sigma^2)$, and σ^2 is the noise power.

For discrete sampled signals, assume a symbol rate of B . The sampling interval is $T_s = 1/B$ and the received signal can be expressed as:

$$\mathbf{y}[n] = \sum_{k=0}^{L-1} \mathbf{h}[k] \mathbf{s}[n-k] + \mathbf{n}[n] \quad (2)$$

where L is the channel memory length. The mathematical form of the modulated signal $\mathbf{s}[n]$ depends on the modulation

method. Take the example of common modulation types:

PSK (Phase Shift Keying):

$$\mathbf{s}[n] = e^{j(2\pi m/M + \phi_0)}, m \in \{0, 1, \dots, M-1\} \quad (3)$$

QAM (Quadrature Amplitude Modulation):

$$\mathbf{s}[n] = a_m + jb_m \quad (4)$$

(a_m, b_m) is the coordinates of the constellation points.

FSK (Frequency Shift Keying):

$$\mathbf{s}[n] = e^{j2\pi f_m n T_s} \quad (5)$$

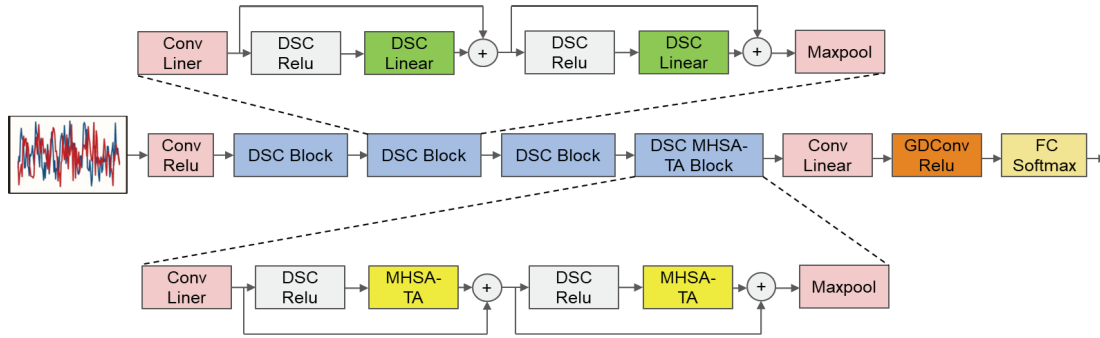
f_m is the discrete frequency point.

The datasets signal format used in the experiments is I/Q signals. The objective of modulation identification is to predict the kind of $\mathbf{s}[n]$ based on the received IQ signal.

3.2 ResDBOTTA model structure

The ResDBoTTA model proposed in this paper consists of the following three main modules: a feature extraction module using deep separable convolution with residual concatenation and a multi-head self-attention mechanism based on temporal trend sensing, a feature reconstruction module using a global deep convolution algorithm, and a classification module. The specific structure is shown in Figure 1.

Figure 1: The structure of ResDBOTTA model



3.3 Feature Extraction Layers

3.3.1 Residual Connected DSC Block

The jump connection method of residual network species can not only solve the problem of gradient vanishing, but also improve the network performance. The ResDBoTTA proposed in this paper designs a DSC residual block by introducing the jump connection method while using DSC instead of ordinary convolution. In order to extract the features of the signal, we first go through a linear convolution, then through two DSC residuals in parallel, and finally through a maximum pooling operation. Fig. 2 shows their structure.

The DSC operation process is comprised of two steps. Initially, the deep convolutional layer performs the convolutional operation on each channel of the input individually. Subsequently, the dot convolutional layer fuses all the channels. The operation of the deep convolutional layer is shown in Fig. 2. The deep convolution kernels are single-channel, and that the number of kernels is equal to the inputs. The height and width are still 5×5 . The point convolution layer is characterized by a convolution kernel height and width of 1 standard convolution layer, a single convolution kernel size of $1 \times 1 \times 3$, and a single point convolution kernel operation of the input and the output in the height and width of the dimension of the same, but the output into a single channel of the form of $8 \times 8 \times 1$, as illustrated in Figure 2.

3.3.2 Time trend-aware multi-head self-attention mechanism

The conventional self-attention mechanism calculates the similarity between the query and the key by point-by-point (point-wise) without leveraging contextual information. In the event that the observation at a given time step is an outlier, it will incorrectly match the relevant points. This results in an inaccurate weight assignment for self-attention. As illustrated in the left panel of Fig. 3, point A is incorrectly matched with point B, rather than being matched with the more similar point C. In this study, we propose a temporal trend-aware attention mechanism (MHSA-TA) to extract the local context information of

each node so that each node has the ability to perceive the context, and then construct an attention matrix by capturing the local trend. The proposed MHSA-TA model contributes to more accurate predictions and is more stable than the traditional attention mechanism.

Figure 2: DSC residual block and DSC operation

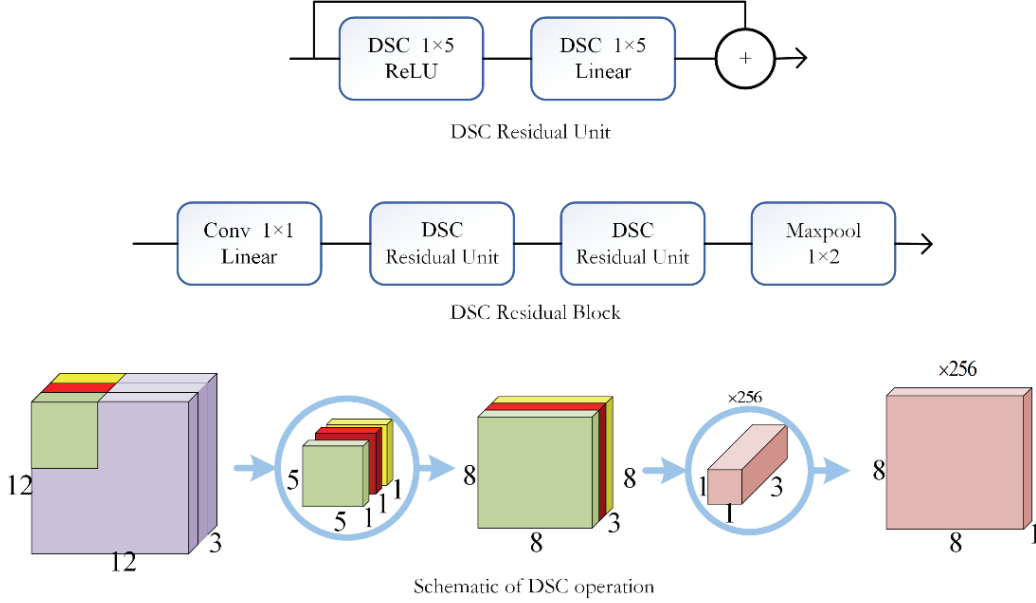
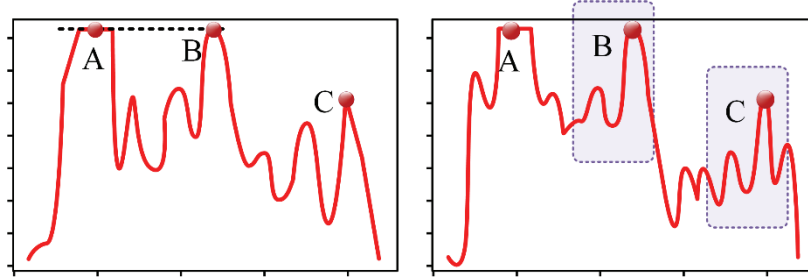


Figure 3: Conventional self-attention (left) and time-trend-aware attention (right)



The ResDBoTTA network employs a temporal trend-aware multi-head based self-attention mechanism module in the last residual block, replacing the conventional convolution operation. This modification enhances the modulation pattern recognition anti-noise capability, thereby functioning as a signal smoothing filter. The module has been shown to increase the robustness of the model in the face of noise and disturbances by reducing its sensitivity to outliers or extreme values. Furthermore, the module has been shown to reduce the mis allocation of weights and improve the generalization of the model over unknown data. In modulation recognition, time-trending self-attention ensures that the model maintains a high recognition accuracy even in the presence of poor signal quality.

MHSA-TA is a variant of MHSA that replaces the query and key projection operations in MHSA with 1D convolution, as shown in Fig. 4. The use of convolution here allows the contextual relationships of the time domain signals to be extracted as inputs to the MHSA, enabling our model to capture trends hidden in the time data series and to ignore anomalous changes. The definition of MHSA-TA is as follows.

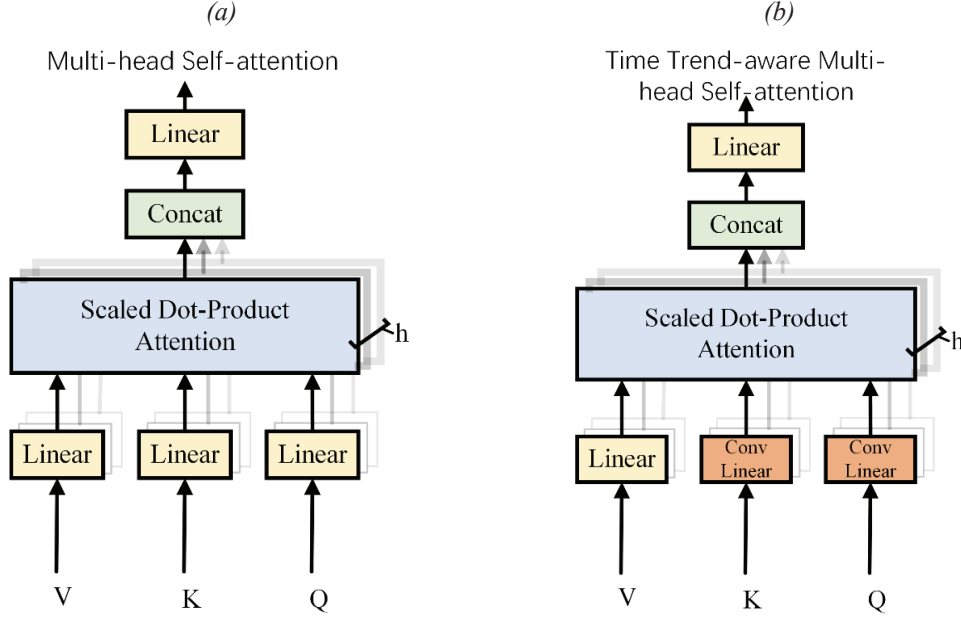
For an input $X \in \mathbb{R}^{N \times T \times C}$, where N represents the number of sequences, T denotes the sequence length, and C is the number of channels. The context relationship is captured by two 1D causal convolution of $1 \times k$. These layers generate the query Q and key K representations. Additionally, a separate 1×1 convolution layer is employed to generate the value V .

$$\begin{aligned} Q &\in \mathbb{R}^{N \times T \times C} = \text{Conv}_{1 \times k}(X) \\ K &\in \mathbb{R}^{N \times T \times C} = \text{Conv}_{1 \times k}(X) \\ V &\in \mathbb{R}^{N \times T \times C} = \text{Conv}_{1 \times 1}(X) \end{aligned} \quad (6)$$

The matrix of Q, K, V is reshaped by the reshape operation to obtain a representation in the form of multi-head attention:

$$\begin{aligned}
\mathbf{Q} &\in \mathbb{R}^{N \times h \times T \times d} = \text{reshape}(\mathbf{Q}) \\
\mathbf{K} &\in \mathbb{R}^{N \times h \times T \times d} = \text{reshape}(\mathbf{K}) \\
\mathbf{V} &\in \mathbb{R}^{N \times h \times T \times d} = \text{reshape}(\mathbf{V})
\end{aligned} \tag{7}$$

Figure 4: The multi-head attention structure (a) and Time trend-aware multi-head self-attention structure (b)



Then, the attention matrix is computed in a localized contextual and normalized by softmax:

$$\mathbf{A} \in \mathbb{R}^{N \times h \times T \times T} = \text{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{d_k}}\right) \tag{8}$$

where h represents the number of attention heads, d represents the number of channels per attention head, and $C = h \times d$.

Finally, $\mathbf{V}$ is weighted and summed and restored to the same shape as the input:

$$\begin{aligned}
\mathbf{Y} &\in \mathbb{R}^{N \times h \times T \times d} = \mathbf{A} \cdot \mathbf{V} \\
\mathbf{Y} &\in \mathbb{R}^{N \times T \times C} = \text{reshape}(\mathbf{Y})
\end{aligned} \tag{9}$$

$\mathbf{Y} \in \mathbb{R}^{N \times T \times C}$ is the output of the MHSA-TA.

3.4 Feature Reconstruction

After DSC operation, in the feature reconstruction part, ResDBoTTA proposes to utilize GDConv operation to learn the importance of features at different locations, after which the classification ability of reconstructed features is enhanced by ReLU activation function and bias term. The global deep convolutional layer is a deep convolutional layer with a convolutional kernel dimension equal to the size of the input feature map. The output of the global deep convolutional layer can be expressed as:

$$\tilde{G}_m = \sigma_{\text{ReLU}}\left(\sum_{i,j} \tilde{K}_{i,j,m} \cdot \tilde{F}_{i,j,m} + \tilde{b}_m\right) \tag{10}$$

In Eq. (10) $\tilde{F}$ is the feature map (Last Feature Map, LFM) with size $W \times H \times M$; $\tilde{K}$ is a global deep convolution kernel with size $W \times H \times M$; $\tilde{b}$ is the bias term; $\tilde{G}$ is the output classification feature vector of size $1 \times 1 \times M$; (i,j) denotes the spatial location index; m denotes the channel index. Looking further, the size of the global deep convolution kernel $\tilde{K}$ is equal to the size of $\tilde{F}$ in the input LFM, The feature reconstruction operation occurs between the channels corresponding to $\tilde{K}$ and $\tilde{F}$. For a particular channel m , the values of $\tilde{K}_m$ and $\tilde{F}_m$ at the same position (i,j) are first multiplied, then all dot product results are summed plus a bias term $\tilde{b}_m$, and finally a classification feature is obtained by the output of the ReLU activation function. Adding an FC layer activated by Softmax (normalized exponential function) after feature reconstruction transforms the predictions into probabilistic outputs.

3.5 ResDBOTTA model parameters

ResDBoTTA is implemented based on three parts: the DSC residual block, the MHSA module and the global deep convolutional feature reconstruction module. The specific configuration of the network is shown in Table 1. Four DSC residual blocks are deployed in the feature extraction part, the MHSA module is utilized in the fourth DSC residual block instead of the original convolution. The residual units are convolved and passed through the MHSA module before being summed up, and finally go through the maximum pooling. A layer of standard convolution is deployed at the front and back of the four DSC residual blocks. The first two DSC residual blocks are designed with 32 convolution kernels to decrease both the parameter count and computational load. And in order to enhance the feature dimension and extract sufficient information from the current input for feature extraction, the number of convolution kernels in the first standard convolutional layer and the last two DSC residual blocks are both designed to be 64, i.e., from 32 classes of features to 64 classes of features.

Table 1: ResDBoTTA parameter configuration

Network layer	Output dimension	function
Input	2×1024×1	--
First Conv 2×5 ReLU	1×1024×64	Feature Extraction
DSC Residual Block	1×512×32	
DSC Residual Block	1×256×32	
DSC Residual Block	1×128×64	
DSC MHSA-TA Block	1×64×64	Feature Reconstruction
Conv 1×1 Linear	1×16×64	
GDCov 1×16 ReLU	1×64	
FC Softmax	24	

The loss function uses the categorical cross-entropy loss function `categorical_crossentropy` that works better with Softmax for classification tasks, and its definition can be expressed as:

$$Loss = - \sum_{i=1}^{output} y_i \cdot \log \hat{y}_i \quad (11)$$

From Eq. (11), we can see that the result is 0 when y_i is 0 and the correct result output exists when and only when y_i is 1. This loss function is mainly used for comparing the two probability distributions, and in the back side it is connected to Softmax to rescale the model output so that it has the correct attributes, and the network finally outputs the result through an FC layer activated by Softmax.

4. Experimental Results and Discussion

4.1 Dataset Description

The DeepSig RadioML2016.10A^[31] and RadioML2018.01A^[32] datasets were used in this paper, respectively. The DeepSig RadioML 2016.10A dataset contains 8 digital modulation methods and 3 analog modulation methods, with a total of 220,000 samples. The signal-to-noise ratio ranges from -20dB to 18dB with 2dB spacing. The data format is 2×128 IQ signals.

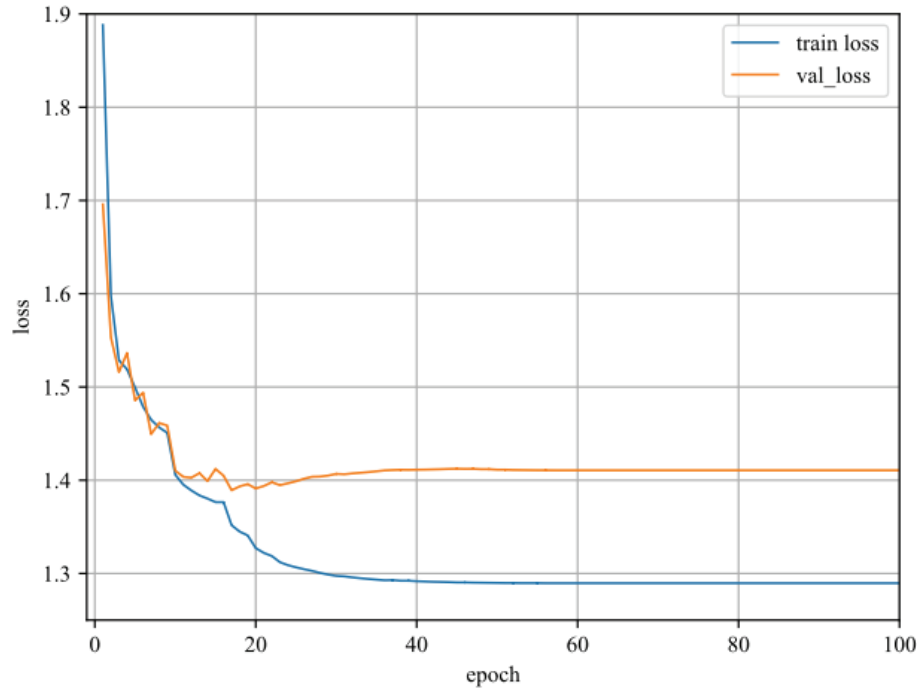
The DeepSig RadioML 2018.01A dataset contains a total of 24 signal modulation types, and there are 4096 data for each modulation type at each signal ratio in the range of signal ratios from -20dB to 30dB with a step value of 2dB, for a total of 2555904 data. The signal data is divided into I and Q and the data format is (1024,2). By sampling, 2064384 data points with signal-to-noise ratios from -20dB to 20dB are used, of which 1548288 are training samples and 516096 are test samples.

4.2 Experimental Settings

The ResDBoTTA network is trained with epoch of 100 and batch_size size of 512. Lazy Adam is used as the model optimizer during the training process instead of the commonly used SGD optimizer because in this structure the SGD optimizer tends to fall into local minima and has a slower rate of descent. Lazy Adam can handle sparse updates more efficiently. The ResDBoTTA network employs a categorical cross-entropy loss function, `categorical_crossentropy`, which is used to measure

the extent to which the model's predicted value $f(x)$ differs from the true value y . It guides the direction of the next step of training by calculating the gap between the forward computed result and the true value in each iteration. The loss function of the ResDBoTTA network when training the RadioML 2018.01.A datasets is shown in Figure. 5.

Figure 5. Loss function for RadioML 2018.01.A datasets



4.3 Baseline Models

Table 2. Baseline models of RadioML 2016.10A datasets

Datasets	Model	Parameters
RadioML 2016.10A	CNN	0.85M
	ResNet	3.1M
	MCLDNN	0.41M
	1DCMMPF	0.17M
	CNN-atten	0.24M
	ResDBOTTA	0.17M

CNN^[32] model: The classical CNN architecture proposed by O'Shea et al. contains 3 convolutional and 2 fully connected layers, with the input being a joint time-domain amplitude-phase representation of the I/Q signal. The model achieves 89% accuracy on 12 classes of modulated signals, but the timing dependence is not considered.

ResNet^[11] model: A deep convolutional network with residual connections is used to mitigate the gradient vanishing problem by jump connections. The model has a classification accuracy of 93.2% (SNR = 20 dB) for 24 classes of signals on the RML2018 dataset.

MCLDNN^[33]: The article proposes a spatio-temporal multi-channel learning framework. This model is able to receive different types of input data and process different spatio-temporal features through a multi-channel model.

CNN-atten^[24]: A radio signal recognition method based on the combination of convolutional neural network (CNN) and Self-Attention Mechanism is proposed in this paper. The accuracy of recognition is enhanced in low signal-to-noise conditions.

1DCMMPF^[34]: This article proposes an AMC model that combines the outputs of multiple CNN models to improve the accuracy and robustness of modulation recognition through parallel fusion.

The comparison baseline model of data set RadioML 2018.01A is shown in the Table 3:

Table 3. Baseline models of RadioML 2018.01A datasets

Datasets	Model	Parameters
RadioML 2018.01A	CNN	1.7M
	ResNet	21M
	MCNET	0.13M
	PETCGDNN	0.075M
	CNN-atten	0.25M
	ResDBOTTA	0.18M

MCNET^[35]: The MCNET proposed in the article is an improved CNN modulation recognition model that can reduce the consumption of computational resources while ensuring high classification accuracy.

PETCGDNN^[36]: The model proposed in the article combines graph convolutional network (GCN), temporal convolutional network (TCN) and pyramid structure for multilevel feature learning. The accuracy of recognition is improved and the number of parameters is reduced.

The same models as in the previous RadioML2016.10A experiment are CNN-atten, ResNet and CNN.

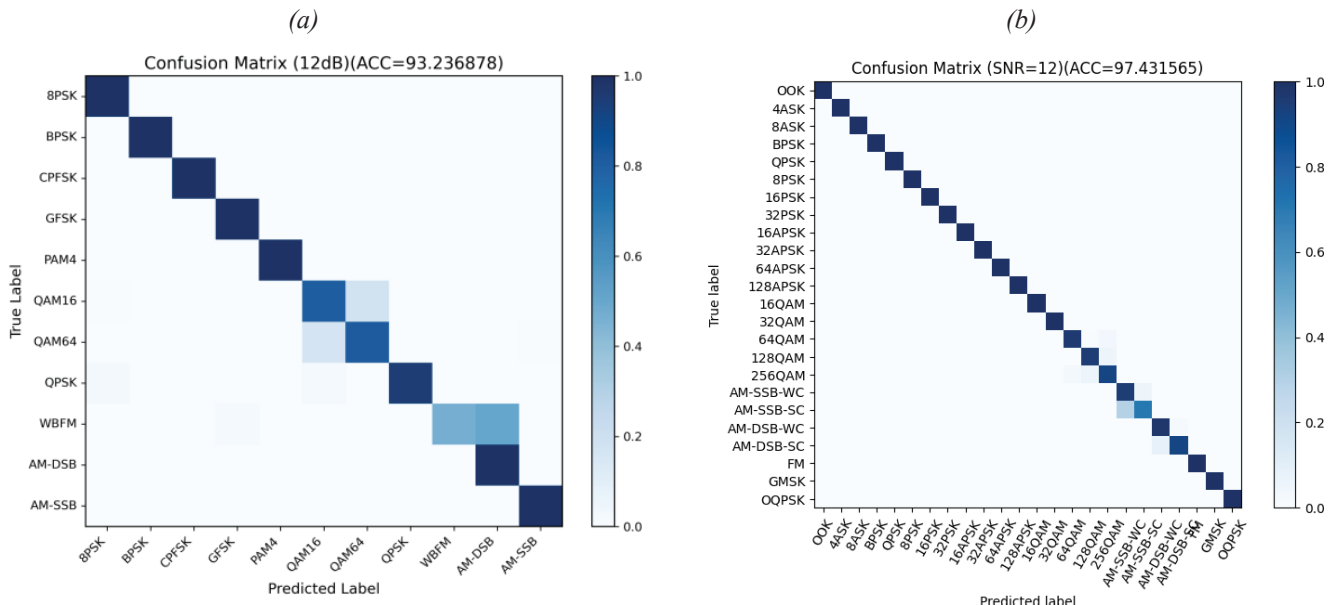
4.4 Comparative Experiments of Different Networks

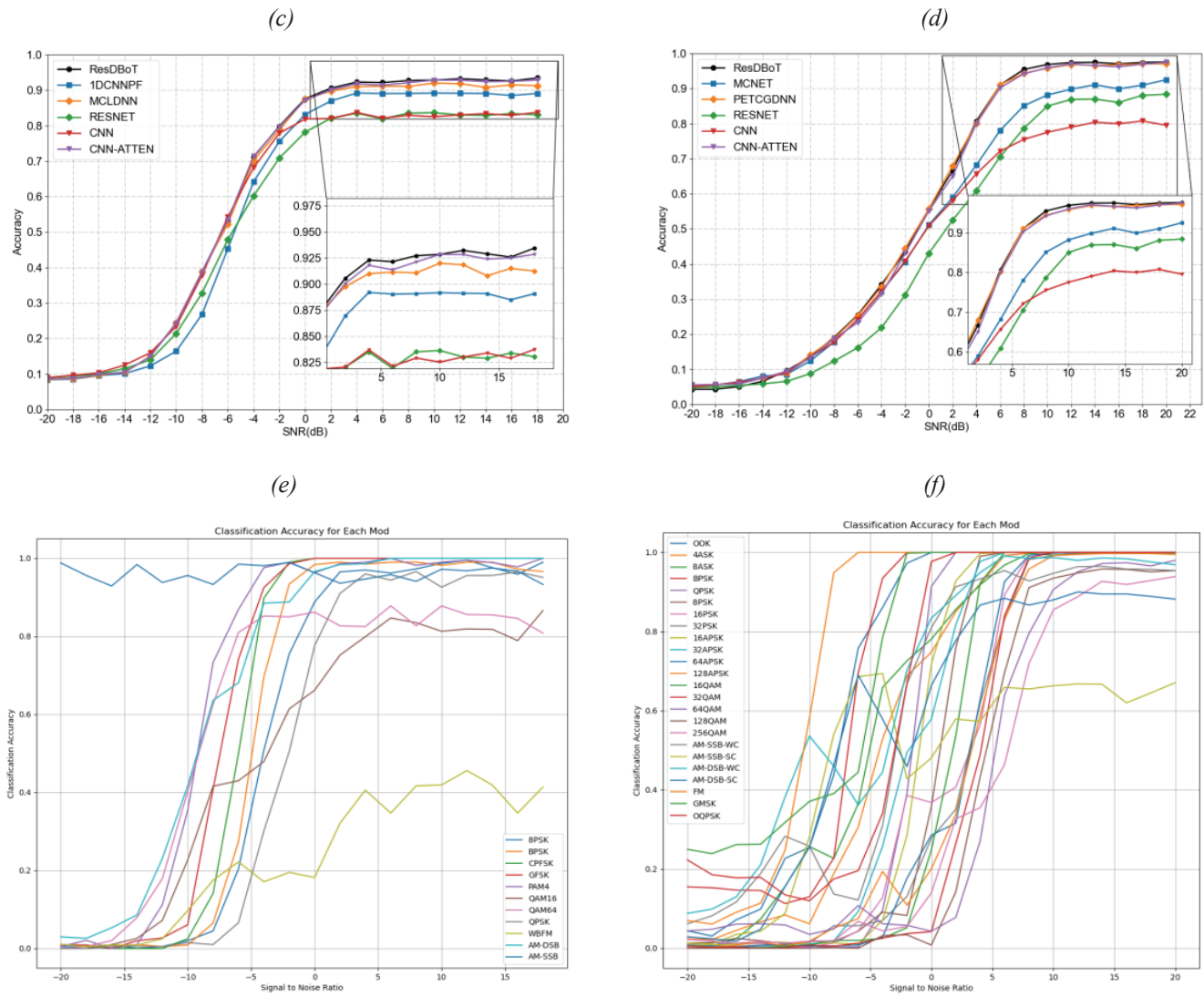
As shown in Figure 6, the confusion matrix shows the results of the proposed model in identifying the RML2016.10A (a) and RML2018.01A (b) datasets at a signal-to-noise ratio of 12dB. (c) and (d) respectively describe the recognition accuracy curves of the comparative test with the baseline model on the RML2016.10A datasets and the RML2018.01A datasets. The following sections (e) and (f) present the recognition accuracy of the ResDBoTTA for different modulation types on the RML2016.10A and RML2018.01A datasets, respectively. As shown in Figure 6(c), for the RML2016.10A dataset, when the signal-to-noise ratio is greater than 2 dB, the recognition accuracies of all the compared models reach more than 80%, among which MCLDNN, cnn-atten, and resdbot all exceed 90% accuracy.

For the RML2018.01A dataset in Figure. 6(d), at 10 dB, all except CNN achieve over 85% accuracy. As demonstrated in Figure. 6(e) and Figure. 6(f), the proposed model demonstrates commendable recognition accuracy for various modulations, except for two modulations: WBFM and AM-SSB-SC.

As shown in Figure. 6(a) and Figure. 6(b), the ResDBoTTA network achieved better recognition results in the high signal-to-noise ratio interval, with 93.23% recognition accuracy for the RML2016.10A dataset and 97.43% for the RML2018.01A dataset under the condition of a signal-to-noise ratio of 12 dB. Under the condition of higher signal-to-noise ratio, the model proposed in this paper has higher recognition accuracy compared with other models.

Figure 6. Experiments results





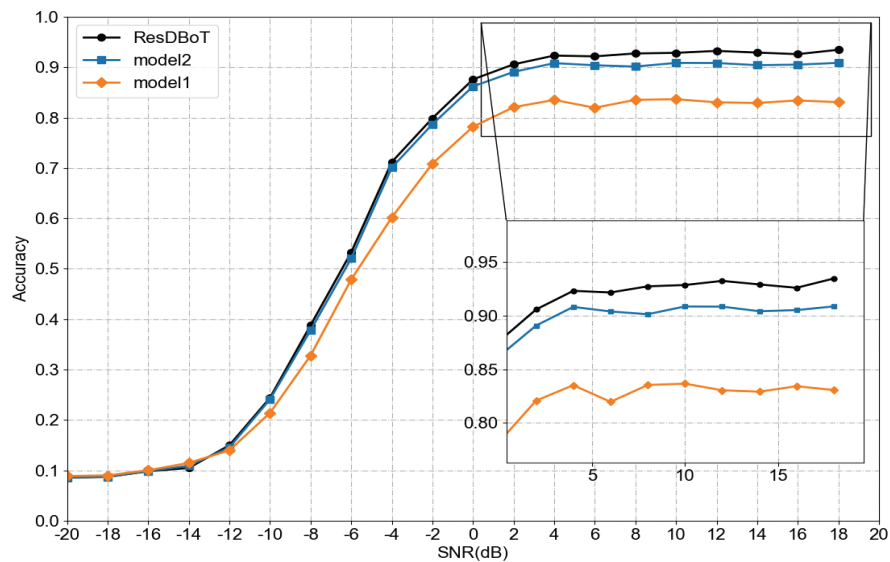
4.5 Ablation Experiment

An ablation study was conducted on the RadioML2018.01A dataset to test the performance of the module using two models:

1st model: Using the ResDBoTTA model without multi-head attention.

2nd model: Added multi-head attention to the 1st model, but no time trend-aware.

Figure 7. Recognition accuracy of the three models



Based on the results of the ablation experiments in Figure. 7, we can get the following conclusion. All models based on the DSC residual block structure have good recognition accuracy. The multi-head self-attention mechanism gives a great improvement in accuracy compared with 1st model and 2nd model. On this basis, the trend-aware self-attention mechanism can further improve the recognition accuracy in the case of high signal-to-noise ratio.

Conclusion

In this study, a new ResDBoTTA net is proposed to recognize the modulation of communication signals. The proposed ResDBoTTA model consists of three parts: depth-separable convolutional residual block to achieve multi-scale feature extraction, time trend-aware self-attention mechanism, and feature reconstruction using global depth convolution. The self-attention mechanism using time trend awareness helps to substantially improve the modulation pattern recognition accuracy of the ResDBoTTA model. In order to examine the modulation pattern recognition capability of ResDBoTTA net, a detailed experimental analysis is conducted and the results are examined from several aspects. The experimental results show that the modulation pattern recognition accuracy of ResDBoTTA net is improved compared to the existing techniques. Further research can be done in the future to reduce the model complexity while ensuring the accuracy.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Mohsen, S., Ali, A. M., & Emam, A. (2023). Automatic modulation recognition using CNN deep learning models. *Multimedia Tools and Applications*, 83(3), 7035–7056. <https://doi.org/10.1007/s11042-023-15814-y>
- [2] Das, D., Bora, P. K., & Bhattacharjee, R. (2021). Automatic modulation classification over MIMO amplify and forward (AF)-relay fading channels. *Physical Communication*, 47, 101399. <https://doi.org/10.1016/j.phycom.2021.101399>
- [3] Chen, Q., Meng, W., Han, S., Li, C., & Chen, H.-H. (2022). Robust task scheduling for delay-aware IoT applications in civil aircraft-augmented SAGIN. *IEEE Transactions on Communications*, 70(8), 5368–5385. <https://doi.org/10.1109/tcomm.2022.3186997>
- [4] Moulay, H., Djebbar, A. B., Dehri, B., & Besseghier, M. (2024). Dendrogram-based heterogeneous learners for automatic modulation classification in DSTBC-OFDM systems. *Physical Communication*, 62, 102241. <https://doi.org/10.1016/j.phycom.2023.102241>
- [5] Li, J., Chen, Q., Long, Z., Wang, W., Zhu, H., & Wang, L. (2021). Spectrum sensing with non-Gaussian noise over multi-path fading channels towards smart cities with IoT. *IEEE Access*, 9, 11194–11202. <https://doi.org/10.1109/access.2021.3051719>
- [6] Dobre, O. A., Abdi, A., Bar-Ness, Y., & Su, W. (2007). Survey of automatic modulation classification techniques: Classical approaches and new trends. *IET Communications*, 1(2), 137. <https://doi.org/10.1049/iet-com:20050176>
- [7] D., Zhang, M., Li, J., Li, Z., Li, J., Song, C., & Chen, X. (2017). Intelligent constellation diagram analyzer using convolutional neural network-based deep learning. *Optics Express*, 25(15), 17150. <https://doi.org/10.1364/oe.25.017150>
- [8] Hassan, K., Dayoub, I., Hamouda, W., & Berbineau, M. (2010). Automatic modulation recognition using wavelet transform and neural networks in wireless systems. *EURASIP Journal on Advances in Signal Processing*, 2010(1). <https://doi.org/10.1155/2010/532898>
- [9] Parmar, A., Chouhan, A., Captain, K., & Patel, J. (2024). Deep multilevel architecture for automatic modulation classification. *Physical Communication*, 64, 102361. <https://doi.org/10.1016/j.phycom.2024.102361>
- [10] O'Shea, T. J., Corgan, J., & Clancy, T. C. (2016). Convolutional radio modulation recognition networks. In *Engineering Applications of Neural Networks* (pp. 213–226). https://doi.org/10.1007/978-3-319-44188-7_16
- [11] Liu, X., Yang, D., & El Gamal, A. (2017). Deep neural network architectures for modulation classification. In *2017 51st Asilomar Conference on Signals, Systems, and Computers* (pp. 1–5). <https://doi.org/10.1109/acssc.2017.8335483>

- [12] Zhou, F., Li, J., & Wang, Y. (2023). An improved CNN-LSTM network for modulation identification relying on periodic features of signal. *IET Communications*, 17(18), 2097–2106. <https://doi.org/10.1049/cmu2.12682>
- [13] Hamza, M. A., Alghamdi, A. M., Alzahrani, J. S., Alharbi, M. T., & Al-Turki, Y. (2022). Optimal bidirectional LSTM for modulation signal classification in communication systems. *Computers, Materials & Continua*, 72(2), 3055–3071. <https://doi.org/10.32604/cmc.2022.024490>
- [14] Zhu, Z., Sun, D., Gong, K., Wang, W., & Sun, P. (2021). A lightweight CNN architecture for automatic modulation classification. *Electronics*, 10(21), 2679. <https://doi.org/10.3390/electronics10212679>
- [15] Wei, W., & Mendel, J. M. (2000). Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Transactions on Communications*, 48(2), 189–193. <https://doi.org/10.1109/26.823550>
- [16] Panagiotou, P., Anastasopoulos, A., & Polydoros, A. (2000). Likelihood ratio tests for modulation classification. In *MILCOM 2000 Proceedings. In 21st Century Military Communications. Architectures and Technologies for Information Superiority* (Vol. 2, pp. 670–674). <https://doi.org/10.1109/milcom.2000.904013>
- [17] Wu, H.-C., Saquib, M., & Yun, Z. (2008). Novel automatic modulation classification using cumulant features for communications via multipath channels. *IEEE Transactions on Wireless Communications*, 7(8), 3098–3105. <https://doi.org/10.1109/twc.2008.070015>
- [18] Ramkumar, B. (2009). Automatic modulation classification for cognitive radios using cyclic feature detection. *IEEE Circuits and Systems Magazine*, 9(2), 27–45. <https://doi.org/10.1109/mcas.2008.931739>
- [19] Pawar, S. U., & Doherty, J. F. (2011). Modulation recognition in continuous phase modulation using approximate entropy. *IEEE Transactions on Information Forensics and Security*, 6(3), 843–852. <https://doi.org/10.1109/tifs.2011.2159000>
- [20] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
- [21] Srinivas, A., Lin, T.-Y., Parmar, N., Shlens, J., Abbeel, P., & Vaswani, A. (2021). Bottleneck transformers for visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16519–16529). <https://ieeexplore.ieee.org/abstract/document/9577771>
- [22] Lin, S., Zeng, Y., & Gong, Y. (2022). Learning of time-frequency attention mechanism for automatic modulation recognition. *IEEE Wireless Communications Letters*, 11(4), 707–711. <https://doi.org/10.1109/lwc.2022.3140828>
- [23] Zhang, W., Sun, Y., Xue, K., & Yao, A. (2023). Research on modulation recognition algorithm based on channel and spatial self-attention mechanism. *IEEE Access*, 11, 68617–68631. <https://doi.org/10.1109/access.2023.3292408>
- [24] Liang, Z., Tao, M., Xie, J., Yang, X., & Wang, L. (2022). A radio signal recognition approach based on complex-valued CNN and self-attention mechanism. *IEEE Transactions on Cognitive Communications and Networking*, 8(3), 1358–1373. <https://doi.org/10.1109/tccn.2022.3179450>
- [25] Qi, M., Shi, N., Wang, G., & Shao, H. (2024). Data-transform multi-channel hybrid deep learning for automatic modulation recognition. *IEEE Access*, 12, 59113–59121. <https://doi.org/10.1109/access.2024.3393481>
- [26] Feng, Y., Duan, R., Li, S., Cheng, P., & Liu, W. (2025). A dual-branch network with feature assistance for automatic modulation recognition. *IEEE Signal Processing Letters*, 32, 701–705. <https://doi.org/10.1109/lsp.2025.3527901>
- [27] Zhang, W., Xue, K., Yao, A., & Sun, Y. (2024). CTRNet: An automatic modulation recognition based on transformer-CNN neural network. *Electronics*, 13(17), 3408. <https://doi.org/10.3390/electronics13173408>
- [28] Duan, R., Zhang, S., Wang, X., Li, Y., Liu, W., & Feng, Y. (2023). A multi-modal modulation recognition method with SNR segmentation based on time domain signals and constellation diagrams. *Electronics*, 12(14), 3175. <https://doi.org/10.3390/electronics12143175>
- [29] Luo, Z., Xiao, W., Zhang, X., Zhu, L., & Xiong, X. (2024). RLITNN: A multi-channel modulation recognition model combining multi-modal features. *IEEE Transactions on Wireless Communications*, 23, 1–15. <https://doi.org/10.1109/twc.2024.3478752>
- [30] Kong, W., Jiao, X., Xu, Y., Zhang, B., & Yang, Q. (2023). A transformer-based contrastive semi-supervised

- learning framework for automatic modulation recognition. *IEEE Transactions on Cognitive Communications and Networking*, 9(4), 950–962. <https://doi.org/10.1109/tccn.2023.3264908>
- [31] O'Shea, T. J., Roy, T., & Clancy, T. C. (2018). Over-the-air deep learning based radio signal classification. *IEEE Journal of Selected Topics in Signal Processing*, 12(1), 168–179. <https://doi.org/10.1109/jstsp.2018.2797022>
- [32] O'Shea, T. J., & West, N. (2016). Radio machine learning dataset generation with GNU Radio. *Proceedings of the GNU Radio Conference*, 1(1). <https://pubs.gnuradio.org/index.php/grcon/article/view/11>
- [33] Xu, J., Luo, C., Parr, G., & Luo, Y. (2020). A spatiotemporal multi-channel learning framework for automatic modulation recognition. *IEEE Wireless Communications Letters*, 9(10), 1629–1632. <https://doi.org/10.1109/lwc.2020.2999453>
- [34] Perenda, E., Rajendran, S., & Pollin, S. (2019). Automatic modulation classification using parallel fusion of convolutional neural networks. *IEEE Transactions on Vehicular Technology*, 69(9), 9825–9837. <https://doi.org/10.1109/tvt.2020.3000148>
- [35] Huynh-The, T., Hua, C.-H., Pham, Q.-V., & Kim, D.-S. (2020). MCNet: An efficient CNN architecture for robust automatic modulation classification. *IEEE Communications Letters*, 24(4), 811–815. <https://doi.org/10.1109/lcomm.2020.2968030>
- [36] Zhang, F., Luo, C., Xu, J., Luo, Y., & Zheng, F.-C. (2022). Deep learning based automatic modulation recognition: Models, datasets, and challenges. *Digital Signal Processing*, 129, 103650. <https://doi.org/10.1016/j.dsp.2022.103650>

Progress on CNT/Al Composites

Lian Xue^{1,2}, Kangdi Zhong^{1,2,3*}, Bingbing Wang^{1,2}, Lei Sha^{1,2}, Yuanxiao Dang^{1,2}

1. School of Aeronautics, Changji University, Changji, 831100, China

2. Xinjiang Key Laboratory of Intelligent Control and Safety Assurance Technology for Low-Altitude Airspace, Changji University, Changji, 831100, China

3. Key Laboratory of Aircraft Design, Manufacture and Maintenance, Changji University, Changji, 831100, China

*Corresponding author: Kangdi Zhong, kangdizh@163.com

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Carbon nanotube reinforced aluminum matrix composites have the advantages of light weight and high specific strength, and have broad application prospects in aerospace, automotive lightweight and other fields. In this paper, the preparation process of CNT/Al composites, the surface modification of CNT and the alloying of aluminum matrix are systematically reviewed. The mechanism and research progress of each method are analyzed. The key properties of CNT/Al composites, such as mechanics, thermal expansion, electrical conductivity and wear resistance, are summarized. The development trend and key research directions in this field are prospected.

Keywords: Carbon Nanotube; Aluminum Matrix Composites; Preparation Technology; Mechanical Properties

Published: Apr 17, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1290>

1. Introduction

As a commonly used lightweight metal material, aluminum alloy is widely used in various fields due to its low density, excellent processing performance and controllable cost. However, the strength, hardness and wear resistance of a single aluminum alloy are difficult to meet the needs of high-end scenarios. Researchers have long been committed to improving the properties of aluminum alloys by adjusting the thermal deformation parameters and optimizing the heat treatment process. Although the mechanical and physical and chemical properties of aluminum alloys can be improved to a certain extent, due to the structural characteristics of the material itself, it has failed to achieve a leap-forward breakthrough in performance. In this context, aluminum matrix composites have become an important research direction to solve this bottleneck problem due to their advantages of strong designability and excellent comprehensive performance^[1].

As a new type of nano-reinforcement, carbon nanotube (CNT) has extremely high elastic modulus, tensile strength, excellent thermal conductivity and electrical conductivity. The CNT/Al composites prepared by using CNT as reinforcement can effectively combine the lightweight characteristics of aluminum matrix with the excellent reinforcement effect of CNT, and show broad application potential in high-end fields such as aerospace parts and automobile lightweight structural parts. However, due to the strong surface inertia of CNT and the large difference in physical and chemical properties between CNT and aluminum matrix, there is a problem of weak interface bonding between them, which seriously restricts the full play of CNT's excellent performance. As a connecting bridge between the matrix and the reinforcement in the composite, the structural characteristics and formation mechanism of the interface directly determine the comprehensive properties of the

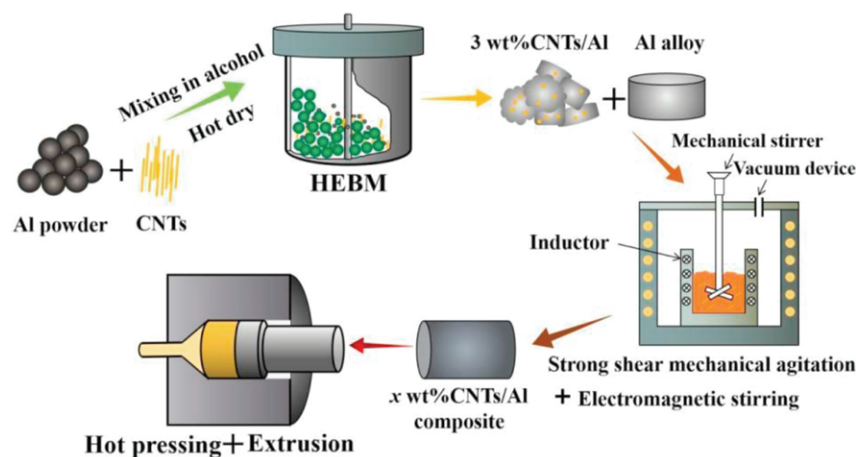
composite, such as mechanics, thermal conductivity and electrical conductivity. Therefore, interface regulation has become the core key to the performance optimization of CNT/Al composites. At present, the research on the interface control of CNT/Al composites mainly focuses on three directions: optimization of composite preparation process, CNT surface modification and aluminum matrix alloying. This paper will systematically review the above research directions and the properties of composites.

2. Preparation Process of CNT/Al Composites

2.1 Liquid Phase Method

The liquid phase method is to directly add the reinforcement particles or first make the composite preform into the metal melt, and then stir and finally cool to obtain the composite material. However, the interfacial reaction during the preparation process is not easy to control, especially the reaction between the reinforcing particles and the aluminum matrix. In the liquid phase method, the liquid metal can completely encapsulate the carbon nanotubes, and there is enough surface contact to form a strong interfacial bonding force. However, at higher temperatures, serious interfacial reactions may occur, and the formation of brittle interfacial phases affects the properties of the composites. Zhang et al.^[2] used high-energy ball milling (HEBM) -semi-solid stirring casting (SSC) process to prepare CNT reinforced Al-Cu-Mg composites, and explored the effect of CNT on the microstructure and properties after hot extrusion, as shown in Fig.1. With the multiple strengthening effect induced by CNT, when the CNT content increased from 0 to 1.5 wt. %, the tensile strength and yield strength of the composites increased from 319 MPa and 261 MPa to 523 MPa and 476 MPa, respectively.

Figure 1: The schematic diagram of the preparation process of CNT/Al copper magnesium composite material^[2].



2.2 Solid-state Method

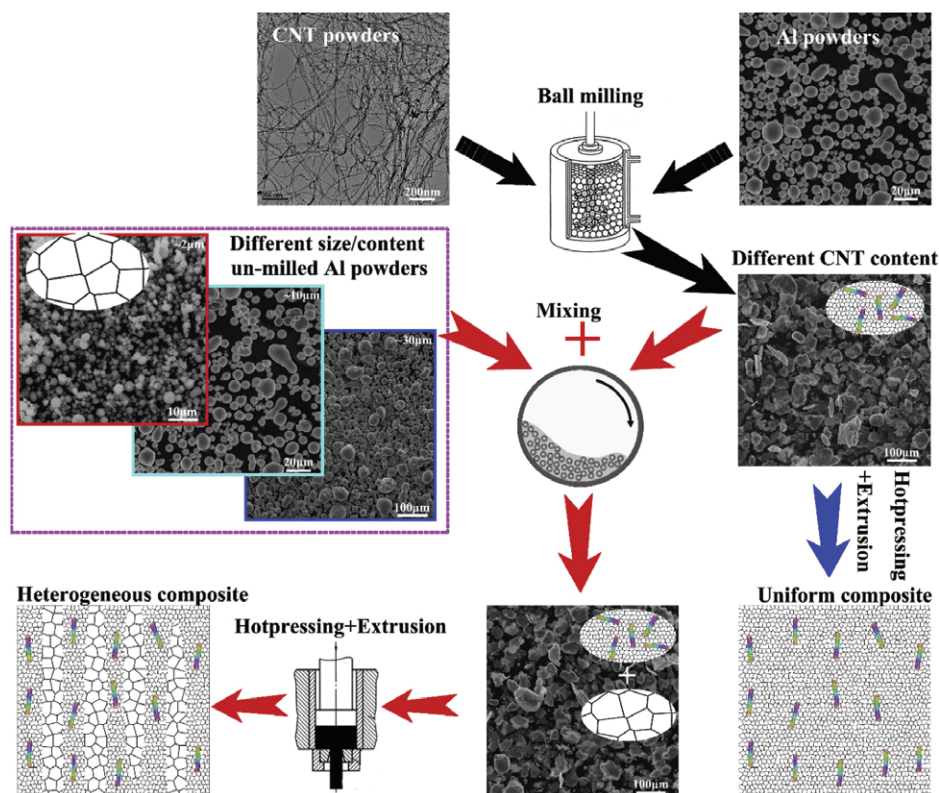
The solid phase method mainly includes powder metallurgy, hot extrusion, cold pressing sintering and spark plasma sintering (SPS). In order to improve the density, carbon nanotube dispersion and mechanical properties of materials, it is often supplemented by single or composite secondary processing processes such as extrusion, rolling, friction stir welding and forging.

The powder metallurgy method achieves uniform dispersion of the reinforcing phase by mixing the matrix powder with the reinforcing phase particles. After mixing, the green body is obtained by cold pressing, and the density can reach 80% of the theoretical density. Subsequently, the green body was sealed and degassed to remove the water adsorbed on the surface of the powder; finally, a high-density composite material was prepared by isostatic pressing or uniaxial sintering^[3]. Choi et al.^[4] found that when preparing aluminum matrix composites by powder metallurgy method, with the increase of ball milling strength, carbon nanotubes and aluminum matrix formed a good interface bonding through mechanical occlusion. Singhal et al.^[5] prepared aminated carbon nanotubes reinforced aluminum matrix composites by powder metallurgy method. The results show that the dispersion of aminated carbon nanotubes in aluminum matrix is significantly better than that of unmodified carbon nanotubes, which can be uniformly dispersed in aluminum matrix without obvious agglomeration.

The hot extrusion method refers to the processing technology of extruding the material powder from the mold hole by high

pressure under the heating state. Wu et al.^[6] used semi-solid powder metallurgy process to achieve partial liquid phase sintering by raising the temperature near the melting point of aluminum powder during hot pressing. It was found that the increase of liquid phase composition could significantly improve the density and mechanical properties of the material. The hardness of CNT/6061Al alloy composite reached 87.5 HV after sintering at 620 °C, which was nearly twice as high as that of the matrix alloy. Esawi et al.^[7] mixed 5wt. % CNT with aluminum powder by ball milling, cold pressing and hot extrusion sintering. The results showed that the stiffness of the composite was 23% higher than that of pure aluminum, and the tensile strength could be increased by 50%. When the CNT content exceeded 2 wt. %, the dispersion difficulty increased, and the mechanical properties did not meet the expectations. Liu et al.^[8] used a bimodal structure design (Fig. 2), and added unmilled aluminum powder to the CNT/Al composite powder, so that the elongation of the composite was 100% higher than that of the uniform CNT/Al composite, and the tensile strength remained basically unchanged. Bakshi et al.^[9] found that in the densification process of composite powder, when the extrusion ratio is greater than 20 and the pressing force reaches the order of GPa, the composite material has better performance ; the greater the extrusion ratio and pressing force, the higher the deformation degree of the material, and the hot extrusion rate and temperature need to be reasonably controlled, otherwise it is easy to cause processing instability and surface cracking.

Figure 2: A schematic diagram of the preparation process of CNT/2009Al composites with heterogeneous and uniform structure^[8].



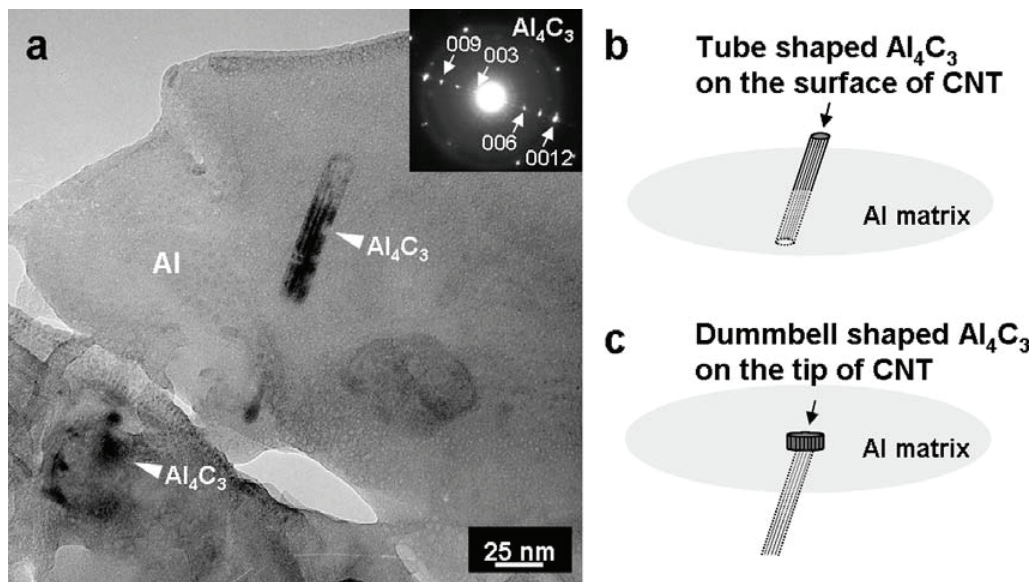
George et al.^[10] prepared CNT/Al powder mixture by ball milling, and sintered in nitrogen atmosphere after 120 kN pressure cold pressing. The results showed that the mechanical properties such as Young's modulus of the composite were significantly improved. Xu et al.^[11] found that the higher the input energy of ball milling and the longer the time, the more serious the damage of CNT, and the defects will promote the formation of Al_4C_3 in the subsequent sintering or heat treatment process. The interfacial reaction rate of high-speed ball milling composites is 61.85%, which is much higher than 2.42% of low-speed ball milling. Although high-speed ball milling can make CNT embedded in Al matrix to achieve good bonding, it will seriously damage the crystal structure of CNT. Chen et al.^[12] showed that amorphous natural Al_2O_3 could be transformed into γ - Al_2O_3 nanoparticles under the suitable SPS conditions of 540~635°C, 30~300 MPa, and 60 min, thereby improving the strain hardening ability of the composites. Chen et al.^[13] used the cumulative extrusion bonding (AEB) process to process ultrafine-grained aluminum. The microstructure characterization showed that a good interface bonding was obtained after 6

passes of processing, and the average grain size was ≤ 440 nm. Compared with the traditional accumulative rolling bonding process, this process can refine the grains more efficiently and significantly improve the mechanical properties of the samples.

2.3 CNT In-Situ Generation Method

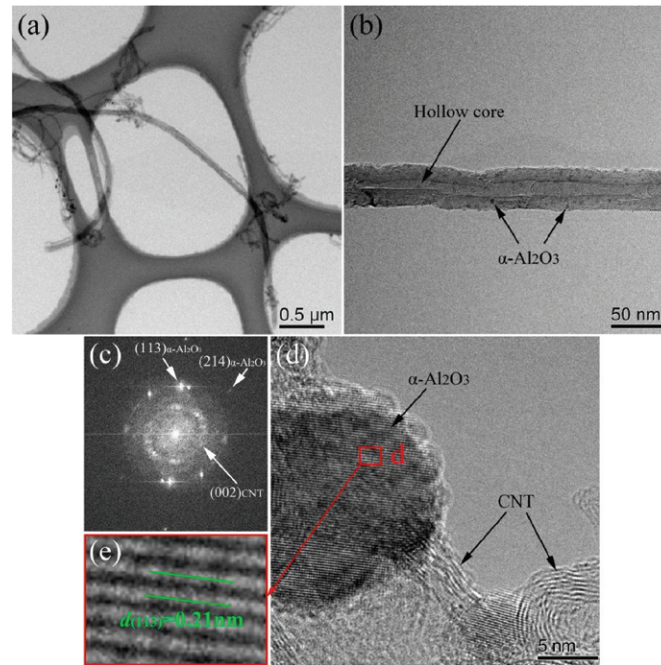
CNT in-situ formation method refers to the process of directly in-situ growth of CNT on the surface of metal particles. However, the carbon source used in this method to generate CNT is combustible gas, which not only has certain safety hazards, but also has low yield, thus limiting its practical application. Kwon et al.^[14] prepared Al/CNTs composites by SPS combined with hot extrusion process. It was found that the addition of CNT can improve the tensile strength without reducing the elongation. The reason is that CNT exists in the grain boundary layer, and the formation of Al_4C_3 and the good orientation of CNT along the extrusion direction achieve effective stress transfer (Figure 3). Yang et al.^[15] synthesized CNT in situ after depositing 5% (mass fraction) nickel on the surface of aluminum powder by impregnation method. The results show that the high crystallinity CNT synthesized in situ has a bamboo structure, which can avoid the interface reaction with aluminum matrix below 1,000 °C. The yield strength and elastic modulus of the composite with CNT content of 1.5% prepared by hot pressing are 2.2 times and 3 times that of the pure aluminum matrix, respectively.

Figure 3: High resolution transmission electron microscopy images of the microstructure of CNT/Al composites after extrusion^[14]. (a) Al_4C_3 morphology, and the crystal structure of Al_4C_3 was confirmed by illustrations. The surface of (b) and (c) defective carbon nanotubes (CNT) forms tubular Al_4C_3 , while the CNT tip forms granular Al_4C_3



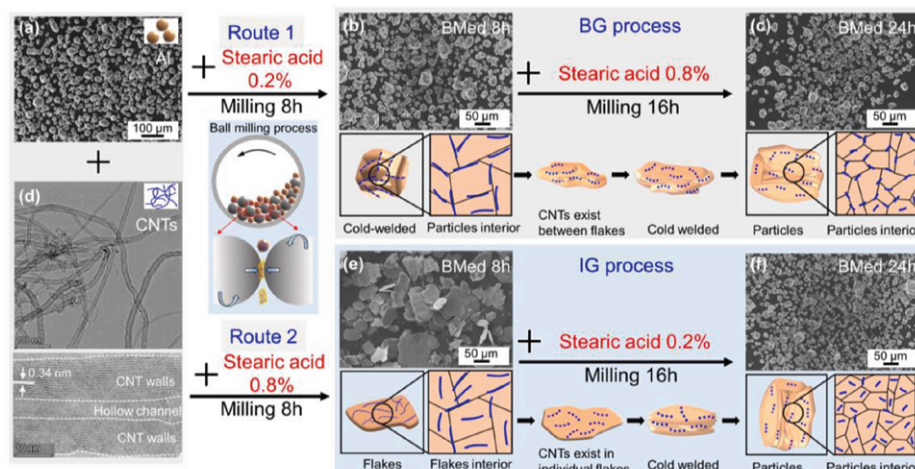
3. Surface Modification of CNT

The core of CNT surface modification is to improve the surface activity by regulating its surface state and microstructure, thereby improving its compatibility with the aluminum matrix and inhibiting the agglomeration of CNT, mainly including surface oxidation modification and surface coating treatment. In order to realize the surface functionalization of the original multi-walled CNT without destroying its structural integrity, Zhou et al.^[16] modified the CNT by mild acid treatment. The appropriate amount of acid-treated CNT and Al powder were dispersed in ethanol respectively, and the suspension was mixed after 1 h ultrasonic treatment. The results showed that the modified CNT could be uniformly dispersed in the aluminum matrix. So et al.^[17] coated silicon nanoparticles on the surface of CNT and formed a SiC layer after high temperature annealing. The SiC layer not only reduced the wetting angle between CNT and aluminum matrix and improved the interfacial wettability, but also significantly enhanced the interfacial bonding strength by forming SiC covalent bonds. Chen et al.^[18] prepared 2195Al/CNT composites by Al_2O_3 partially coated modified CNT, combined with melting casting and hot extrusion. After aging treatment, the tensile strength, yield strength and elongation reach 697 MPa, 590 MPa and 11.3%, respectively, and the mechanical properties are significantly improved compared with those of 2195 alloy under the same conditions. CNT is mainly distributed in the grain boundary, and there is a small amount of fluorescent rod-like Al_4C_3 phase near it, which is formed by the segregation of the solute element interface of the alloy, as shown in Figure 4.

Figure 4: TEM images^[18]: (a-b) TEM images, (c) SAED patterns, (d-e) HRTEM images.

4. Aluminum Matrix Alloying

Matrix alloying refers to the addition of a small amount of alloying elements (Si, Cr, etc.) to the aluminum matrix. Through the interfacial chemical reaction between the alloying elements and CNT, a certain thickness of the interfacial reaction layer is formed, thereby improving the interfacial bonding strength and inhibiting the reaction between CNT and the aluminum matrix to form Al_4C_3 phase which is easy to hydrolyze. Yuan et al.^[19] introduced magnesium into aluminum powder when preparing CNT reinforced aluminum matrix composites. It was found that micro-nano strengthening phases such as Al_4C_3 , MgAl_2O_4 and Al_2O_3 were formed at the bottom of the pits of the composites. At the same time, the Al_2O_3 layer is replaced by spinel phase (MgAl_2O_4), which is distributed both around the interface and inside the layered grains. Such fine nano-grains can effectively alleviate the stress concentration, eliminate the early crack source during the tensile process, and significantly improve the extensibility of the material. Cao et al.^[20] reported a strategy to improve the ductility and tensile strength of CNTs/Al composites by embedding CNT into the matrix grains. The preparation process is shown in Figure 5. Two kinds of composites were prepared by modified ball milling powder metallurgy process, which were CNT mainly dispersed in the grain interior and mainly distributed in the grain boundary. The results show that compared with the composites with CNT distributed at grain boundaries and the reference composites, the composites with CNT dispersed inside the grains have higher tensile strength and ductility at different CNT contents.

Figure 5: Preparation schematic diagram of CNTs/Al composite powder with different carbon nanotube distribution^[20]

5. Properties of CNT Reinforced Aluminum Matrix Composites

CNT reinforced aluminum matrix composites have excellent comprehensive mechanical properties. The uniform dispersion of CNT in aluminum matrix plays an increasingly important role in improving the comprehensive properties of composites. At the same time, the introduction of CNT can significantly reduce the thermal expansion coefficient of the composite material. On the one hand, the thermal expansion coefficient of CNT itself is almost zero, on the other hand, the CNT at the interface has a restraining effect on the aluminum matrix, thus inhibiting the thermal expansion of the composite material^[21-23]. Since the conductivity of CNT is lower than that of aluminum matrix, its addition will lead to a decrease in the conductivity of the composites, and the conductivity decreases with the increase of CNT content. The study of wear performance shows that the friction coefficient and wear loss of the composites decrease with the increase of CNT content due to the self-lubricating effect of CNT. At the same time, the introduction of CNT improves the deformation resistance of the aluminum matrix, and the synergistic effect of the two makes the composite obtain excellent wear resistance^[24-27].

Conclusion

In this paper, the interface control and properties of CNT reinforced aluminum matrix composites are systematically reviewed. The three paths of preparation process, surface modification and matrix alloying work together to effectively improve the CNT dispersion and interface bonding strength, and achieve the improvement of material properties. At present, there are still some problems such as uneven dispersion of CNT and insufficient controllability of interface reaction. In the future, it is necessary to focus on process optimization, interface reaction regulation and performance balance to promote its development to engineering scale application.

Funding

Supported by Natural Science Foundation of Xinjiang Uygur Autonomous Region (2024D01C210); Science and Technology Plan Project of Changji Prefecture (2025S01-03); Research Project of Changji University (KY2024024).

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Liu, J., Khan, U., Coleman, J. et al. (2016). Graphene oxide and graphene nanosheet reinforced aluminium matrix composites: Powder synthesis and prepared composite characteristics. *Materials & Design*, 94: 87-94. 10.1016/j.matdes.2016.01.031.
- [2] Zhang, Z., Xiao, Y., Xu, J. et al. (2022). Understanding the influencing mechanism of CNTs on the microstructure and mechanical properties of semi-solid stir casting Al-Cu-Mg alloys. *Journal of Materials Research and Technology*, 18: 3949-3960. <https://doi.org/10.1016/j.jmrt.2022.04.087>.
- [3] Fang, Z. Z., Paramore, J. D., Sun, P. et al. (2017). Powder metallurgy of titanium – past, present, and future. *International Materials Reviews*, 63(7): 407-459. 10.1080/09506608.2017.1366003.
- [4] Choi, H. J., Shin, J. H., Bae, D. H. (2012). The effect of milling conditions on microstructures and mechanical properties of Al/MWCNT composites. *Composites Part a: Applied Science and Manufacturing*, 43(7): 1061-1072. 10.1016/j.compositesa.2012.02.008.
- [5] Singhal, S. K., Pasricha, R., Jangra, M. et al. (2012). Carbon nanotubes: Amino functionalization and its application in the fabrication of Al-matrix composites. *Powder Technology*, 215-216: 254-263. 10.1016/j.powtec.2011.10.013.
- [6] Wu, Y., Kim, G. (2011). Carbon nanotube reinforced aluminum composite fabricated by semi-solid powder processing. *Journal of Materials Processing Technology*, 211(8): 1341-1347. 10.1016/j.jmatprotec.2011.03.007.
- [7] Esawi, A. M. K., Morsi, K., Sayed, A. et al. (2010). Effect of carbon nanotube (CNT) content on the mechanical properties of CNT-reinforced aluminium composites. *Composites Science and Technology*, 70(16): 2237-2241. 10.1016/j.compscitech.2010.05.004.
- [8] Liu, Z. Y., Ma, K., Fan, G. H. et al. (2020). Enhancement of the strength-ductility relationship for carbon nanotube/Al–

- Cu–Mg nanocomposites by material parameter optimisation. *Carbon*, 157: 602-613. 10.1016/j.carbon.2019.10.080.
- [9] Bakshi, S. R., Agarwal, A. (2011). An analysis of the factors affecting strengthening in carbon nanotube reinforced aluminum composites. *Carbon*, 49(2): 533-544. 10.1016/j.carbon.2010.09.054.
- [10] George, R., Kashyap, K. T., Rahul, R. et al. (2005). Strengthening in carbon nanotube/aluminium (CNT/Al) composites. *Scripta Materialia*, 53(10): 1159-1163. 10.1016/j.scriptamat.2005.07.022.
- [11] Xu, R., Tan, Z., Xiong, D. et al. (2017). Balanced strength and ductility in CNT/Al composites achieved by flake powder metallurgy via shift-speed ball milling. *Composites Part a: Applied Science and Manufacturing*, 96: 57-66. 10.1016/j.compositesa.2017.02.017.
- [12] Chen, B., Zhou, X. Y., Zhang, B. et al. (2020). Microstructure, tensile properties and deformation behaviors of aluminium metal matrix composites co-reinforced by ex-situ carbon nanotubes and in-situ alumina nanoparticles. *Materials Science and Engineering: A*, 795: 139930. 10.1016/j.msea.2020.139930.
- [13] Chen, X., Huang, G., Liu, S. et al. (2019). Grain refinement and mechanical properties of pure aluminum processed by accumulative extrusion bonding. *Transactions of Nonferrous Metals Society of China*, 29(3): 437-447. 10.1016/S1003-6326(19)64953-8.
- [14] Kwon, H., Park, D. H., Silvain, J. F. et al. (2010). Investigation of carbon nanotube reinforced aluminum matrix composite materials. *Composites Science and Technology*, 70(3): 546-550. 10.1016/j.compscitech.2009.11.025.
- [15] Yang, X., Shi, C., He, C. et al. (2011). Synthesis of uniformly dispersed carbon nanotube reinforcement in Al powder for preparing reinforced Al composites. *Composites Part a: Applied Science and Manufacturing*, 42(11): 1833-1839. 10.1016/j.compositesa.2011.08.009.
- [16] Zhou, W., Sasaki, S., Kawasaki, A. (2014). Effective control of nanodefects in multiwalled carbon nanotubes by acid treatment. *Carbon*, 78: 121-129. 10.1016/j.carbon.2014.06.055.
- [17] So, K. P., Jeong, J. C., Park, J. G. et al. (2013). SiC formation on carbon nanotube surface for improving wettability with aluminum. *Composites Science and Technology*, 74: 6-13. 10.1016/j.compscitech.2012.09.014.
- [18] Chen, X., Liu, K., Peng, S. et al. (2025). Enhanced mechanical properties of the surface-modified CNTs reinforced 2195 aluminum-based composite. *Materials Science and Engineering: A*, 922: 147623. 10.1016/j.msea.2024.147623.
- [19] Yuan, C., Zhang, Z., Tan, Z. et al. (2021). Enhanced ductility by Mg addition in the CNT/Al-Cu composites via flake powder metallurgy. *Materials Today Communications*, 26: 101854. 10.1016/j.mtcomm.2020.101854.
- [20] Cao, L., Chen, B., Wan, J. et al. (2025). Simultaneously improving strength and ductility of carbon nanotube (CNT)-reinforced aluminum matrix composites by embedding CNTs inside matrix grains. *Composites Part B: Engineering*, 296: 112240. 10.1016/j.compositesb.2025.112240.
- [21] Maniwa, Y., Fujiwara, R., Kira, H. et al. (2001). Thermal expansion of single-walled carbon nanotube (SWNT) bundles: X-ray diffraction studies. *Physical Review B*, 64(24). 10.1103/physrevb.64.241402.
- [22] Deng, C. F., Ma, Y. X., Zhang, P. et al. (2008). Thermal expansion behaviors of aluminum composite reinforced with carbon nanotubes. *Materials Letters*, 62(15): 2301-2303. 10.1016/j.matlet.2007.11.086.
- [23] Zhong, K., Zhou, J., Zhao, C. et al. (2023). Inhibiting of interface thermal cracking with the interlayer nanostructure in fiber aluminum hierarchical composite. *Composites Communications*, 40: 101599. 10.1016/j.coco.2023.101599.
- [24] Choi, H. J., Lee, S. M., Bae, D. H. (2010). Wear characteristic of aluminum-based composites containing multi-walled carbon nanotubes. *Wear*, 270(1-2): 12-18. 10.1016/j.wear.2010.08.024.
- [25] Kim, I., Lee, J., Lee, G. et al. (2009). Friction and wear characteristics of the carbon nanotube–aluminum composites with different manufacturing conditions. *Wear*, 267(1-4): 593-598. 10.1016/j.wear.2008.12.096.
- [26] Liu, Z. Y., Xiao, B. L., Wang, W. G. et al. (2013). Developing high-performance aluminum matrix composites with directionally aligned carbon nanotubes by combining friction stir processing and subsequent rolling. *Carbon*, 62: 35-42. 10.1016/j.carbon.2013.05.049.
- [27] Zuo, G., Bai, Y., Tan, Z. et al. (2024). Temperature effects on solid state bonding joints of ultrafine-grained CNT/Al–Cu–Mg composites. *Composites Part B: Engineering*, 284: 111681. 10.1016/j.compositesb.2024.111681.

GASP-D3QN: Geometry-Aware Safety-Prioritized Dueling Double Q-Network for Online UAV Path Planning

Kexiao Wu¹, Bingyu Yang², Mingshen Xu^{1*}

1. North China Electric Power University, Baoding, Hebei, 071003, China

2. Guizhou University, Guiyang, Guizhou, 550025, China

*Corresponding author: Mingshen Xu, 3165865378@qq.com

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Online three-dimensional path planning for unmanned aerial vehicles (UAVs) in cluttered environments remains challenging because decision-making must be performed under partial geometric observation, dynamic uncertainty, and strict safety constraints. This paper presents GASP-D3QN, a value-based reinforcement learning framework that integrates a geometry-aware hybrid encoder, a hard safety shield, a prior-guided action selector, and a dueling double-Q backbone with prioritized replay. The proposed design explicitly separates global kinematic cues from a local occupancy cube and introduces task priors related to goal progress, clearance, energy cost, and heading stability. To ensure an application-consistent evaluation, the comparison includes both learning-based online baselines and a classical reference method under the same observation and action interface. On the standard benchmark, GASP-D3QN achieves the best overall performance among the learning-based methods, with a success rate of 0.50 and an average return of 383.19, while maintaining competitive energy consumption. Additional experiments under denser obstacles and out-of-distribution wind disturbances preserve the same advantage over the learning-based baselines. Ablation studies further show that the geometry-aware encoder, safety shield, and action prior each contribute materially to the final result. These findings indicate that explicit geometric modeling and safety-aware action filtering provide an effective and practical recipe for learned online UAV navigation.

Keywords: UAV Path Planning; Deep Reinforcement Learning; Dueling Double Q-Network; Safety Shield; Geometry-Aware Representation; Online Decision Making

Published: Apr 17, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1272>

1. Introduction

Autonomous path planning for unmanned aerial vehicles has long been a central problem in robotics and intelligent control. Classical approaches were established on heuristic search, artificial potential fields, and sampling-based motion planning, and many modern planning systems continue to inherit these design principles^[1-6]. Although such methods remain highly effective in structured environments, their performance may degrade when the operating space changes online, when sensing is restricted to local observations, or when the vehicle must react rapidly to uncertainty.

Recent work on UAV planning and autonomous navigation highlights this difficulty from both algorithmic and application perspectives. Survey studies report that path-planning performance is strongly affected by environmental uncertainty, onboard sensing limitations, and the tension between safety, efficiency, and computational cost^[7-8]. In parallel, reinforcement learning (RL) has emerged as a promising framework for sequential robotic decision-making when explicit environment modeling is

incomplete or costly, and when the controller must learn a direct mapping from observations to actions.

Within the RL literature, value-based methods are especially relevant to discrete motion-planning tasks. Deep Q-Networks demonstrated that high-dimensional observations can be converted into action-value estimates^[9]. Subsequent advances improved this framework substantially: Double Q-learning reduced value overestimation^[10], dueling architectures enhanced value-advantage decomposition^[11], and prioritized replay improved sample utilization^[12]. However, these generic improvements do not by themselves resolve the core difficulties of geometric navigation, namely representation quality, safety assurance, and action efficiency.

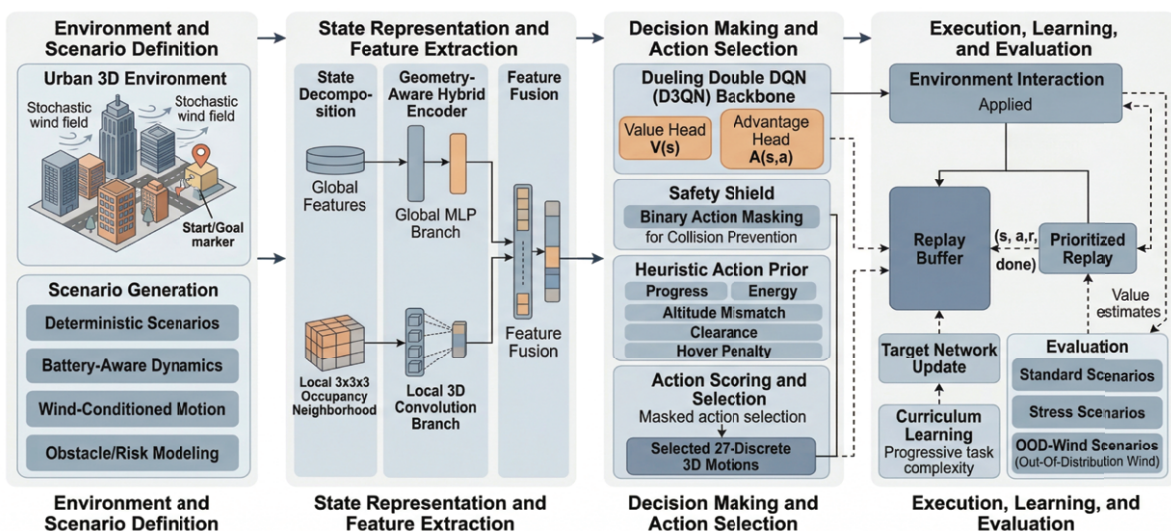
A parallel research direction has explored learned UAV navigation, hybrid planning-learning systems, and safety-aware decision-making. Air Learning emphasized that evaluating autonomous aerial navigation requires not only success statistics but also flight-quality indicators such as energy expenditure and control efficiency^[13]. Related studies have investigated UAV path planning with reinforcement learning in field prospecting and complex dynamic environments^[14-16]. In the broader safe RL literature, explicit constraints and safety-aware optimization have also been shown to improve reliability during learning and deployment^[17-22].

Despite these advances, three limitations remain prominent in the target setting considered in this study. First, many learning-based planners still collapse heterogeneous observations into a single flattened representation, thereby weakening the model's ability to reason about local geometry. Second, safety is often handled exclusively through reward penalties, even though a subset of actions can be identified as clearly unsafe before execution. Third, domain knowledge such as goal progress, freespace preference, and energy awareness is frequently embedded only in reward shaping rather than being used to guide online action selection directly.

To address these issues, this paper proposes GASP-D3QN (Geometry-Aware Safety-Prioritized Dueling Double Q-Network), a reinforcement learning framework for online 3-D UAV path planning. The proposed method combines a hybrid state encoder for global and local geometry, a hard action mask that filters immediately unsafe actions, and a prior-guided action selector built on top of a dueling double-Q learner with prioritized replay. The main contributions of this work are summarized as follows:

1. A geometry-aware hybrid state encoder is introduced to separately process global kinematic information and a local $3 \times 3 \times 3$ occupancy cube before feature fusion and value estimation.
2. A safety-prioritized decision mechanism is designed to eliminate evidently unsafe actions and rank the remaining safe candidates using goal-progress, clearance, energy, and stability priors.
3. A scenario-consistent comparison protocol is established for online decision baselines under a shared observation and action interface, and the proposed method achieves the best overall performance among the learning-based methods.
4. Robustness and ablation experiments verify that geometric perception, explicit safety filtering, and action priors each contribute materially to navigation quality, efficiency, and collision reduction.

Figure 1: Overall architecture of GASP-D3QN.



2. Method

2.1 Problem formulation

The online UAV path-planning task is modeled as a finite-horizon Markov decision process. At time step t , the UAV observes state S_t , executes discrete action a_t , receives reward r_t , and transitions to S_{t+1} . The objective is to maximize the discounted cumulative return:

$$J = \mathbb{E} \left[\sum_{t=0}^{T-1} \gamma^t r_t \right]. \quad (1)$$

The state is decomposed into a global feature vector x_t^g and a local geometry tensor x_t^l :

$$s_t = (x_t^g, x_t^l), x_t^l \in \{0, 1\}^{3 \times 3 \times 3}. \quad (2)$$

The global branch contains relative goal direction, vehicle velocity, altitude, and related kinematic descriptors, whereas the local branch represents the occupancy configuration surrounding the UAV. The action space consists of a finite library of 3-D motion primitives. An episode terminates when the vehicle reaches the goal, collides with an obstacle, or exceeds the maximum allowed number of steps.

2.2 Geometry-aware hybrid encoder

A central design principle of GASP-D3QN is to avoid forcing geometrically distinct observations into a single undifferentiated vector representation. Instead, two dedicated encoders are employed:

$$h_t^g = f_g(x_t^g), h_t^l = f_l(x_t^l), \quad (3)$$

and their outputs are fused as

$$h_t = [h_t^g; h_t^l]. \quad (4)$$

The global branch $f_g(\cdot)$ is implemented as a lightweight multi-layer perceptron suitable for low-dimensional continuous kinematic inputs. The local branch $f_l(\cdot)$ is a dedicated geometry encoder that preserves the structural relationships embedded in the local occupancy cube. This separation is advantageous because metric goal-oriented cues and local binary occupancy patterns differ significantly in scale, sparsity, and statistical structure.

The fused representation h_t is passed to a dueling action-value head:

$$Q_\theta(s_t, a) = V_\theta(h_t) + A_\theta(h_t, a) - \frac{1}{|\mathcal{A}|} \sum_{a'} A_\theta(h_t, a'). \quad (5)$$

This formulation allows the model to estimate state value and action-specific deviations separately, which is particularly beneficial when many actions have similar consequences in a given local state.

2.3 Safety shield and action prior

Relying solely on reward penalties to discourage collisions is often inefficient because clearly unsafe actions may still be sampled repeatedly during learning. To address this issue, a hard action mask is introduced:

$$m_t(a) = \begin{cases} 0, & \text{if } a \text{ intersects occupied cells} \\ & \text{or violates altitude bounds} \\ 1, & \text{otherwise} \end{cases} \quad (6)$$

Only actions satisfying $m_t(a) = 1$ are admitted to the final decision stage.

Among the remaining safe candidates, a prior score is computed:

$$p_t(a) = \lambda_d \Delta d_t(a) + \lambda_c c_t(a) - \lambda_e e_t(a) + \lambda_s \sigma_t(a), \quad (7)$$

where $\Delta d_t(a)$ denotes expected goal-progress gain, $c_t(a)$ denotes predicted local clearance, $e_t(a)$ is an energy-related proxy, and $\sigma_t(a)$ rewards dynamically stable heading changes. The final selected action is

$$a_t = \arg \max_{a \in \mathcal{A}, m_t(a)=1} [Q_\theta(s_t, a) + \beta_t p_t(a)]. \quad (8)$$

The coefficient β_t is annealed during training. This schedule allows the policy to benefit from structured domain guidance during early exploration while gradually shifting toward value-dominant decision-making as learning progresses.

2.4 Double-Q target and prioritized replay

Parameter optimization follows the Double-DQN target formulation:

$$y_t = r_t + \gamma Q_{\bar{\theta}} \left(s_{t+1}, \arg \max_{a'} Q_\theta(s_{t+1}, a') \right). \quad (9)$$

Prioritized experience replay samples transitions according to temporal-difference error magnitude, and the training loss is

defined as

$$\mathcal{L}(\theta) = \mathbb{E}_i[w_i(y_i - Q_\theta(s_i, a_i))^2], \quad (10)$$

where w_i denotes the standard importance-sampling correction weight. The target network $\bar{\theta}$ is updated periodically to stabilize training.

Figure 2: Moving-average training return.

P.S: The proposed model converges faster and reaches a substantially higher performance plateau than the compared baselines.

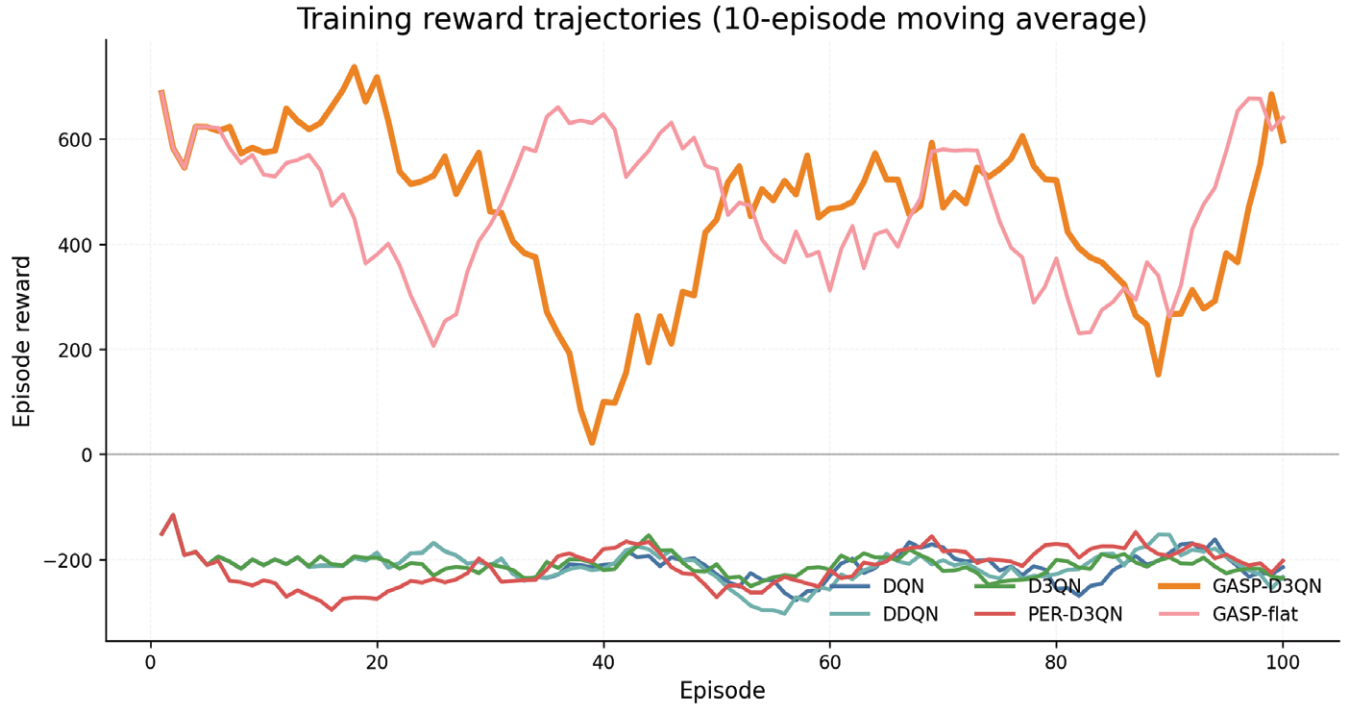
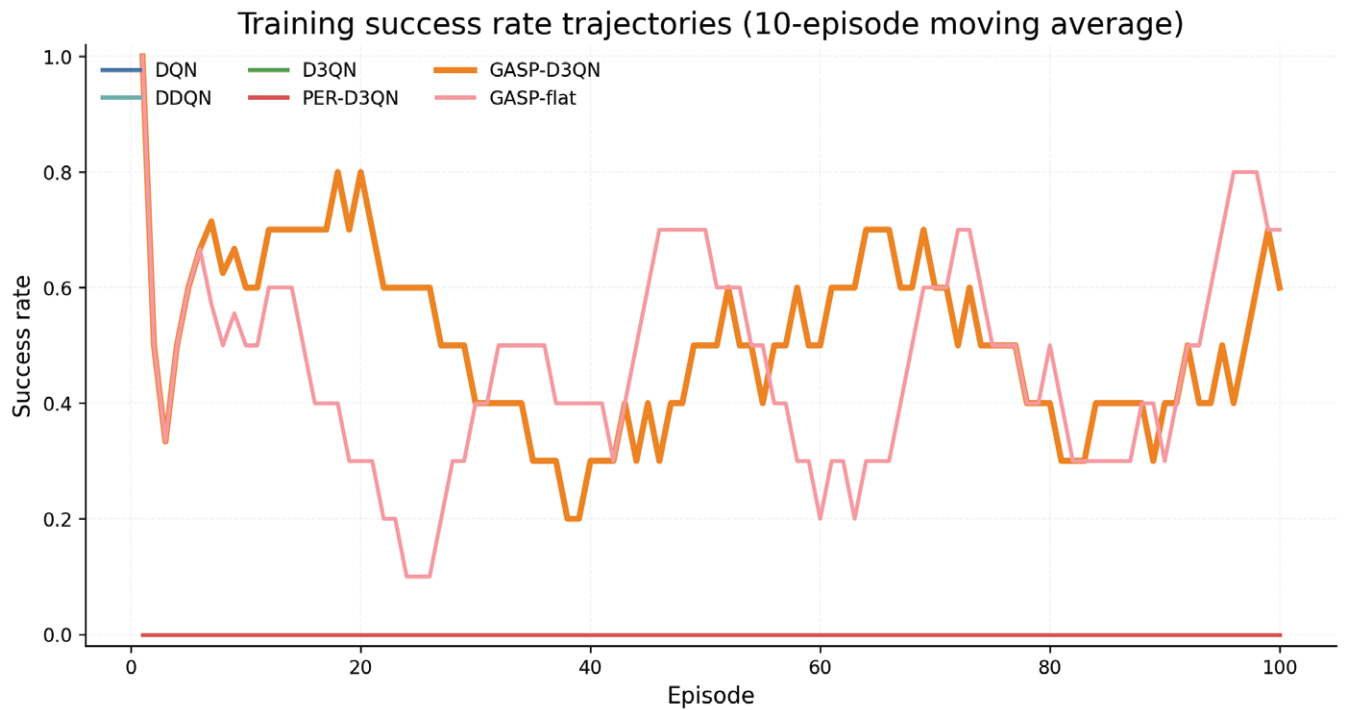


Figure 3: Moving-average training success rate.

P.S: The performance margin between GASP-D3QN and the baseline methods widens notably after the middle training stage.



3. Experimental setup

3.1 Benchmark protocol and comparison set

The benchmark consists of randomized 3-D obstacle fields with start-goal pairs sampled from feasible free space. Training adopts a progressive-difficulty curriculum, illustrated in Fig. 4, in which obstacle density, moving-obstacle ratio, and wind disturbance are gradually increased. Evaluation is performed on three mutually exclusive test sets: a standard set, a stress set, and an out-of-distribution (OOD) wind set.

3.2 Metrics and implementation details

Performance is assessed using success rate, average episode return, energy cost, collision-related outcome ratio, timeout or battery-depletion ratio, and path length. Robustness is evaluated across increased obstacle density and perturbed wind conditions. The reward function combines a terminal success bonus, a collision penalty, a timeout penalty, a per-step energy penalty, and a dense goal-progress term.

All learning-based methods are trained for the same number of episodes with identical replay capacity and target-network update frequency. The checkpoint yielding the best validation return is selected for final test evaluation. Unless otherwise noted, all tables report mean results over the evaluation episodes.

Figure 4: Training curriculum progression under a fixed 100-episode budget.

P.S: The proposed method advances to higher difficulty levels than the compared learning-based variants.

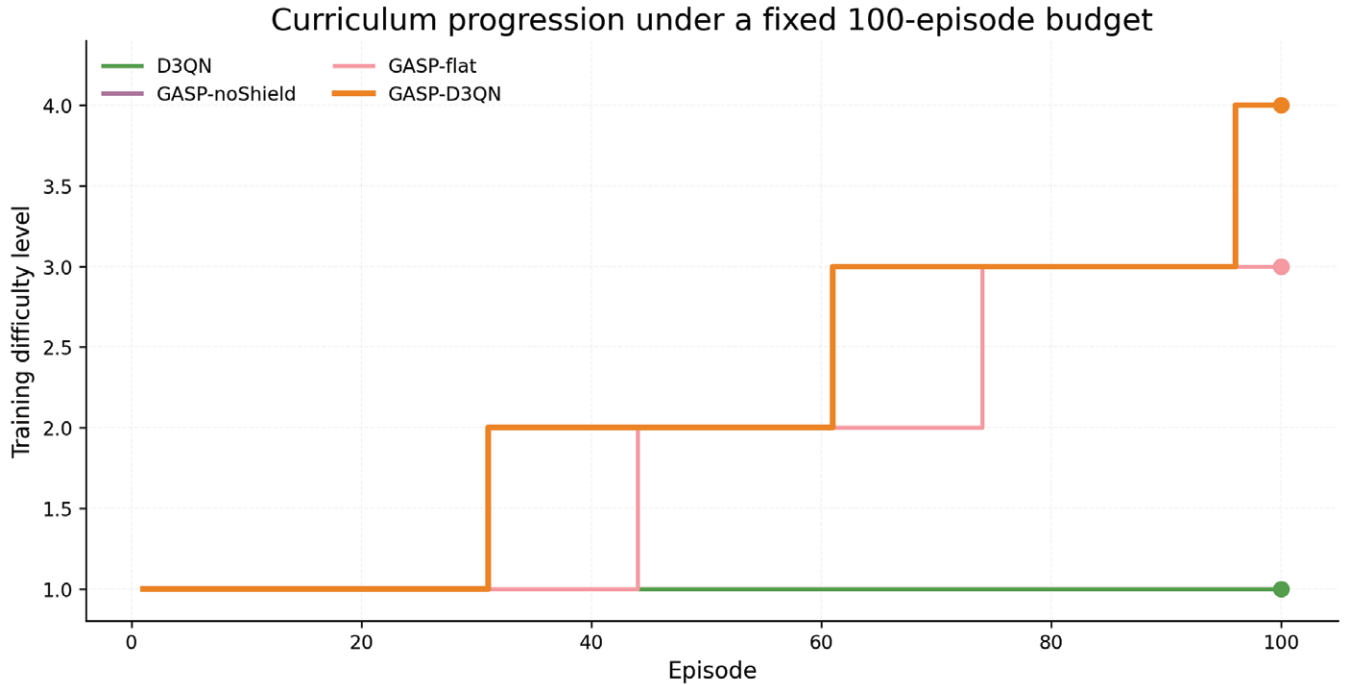


Table 1: Primary benchmark results on the standard test set.

P.S: APF is included as a classical reference; GASP-D3QN is the best-performing learning-based method.

Method	Success	Reward	Energy	Path	Risk
APF	0.62	536.64	495.58	56.12	0.006
DQN	0.02	-68.37	249.4	29.83	0.057
DDQN	0	-81.67	267.74	28.21	0.062
D3QN	0.06	-28.05	326.05	36.77	0.054
PER-D3QN	0	-222.01	123.72	15.07	0.026
Flat-D3QN	0.3	136.51	265.39	32.46	0.048
GASP-D3QN	0.5	383.19	370.61	44.37	0.048

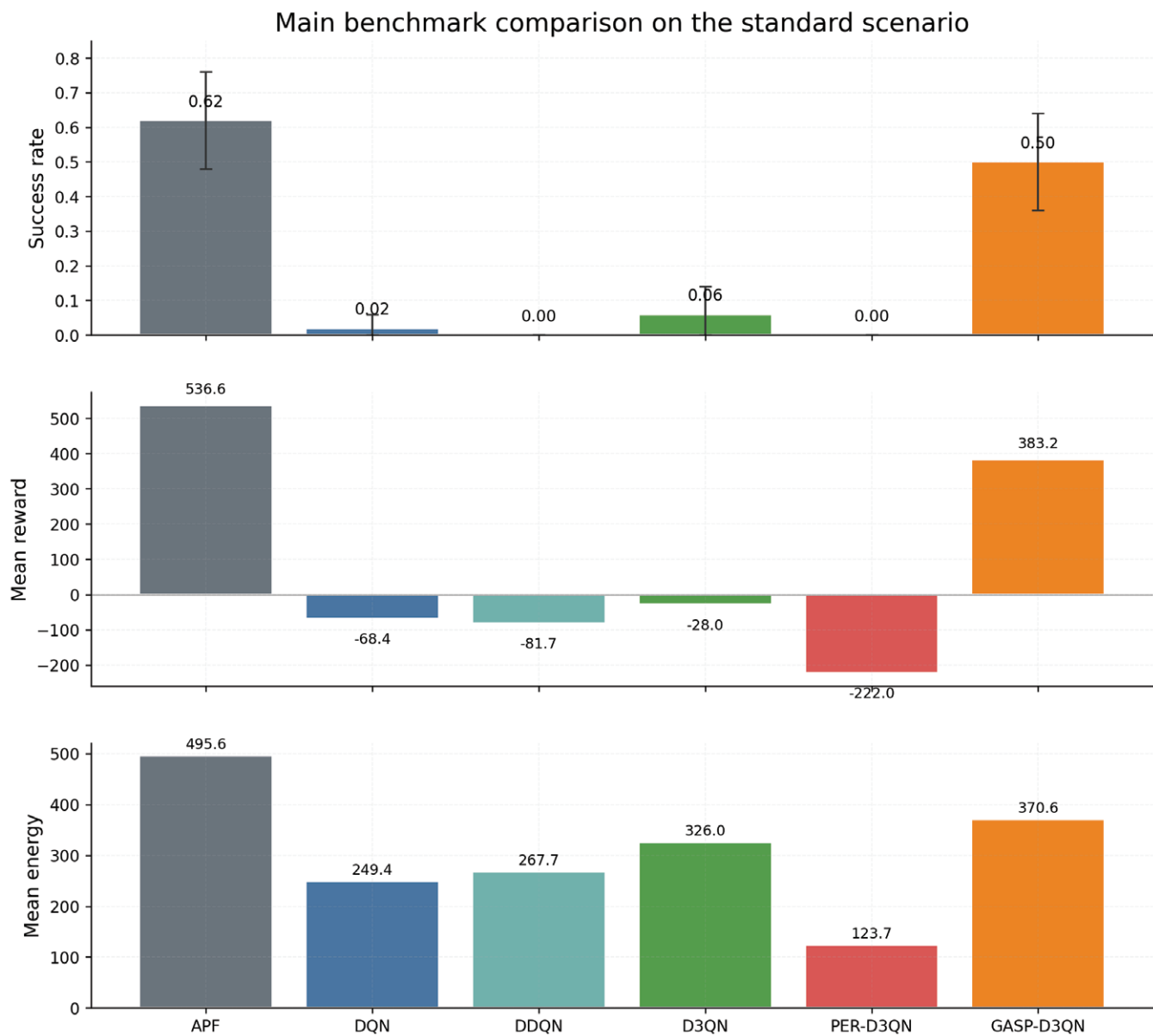
4. Results and analysis

4.1 Primary benchmark

Table 1 and Fig. 5 summarize the results on the standard benchmark. Among the learning-based methods, GASP-D3QN achieves the highest success rate of 0.50, outperforming the strongest structural baseline Flat-D3QN by 20 percentage points and D3QN by 44 percentage points. It also attains the highest average return among the learning-based methods. The APF reference reaches a higher success rate and reward, but it does so with substantially higher energy consumption and belongs to a different methodological family. These results indicate that GASP-D3QN offers the strongest overall trade-off within the learning-based online decision setting.

The poor results of DQN and DDQN suggest that plain value-learning architectures struggle to extract sufficient structure from the 3-D navigation state. D3QN improves upon both, but its performance remains limited. PER-D3QN underperforms because replay prioritization is dominated by early collision-heavy trajectories, which biases learning toward short and frequently unsuccessful episodes. In contrast, GASP-D3QN benefits from the combination of explicit geometry modeling and structured action-space shaping.

Figure 5: Comparison of success rate, average return, and energy cost on the standard benchmark.



4.2 Robustness under stress and wind shift

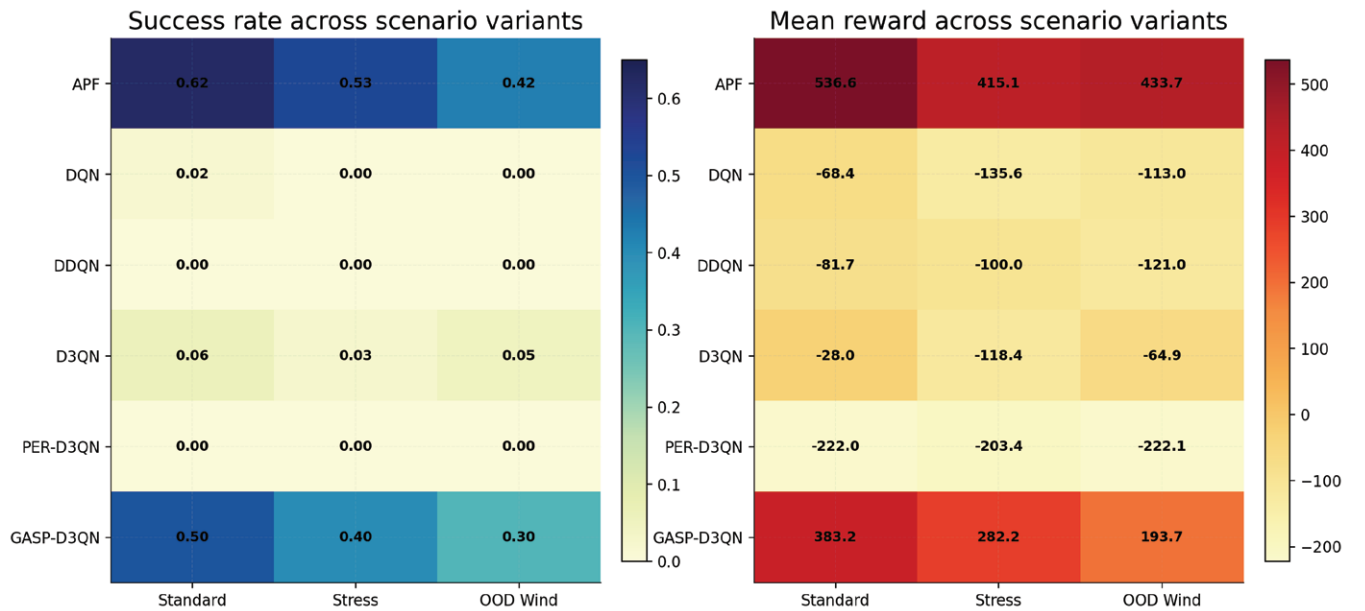
Robustness results are listed in Table 2 and illustrated in Fig. 6. The relative performance ordering remains stable when the

obstacle field becomes denser or when evaluation includes wind conditions outside the nominal training range. GASP-D3QN preserves a success rate of 0.40 on the stress set and 0.30 on the OOD-wind set, whereas the non-structured DQN variants remain close to zero. This behavior suggests that geometry-aware state representation becomes particularly important when free-space corridors are narrow or when drift compensation is necessary.

Table 2: Robustness results across scenario variants.

Method	Standard Succ.	Standard Energy	Stress Succ.	Stress Energy	OOD wind Succ.	OOD wind Energy
APF	0.62	495.58	0.53	440.12	0.42	482.23
DQN	0.02	249.4	0	171.42	0	186.2
DDQN	0	267.74	0	231.11	0	211.54
D3QN	0.06	326.05	0.03	235.05	0.05	257.55
PER-D3QN	0	123.72	0	120.64	0	105.91
GASP-D3QN	0.5	370.61	0.4	310.65	0.3	281.59

Figure 6: Success rate and mean reward across the standard, stress, and OOD-wind settings.



4.3 Outcome composition and ablation

The outcome composition shown in Fig. 7 indicates that GASP-D3QN substantially reduces destructive terminations relative to the weaker learning-based baselines while preserving a comparatively large fraction of successful episodes. This observation is consistent with the intended role of the safety shield, namely filtering evidently unsafe actions before they generate destructive transitions.

The ablation results in Table 3 and Fig. 8 confirm that each component plays a meaningful role. Removing the safety shield decreases standard-set success from 0.50 to 0.06 and sharply degrades average return. Eliminating the action prior causes training to collapse, resulting in zero success on the standard benchmark. Replacing the hybrid encoder with a flat encoder lowers success to 0.30 and reduces return substantially. These results demonstrate that the final performance gain does not arise from a single isolated mechanism, but from the interaction among geometry-aware representation, explicit safety filtering, and action-guided priors.

Figure 7: Outcome composition by method and scenario. The proposed method retains a larger success fraction than the competing learning-based baselines across the evaluated settings.

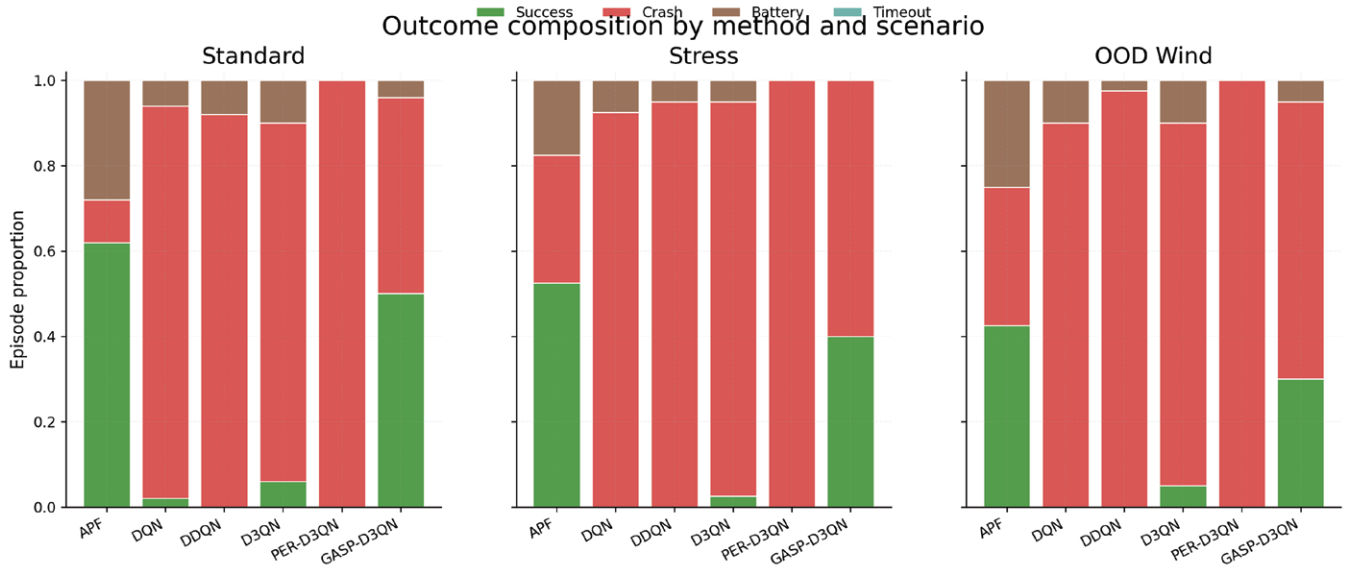


Table 3: Ablation results on the standard benchmark.

Variant	Success	Reward	Energy	Risk
GASP-D3QN	0.5	383.19	370.61	0.048
GASP-noShield	0.06	-60.77	225.62	0.031
GASP-noPrior	0	-239	124.23	0.015
GASP-flat	0.3	136.51	265.39	0.048

4.4 Efficiency frontier

Figure 9 visualizes the trade-off between success rate and energy cost, with bubble size proportional to mean path length. GASP-D3QN occupies the most favorable region among the learning-based methods: it achieves the highest success rate and the strongest return while avoiding the very high energy cost of the APF reference. The training-efficiency statistics reported in Table 4 further show that the proposed model reaches the highest difficulty level under the fixed 100-episode budget, which is advantageous for computationally constrained UAV development cycles.

Table 4: Training-efficiency statistics under the fixed 100-episode budget.

Method	Final level	Last-20 reward	Last-20 success	Last-20 steps
DQN	1	-201.2	0	32.75
DDQN	1	-193.23	0	35
D3QN	1	-215.95	0	33.85
PER-D3QN	1	-198.06	0	41.4
GASP-noShield	3	324.45	0.3	33.5
GASP-noPrior	1	-192.59	0	38.45
GASP-flat	3	451.18	0.5	36.7
GASP-D3QN	4	432.03	0.5	35.05

Figure 8: Ablation trends with respect to success rate and average return.

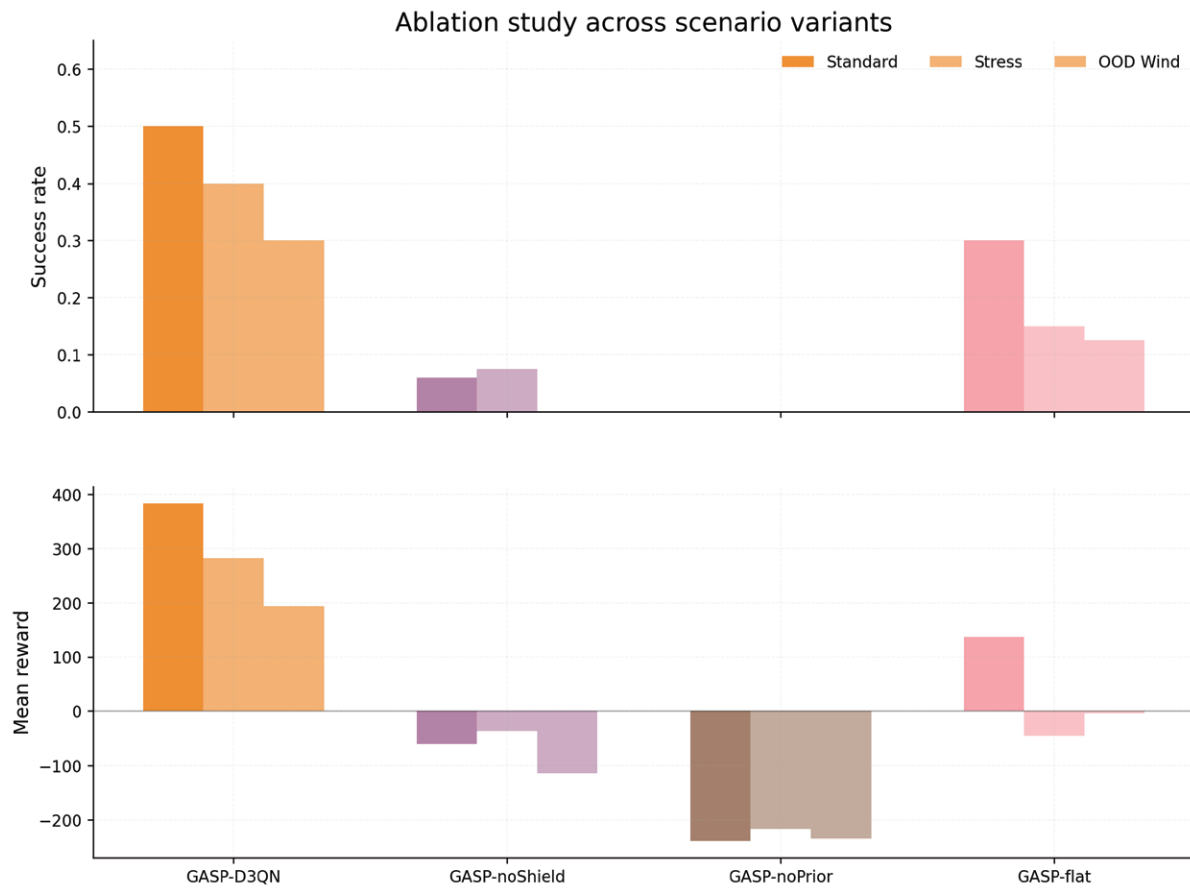
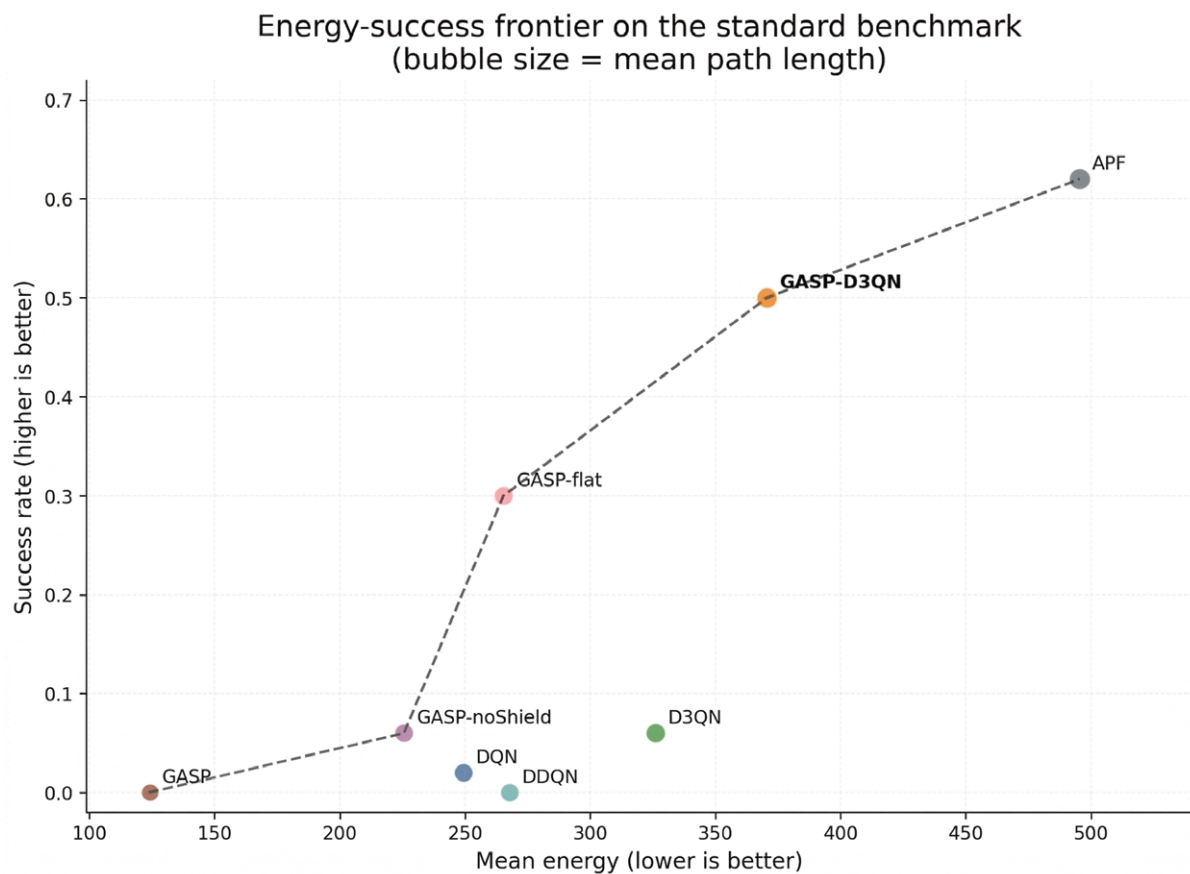


Figure 9: Energy-success frontier on the standard benchmark, where bubble size is proportional to mean path length.



5. Discussion

Two practical observations emerge from the present study. First, hard safety filtering is substantially more effective than delegating all collision handling to reward penalties alone. Although this conclusion is intuitive from a control-theoretic perspective, it is not always reflected in end-to-end value-learning pipelines. Second, lightweight task priors improve data efficiency when they are incorporated directly into the action-selection stage rather than being embedded exclusively in the reward function. In the proposed framework, the safety shield and action prior function as complementary inductive structures: the shield excludes obviously invalid candidates, while the prior ranks the remaining safe actions in a manner that accelerates useful exploration.

This study also has several limitations. The action space is discrete, the benchmark is simulation-based, and the local geometry encoder operates on a compact occupancy cube rather than richer onboard sensing modalities. These limitations do not affect the validity of the comparative results within the present experimental setting, but they point to important future directions, including continuous-control extensions, stronger disturbance models, and hardware-in-the-loop or real-flight validation.

6. Conclusion

This paper presented GASP-D3QN, a geometry-aware and safety-prioritized dueling double-Q framework for online 3-D UAV path planning. By integrating a hybrid state encoder, a hard safety mask, and a prior-guided action selector, the proposed method achieves the best overall performance within the learning-based comparison set. Robustness and ablation experiments further show that the observed gains are systematic and attributable to the coordinated effects of geometric representation, safety-aware action filtering, and structured action priors. These findings suggest that explicit geometric structure and safety-oriented decision mechanisms are essential ingredients for practical learned UAV navigation under partial observation.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Dubins, L. E. (1957). On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents. *American Journal of Mathematics*, 79(3), 497–516.
- [2] Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2), 100–107.
- [3] Khatib, O. (1986). Real-time obstacle avoidance for manipulators and mobile robots. *The International Journal of Robotics Research*, 5(1), 90–98.
- [4] Stentz, A. (1994). Optimal and efficient path planning for partially-known environments. *Proceedings of the IEEE International Conference on Robotics and Automation*, 4, 3310–3317.
- [5] LaValle, S. M., & Kuffner Jr., J. J. (2001). Randomized kinodynamic planning. *The International Journal of Robotics Research*, 20(5), 378–400.
- [6] Karaman, S., & Frazzoli, E. (2011). Sampling-based algorithms for optimal motion planning. *The International Journal of Robotics Research*, 30(7), 846–894.
- [7] Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of the IEEE International Conference on Neural Networks*, 4, 1942–1948.
- [8] Dorigo, M., & Gambardella, L. M. (1997). Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, 1(1), 53–66.
- [9] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., &

- Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.
- [10] van Hasselt, H., Guez, A., & Silver, D. (2016). Deep reinforcement learning with double Q-learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), 2094–2100.
- [11] Wang, Z., Schaul, T., Hessel, M., van Hasselt, H., Lanctot, M., & de Freitas, N. (2016). Dueling network architectures for deep reinforcement learning. *Proceedings of the 33rd International Conference on Machine Learning*, 48, 1995–2003.
- [12] Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2016). Prioritized experience replay. *Proceedings of the International Conference on Learning Representations*.
- [13] Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.
- [14] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
- [15] Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *Proceedings of the 35th International Conference on Machine Learning*, 80, 1861–1870.
- [16] Garcia, J., & Fernandez, F. (2015). A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(42), 1437–1480.
- [17] Achiam, J., Held, D., Tamar, A., & Abbeel, P. (2017). Constrained policy optimization. *Proceedings of the 34th International Conference on Machine Learning*, 70, 22–31.
- [18] Krishnan, S., Boroujerdian, B., Fu, W., Faust, A., & Reddi, V. J. (2021). Air learning: A deep reinforcement learning gym for autonomous aerial robot visual navigation. *Machine Learning*, 110, 2501–2540.
- [19] Puente-Castro, A., Rivero, D., Pazos, A., & Fernandez-Blanco, E. (2022). UAV swarm path planning with reinforcement learning for field prospecting. *Applied Intelligence*, 52, 14101–14118.
- [20] Jiang, Y., Xu, X.-X., Zheng, M.-Y., & Zhan, Z.-H. (2024). Evolutionary computation for unmanned aerial vehicle path planning: A survey. *Artificial Intelligence Review*, 57, Article 267.
- [21] Zhang, D., Xuan, Z., Zhang, Y., Yao, J., Li, X., & Li, X. (2023). Path planning of unmanned aerial vehicle in complex environments based on state-detection twin delayed deep deterministic policy gradient. *Machines*, 11(1), Article 108.
- [22] Liu, J., Luo, W., Zhang, G., & Li, R. (2025). Unmanned aerial vehicle path planning in complex dynamic environments based on deep reinforcement learning. *Machines*, 13(2), Article 162.

A Review of Lightweight Research on 3D Printed Prosthetic Structure Based on Topology Optimization

Kangdi Zhong^{1,2,3*}, Lian Xue^{1,2}, Wen Sun^{1,2}, Haoyu Wang¹

1. School of Aeronautics, Changji University, Changji, 831100, China

2. Xinjiang Key Laboratory of Intelligent Control and Safety Assurance Technology for Low-Altitude Airspace, Changji University, Changji, 831100, China

3. Key Laboratory of Aircraft Design, Manufacture and Maintenance, Changji University, Changji, 831100, China

*Corresponding author: Kangdi Zhong, kangdizh@163.com

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: The lightweight of prosthetic limb is the key to improve the wearing comfort and user experience. The integration of topology optimization and 3D printing technology provides a new path for it. This paper reviews the application status, key technologies and existing problems of topology optimization and 3D printing in the field of lightweight prosthetic limbs. The research shows that the integration of topology optimization and 3D printing is an important direction for the lightweight and customization of prostheses. At present, there is still room for improvement in algorithm adaptation and process coordination. In the future, it is necessary to promote the landing of technology through multidisciplinary cross-disciplinary to provide better rehabilitation aids for the disabled.

Keywords: Topology Optimization; 3D Printing; Prosthetic Limbs; Lightweight Structure

Published: Apr 17, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1289>

1. Introduction

With the increasing demand for rehabilitation of disabled people and the acceleration of population aging, prosthetic limbs have attracted attention as core rehabilitation aids. Although breakthroughs have been made in related research, there are still many bottlenecks in traditional prostheses: solid structure design leads to large weight and discomfort in wearing, traditional manufacturing processes are difficult to achieve complex lightweight structures, and personalized customization has a long cycle and high cost, which cannot meet the user's ergonomic and biomechanical needs.

As an advanced structural optimization design method, topology optimization technology realizes the optimal distribution of structural materials through algorithm iteration under given load and constraint conditions, and can remove redundant materials to the maximum extent without reducing mechanical properties^[1, 2]. 3D printing technology breaks through the limitations of traditional subtractive manufacturing. It has the advantages of high customization, fast forming speed and strong manufacturability of complex structures. It can accurately reproduce the complex lightweight structure of topology optimization design and solve the limitations of traditional manufacturing process^[2-4]. The e-NABLE project launched in 2011 is one of the pioneers of 3D printing low-cost children's prostheses, which proves that functional instruments can be produced at a small cost of traditional prostheses, leading to the popularity of customized robot prostheses^[2, 5]. The deep integration

of topology optimization technology and 3D printing technology has become an effective way to solve the lightweight and customized pain points of traditional prostheses.

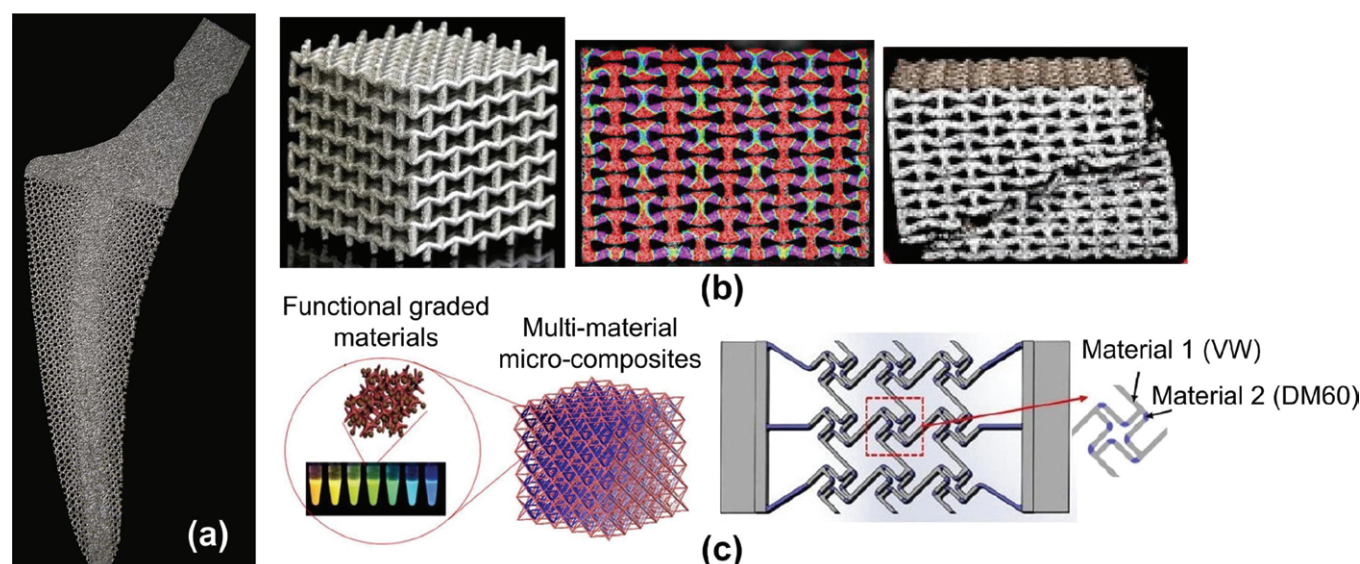
Focusing on the core line of topology optimization-3D printing-lightweight prostheses, this paper systematically reviews the research status in this field, summarizes the key technologies, research results and existing deficiencies in the current research, in order to promote the deep application of topology optimization and 3D printing technology in the field of lightweight prostheses.

2. Topology Optimization of Lightweight Prosthetic Structure

2.1 Principle and Method of Topology Optimization Technology

Topology optimization is an advanced design method based on structural mechanics and mathematical optimization theory. Its core objective is to determine the optimal distribution of materials through optimization algorithm iteration under given design space, load conditions and performance constraints, so as to achieve the balance between “lightest weight” and “optimal performance”. Compared with size optimization and shape optimization, topology optimization does not need to presuppose structural form, can independently generate the optimal structural layout, and is more likely to break through the limitations of traditional design thinking. It is especially suitable for complex structural design such as prostheses that need to take into account both mechanical properties and lightweight. Vogiatzis et al.^[6] used the level set method to optimize the topology of single-material and multi-material negative Poisson's ratio microcomposites. Based on isogeometric analysis, Gao et al.^[7] realized isogeometric topology optimization to obtain negative Poisson's ratio microcomposites. Zhang et al.^[8] combined the density-based topology optimization method with the mixed stress/deformation-driven nonlinear homogenization method to design microcomposites with negative Poisson's ratio behavior in a large strain range. In addition, as shown in Figure 1, functional micro-composites with maximum effective bulk modulus (EBM) have received extensive attention and are widely used in load-bearing structures^[2, 9, 10].

Figure 1: (a) Microstructures in biomaterials applications^[2], (b) Functional microstructures during deformation and failure^[10], and (c) Multi-material microcomposites^[10]



At present, the topology optimization methods for lightweight design mainly include density method^[11], level set method^[12], genetic algorithm^[13], etc., as shown in Table 1. The variable density method has become the most commonly used optimization method in related research because of its simple principle and high computational efficiency. The level set method and genetic algorithm play an important role in the surface accuracy control of the prosthetic limb and multi-objective optimization respectively.

Table 1: Topology optimization method

Typical Topology Optimization Methods	Advantages	Disadvantages
Density Method	Simple principle, high computational efficiency, easy to implement, and wide application	Low boundary precision, requiring subsequent smoothing treatment
Level Set Method	Precise boundary control, suitable for high-precision requirements	Complex calculation, low efficiency, and application in the exploration stage
Genetic Algorithm	Strong global search ability, capable of handling multi-objective optimization	Low computational efficiency, many iterations, affecting the design cycle

2.2 Topology Optimization Constraints of Lightweight Prosthetic Structure

As a rehabilitation assistive device in direct contact with the human body, the topology optimization design of the prosthetic limb needs to take into account multiple constraints such as mechanical properties, ergonomics, and biocompatibility. Mechanical performance constraint is the core. It is necessary to ensure that the prosthetic limb has sufficient strength, stiffness and stability under static load and dynamic load, so as to avoid fracture and deformation during use. Ergonomics constraints require the optimized prosthetic structure to fit the shape of the human body, with accurate size and comfortable wearing, so as to avoid local compression, friction and other problems^[14]. Biocompatibility constraints require the topology optimization process to combine the characteristics of 3D printing materials to ensure that the optimized structure is compatible with the printing material^[15]. RASZEWSKI et al.^[15] have shown that it aims to prepare a bioactive 3D printing material that can be used in transparent removable orthodontic appliances. By adding bioactive glass to the acrylic monomer, the performance and cytotoxicity tests show that the material has stable ion release, good mechanical properties and good cell compatibility, and can be used for orthodontic appliances.

In addition, the topology optimization of the prosthesis also needs to consider the manufacturing constraints, that is, the optimized structure needs to meet the requirements of the 3D printing process, avoid the complex structure that cannot be formed (excessive small support structure, closed cavity, etc.), and ensure that the lightweight structure can be accurately realized by 3D printing. The coupling of multiple constraints makes the topology optimization design of prostheses more complex and challenging, and has become the focus and difficulty of current research.

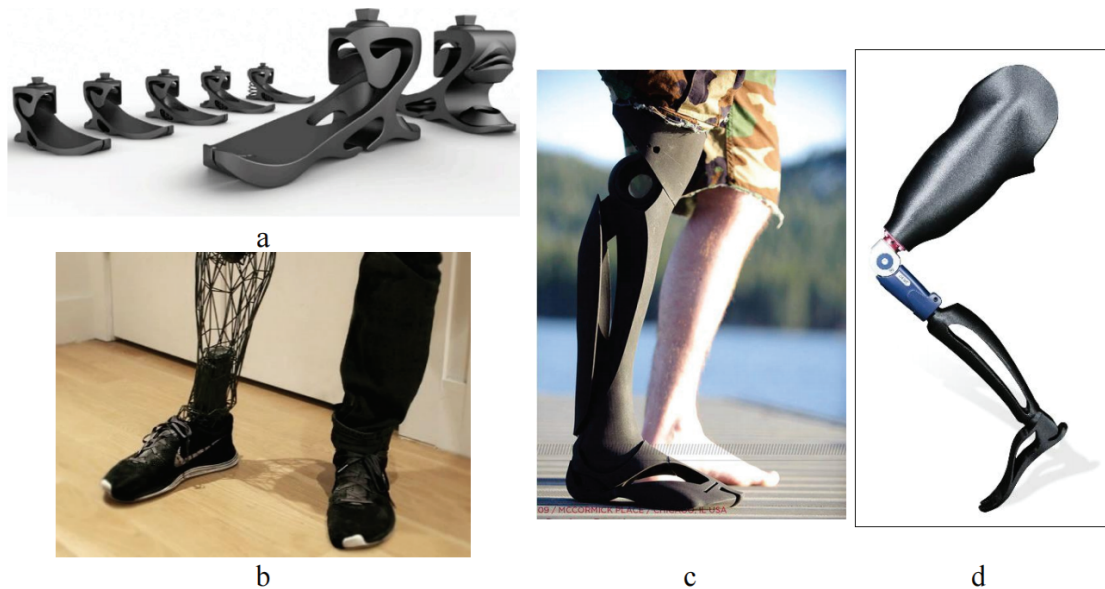
3. Lightweight Design of 3D Printed Prostheses Based on Topology Optimization

At present, the integration of topology optimization and 3D printing technology has become the core path of lightweight and personalized design of prosthetic structure, forming a series of research results, but also showing different research focuses and common problems.

3.1 Lightweight Structure Design and Optimization Method

In the application of lightweight structure design and optimization methods, researchers have focused on combining the biomechanical properties of prosthetic limbs with ergonomic requirements to explore the adaptive application of topology optimization technology. Generico Chair, designed by Marco Hemmerlin et al.^[16], applies the topology optimization idea system to the chair structure design. By optimizing the material distribution, the weight is significantly reduced while ensuring the mechanical reliability. Current research has explored new biomaterials such as shape memory polymers and piezoelectric composites, which are expected to achieve adaptive stiffness adjustment and embedded sensing functions of the next generation of prostheses. By matching material properties with functional requirements such as quality reduction, energy feedback, and interface comfort, the designer successfully developed a prosthetic foot structure that combines performance optimization and patient experience improvement. Figure 2 shows a prototype device based on representative materials^[17].

Figure 2: 3D printed prosthesis example^[17]: (a) polylactic acid (PLA), (b) titanium alloy lattice structure, (c) CFRP, (d) composite



3.2 Personalized Customization and 3D Printing Process

In terms of personalized customization and process adaptation, relying on the customization advantages of 3D printing, combined with topology optimization to achieve precise design and efficient manufacturing of prostheses. Figure 3 is the flow chart of the production of prosthetic limbs using 3D printing technology. Based on biomechanics and 3D printing technology, researchers collected three-dimensional data of human limbs, established personalized biomechanical models, and combined topology optimization with selective laser sintering (SLS), fused deposition modeling (FDM) and other processes (Figure 4) to create lightweight prostheses that meet individual needs^[17]. Some studies focused on process improvement and cost control, and formed targeted results : researchers carried out ankle prosthesis design, combined with topology optimization to optimize joint structure, and improved the activity performance and lightweight level of prostheses^[18]. Another study, from the perspective of embodiment, combined with topology optimization and selective laser sintering process, designed a child prosthesis that takes into account both lightweight and comfort, and solves the pain points of children's prosthetic limbs with high personalized needs and fast growth and development^[19]; 3D technology provides new opportunities for imitating and manufacturing multi-scale, multi-material and multi-functional structures in nature, and provides high-quality material support for topologically optimized prosthetic structures, further improving lightweight effect and biocompatibility^[20].

Figure 3: The process of using 3D printing technology to manufacture prosthetic limbs^[21]

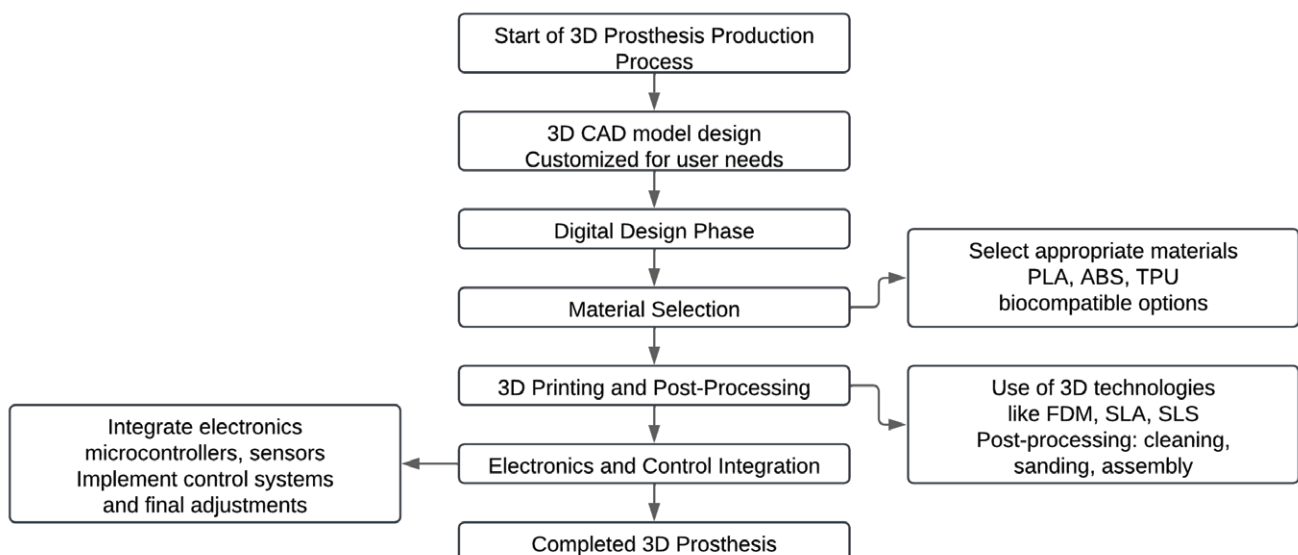
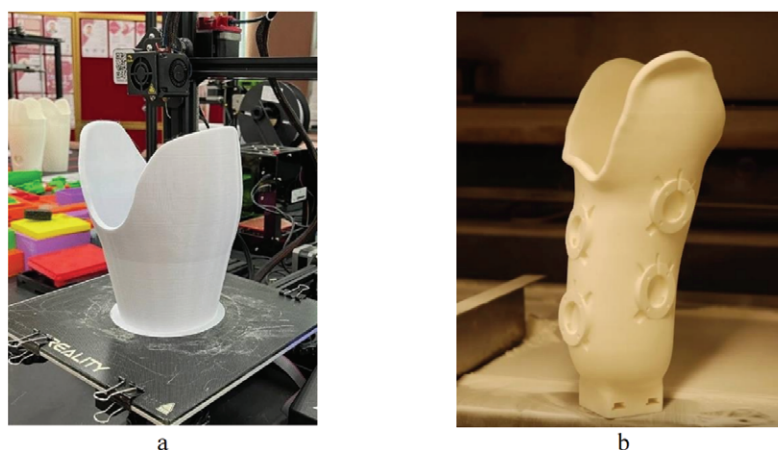


Fig. 4 Representative additive manufacturing cases^[17]: (a) FDM printed prosthesis socket; (b) Prosthetic sockets were made by SLS



4. Key Technologies and Research Hotspots

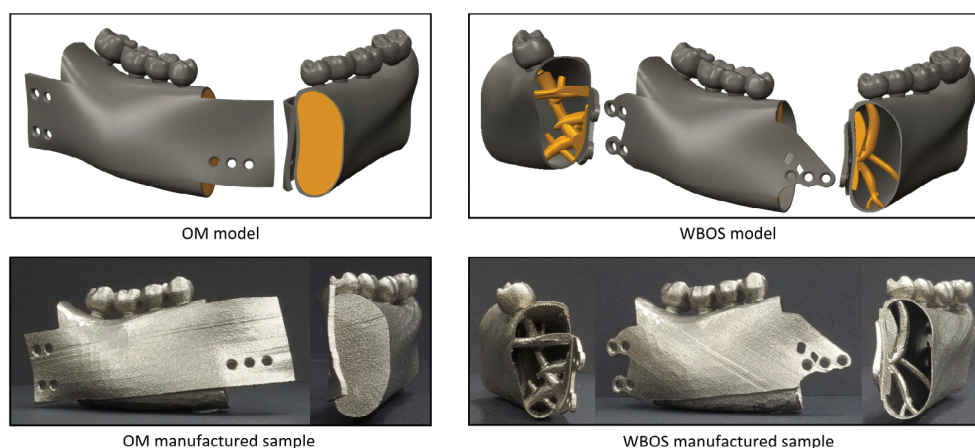
4.1 Multi-objective Collaborative Design of Prosthetic Topology Optimization

At present, the topology optimization of lightweight prosthetic structure has shifted from a single weight optimization to a multi-objective optimization of “lightweight+mechanical properties+ comfort+manufacturing feasibility”, which has become a research hotspot. The core of multi-objective collaborative design is to establish a reasonable optimization objective function and balance the contradiction between the constraints. For example, the balance between lightweight and mechanical properties, the balance between individual adaptation and manufacturing feasibility^[22, 23]. The researchers solved the conflict problem in multi-objective optimization by introducing weighting coefficient method, Pareto optimal solution and other methods to ensure that the optimized prostheses not only meet the lightweight requirements, but also have good mechanical properties and wearing comfort.

4.2 Synergy of Topology Optimization and 3D Printing Process

The collaborative adaptation of topology optimization and 3D printing process is the key to realize the lightweight design of prosthetic limbs. On the one hand, it is necessary to adjust the constraints of topology optimization according to the characteristics of 3D printing process^[24, 25], so as to avoid designing complex structures that cannot be formed. As shown in Figure 5, Li et al.^[24] used a weighted topology optimization method to design a specific intimal implant for patients with severe inferior membrane defects, providing good biomechanical properties and appearance repair for oral rehabilitation. On the other hand, it is necessary to improve the forming quality of the topologically optimized structure by optimizing the 3D printing process parameters to ensure that the formed prosthetic structure is consistent with the optimized design. At present, the commonly used collaborative adaptation methods include topology optimization design under process constraints, matching test of printing parameters and optimized structure, etc., which has become the focus of current research.

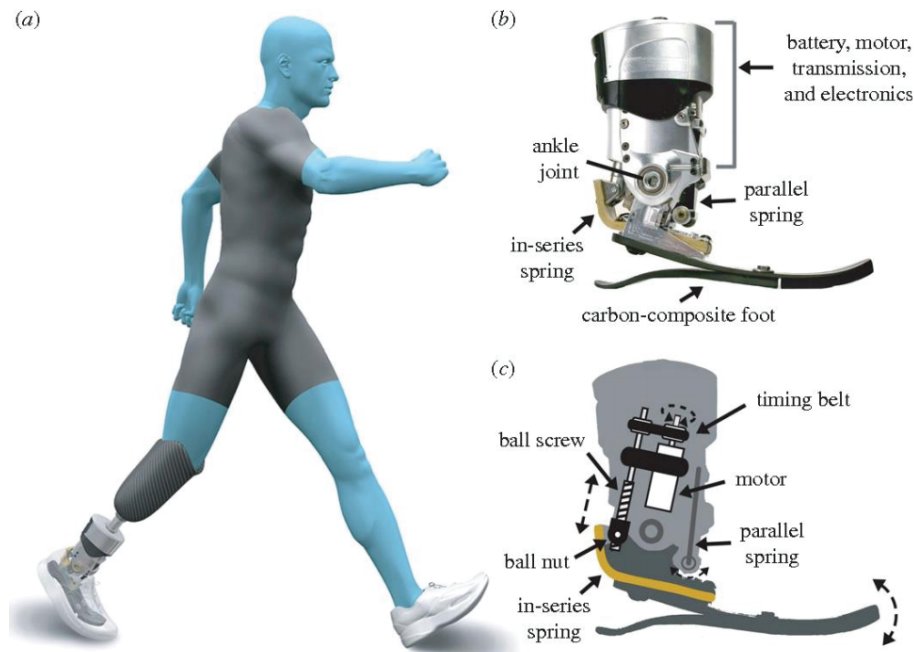
Figure 5: The model before and after the final optimization, and the corresponding physical model was made by metal 3D printing technology^[24]



4.3 Topology Optimization Based on Biomechanics

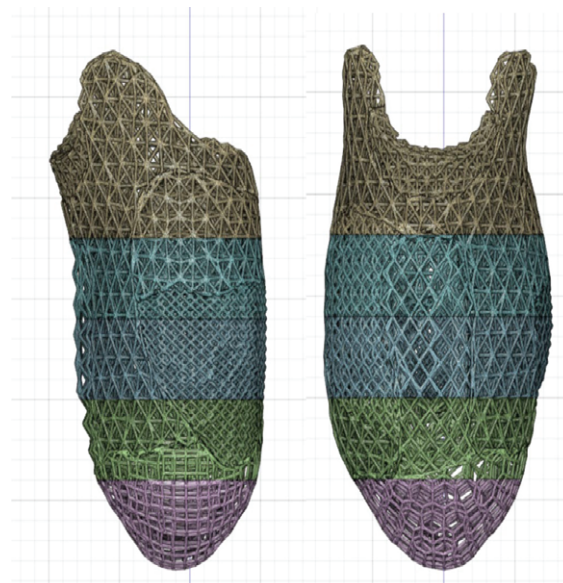
The performance of prostheses is closely related to the biomechanical properties of human body. Personalized topology optimization based on human biomechanical data is the core technology to improve the adaptability of prostheses. Researchers collect multi-dimensional biomechanical information of users, construct personalized models and integrate them into the whole process of topology optimization, so that the prosthetic structure is more suitable for human body shape and exercise habits, and significantly improves wearing comfort and exercise flexibility. Through the wear test of tibial amputees, the fitness, comfort and stability of the prostheses were improved, which confirmed the clinical applicability of the optimized design^[26]. The synergistic effect of the comprehensive verification method ensures that the 3D printing topology-optimized prosthetic foot meets the requirements of performance, durability and user experience (Figure 6).

Figure 6: Assessment of fitness and gait by transtibial amputee test^[17]



In addition, combined with finite element simulation technology, the mechanical response of prostheses under different motion conditions can be simulated, and the structural design can be further optimized to ensure the reliability and safety of prostheses. In the topology optimization study of the prosthetic socket (Figure 7), different lattice structures are proposed to simulate the stiffness of different regions of the flexible socket^[27].

Figure 7: Different lattice structures are used to simulate the stiffness of different regions of the flexible receiving cavity^[27]



5. Development Trends

At present, there are many deficiencies in the lightweight research of 3D printed prosthetic structure based on topology optimization, and the future development needs to be targeted.

- Innovate a dedicated optimization algorithm to improve the adaptability and computational efficiency of the algorithm to the design of prosthetic limbs.
- Promote the deep coordination of topology optimization and 3D printing process, optimize printing materials and process parameters, and solve the problem of structural forming.
- Strengthen clinical verification and industry-university-research cooperation, promote the implementation of technical engineering, reduce manufacturing costs, and achieve large-scale production.

Conclusion

The collaborative integration of topology optimization and 3D printing technology effectively breaks through the bottleneck of traditional prosthetic design and manufacturing, and constructs a new technical system for lightweight and personalized design and manufacturing of prosthetic limbs, which significantly improves the mechanical properties, wearing comfort and customization level of prosthetic limbs. This paper reviews the research results in this field, clarifies the key technologies such as the core method of topology optimization, multi-constraint design and process collaboration, and summarizes the existing problems in algorithm adaptation, process collaboration and clinical verification. In the future, we need to focus on the direction of personalized and engineering development, promote the large-scale implementation of technology through technological innovation and industry-university-research collaboration, realize the high-quality development of prosthetic rehabilitation aids, and provide rehabilitation support that is more suitable for the needs of the disabled.

Funding

Supported by Natural Science Foundation of Xinjiang Uygur Autonomous Region (2024D01C210); Science and Technology Plan Project of Changji Prefecture (2025S01-03); Research Project of Changji University (KY2024024). The Undergraduate Innovation and Entrepreneurship Training Program (S202510997027).

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Eschenauer, H. A., Olhoff, N. (2001). Topology optimization of continuum structures: A review. *Applied Mechanics Reviews*, 54(4): 331-390. 10.1115/1.1388075
- [2] Garner, E., Kolken, H. M. A., Wang, C. C. L. et al. (2019). Compatibility in microstructural optimization for additive manufacturing. *Additive Manufacturing*, 26: 65-75. 10.1016/j.addma.2018.12.007
- [3] Gretsche, K. F., Lather, H. D., Peddada, K. V. et al. (2016). Development of novel 3D-printed robotic prosthetic for transradial amputees. *Prosthetics & Orthotics International*, 40(3): 400-403. 10.1177/0309364615579317
- [4] Vijayavenkataraman, S., Fuh, J., Lu, W. (2017). 3D Printing and 3D Bioprinting in Pediatrics. *Bioengineering*, 4(3): 63. 10.3390/bioengineering4030063
- [5] Chua, C. K., Leong, K. F., 3D Printing and additive manufacturing: Principles and applications (with companion media pack)-of rapid prototyping. World Scientific Publishing Company; 2014.
- [6] Vogiatzis, P., Chen, S., Wang, X. et al. (2017). Topology optimization of multi-material negative Poisson's ratio metamaterials using a reconciled level set method. *Computer Aided Design*, 83: 15-32. 10.1016/j.cad.2016.09.009
- [7] Gao, J., Xue, H., Gao, L. et al. (2019). Topology optimization for auxetic metamaterials based on isogeometric analysis. *Computer Methods in Applied Mechanics and Engineering*, 352: 211-236. 10.1016/j.cma.2019.04.021
- [8] Zhang, G., Khandelwal, K. (2019). Computational design of finite strain auxetic metamaterials via topology optimization and nonlinear homogenization. *Computer Methods in Applied Mechanics and Engineering*, 356: 490-527. 10.1016/j.cma.2019.07.027

- [9] Zhou, H., Zhao, M., Ma, Z. et al. (2020). Sheet and network based functionally graded lattice structures manufactured by selective laser melting: Design, mechanical properties, and simulation. *International Journal of Mechanical Sciences*, 175: 105480. 10.1016/j.ijmecsci.2020.105480
- [10] Chen, Y., Ye, L., Xu, C. et al. (2021). Multi-material topology optimisation of micro-composites with reduced stress concentration for optimal functional performance. *Materials & Design*, 210: 110098. 10.1016/j.matdes.2021.110098
- [11] Ma, W., Lu, Y., Wang, P. et al. (2023). Double Optimization Design of the Formula Racing Car Frame Based on the Variable Density Method and the Joint Variable Method. *Applied Sciences*, 13(18): 10155. 10.3390/app131810155
- [12] Lestari, W. D., Adyono, N., Faizin, A. K. et al. (2024). Optimization of 3D printed parameters for socket prosthetic manufacturing using the taguchi method and response surface methodology. *Results in Engineering*, 21: 101847. 10.1016/j.rineng.2024.101847
- [13] You, W. S., Casebier, J., Mandich, J. et al. (2021). Genetic Algorithm-Based Optimization for the Geometric Design of a Novel Orthopedic Implant. *Ieee Transactions On Bio-Medical Engineering*, 68(12): 3620-3627. 10.1109/TBME.2021.3080226
- [14] Muslim, K., Shafira, A., Kurnianto, R. R. et al. (2025). Enhancing Above-Knee Prosthetic Design for Inclusive Workplaces: Ergonomic Considerations in Manual Material Handling. *Jurnal Optimasi Sistem Industri*, 24(1): 102-120. 10.25077/josi.v24.n1.p102-120.2025
- [15] Raszewski, Z., Chojnacka, K., Kulbacka, J. et al. (2023). Mechanical Properties and Biocompatibility of 3D Printing Acrylic Material with Bioactive Components. *Journal of Functional Biomaterials*, 14(1): 13. 10.3390/jfb14010013
- [16] Hemmerling, M., Nether, U. (2014). Generico: A case study on performance-based design. *Design Proceedings*, 1(8): 126-129. 10.5151/despro-sigradi2014-0021
- [17] Lestari, W. D. L., Ariadi, Y., Putra, A. (2026). Customized Prosthetic Feet via Topology Optimization and 3D Printing: A Critical Review. *Advanced Mechanical and Mechatronic Systems*, 2(1): 33-49. 10.53623/amms.v2i1.776
- [18] Liu, X., Yue, Y., Wang, C. et al. (2022). Topology Optimization Method for Calcaneal Prosthesis. *Journal of Shanghai Jiaotong University (Science)*, 27(2): 240-249. 10.1007/s12204-021-2324-2
- [19] Tang, Z., Yao, X., Feng, W. (2022). Design Exploration of 3D Printed Prosthesis for Children from an Embodied Perspective. *Chinese Art*, (5): 47-53.
- [20] Yang, Y., Song, X., Li, X. et al. (2018). Recent Progress in Biomimetic Additive Manufacturing Technology: From Materials to Functional Structures. *Advanced Materials*, 30(36): 1706539. 10.1002/adma.201706539
- [21] Salazar, M., Portero, P., Zambrano, M. et al. (2025). Review of Robotic Prostheses Manufactured with 3D Printing: Advances, Challenges, and Future Perspectives. *Applied Sciences*, 15(3): 1350. 10.3390/app15031350
- [22] Compton, B. G., Lewis, J. A. (2014). 3D-Printing of Lightweight Cellular Composites. *Advanced Materials*, 26(34): 5930-5935. 10.1002/adma.201401804
- [23] Jin, Z., He, C., Fu, J. et al. (2022). Balancing the customization and standardization: exploration and layout surrounding the regulation of the growing field of 3D-printed medical devices in China. *Bio-Design and Manufacturing*, 5(3): 580-606. 10.1007/s42242-022-00187-2
- [24] Li, C., Wu, C., Lin, C. (2020). Design of a patient-specific mandible reconstruction implant with dental prosthesis for metal 3D printing using integrated weighted topology optimization and finite element analysis. *Journal of the Mechanical Behavior of Biomedical Materials*, 105: 103700. 10.1016/j.jmbbm.2020.103700
- [25] Zhu, J., Hu, J., Zhu, K. et al. (2023). Design of 3D-printed prostheses for reconstruction of periacetabular bone tumors using topology optimization. *Frontiers in Bioengineering and Biotechnology*, 11: 1289363. 10.3389/fbioe.2023.1289363
- [26] Prost, V., Johnson, W. B., Kent, J. A. et al. (2022). Biomechanical evaluation over level ground walking of user-specific prosthetic feet designed using the lower leg trajectory error framework. *Scientific Reports*, 12(1): 5306. 10.1038/s41598-022-09114-y
- [27] Rai, P., Jankiraman, V., Teacher, M. et al. (2022). Design and optimization of a 3D printed prosthetic socket for transtibial amputees. *Materials Today: Proceedings*, 70: 454-464. 10.1016/j.matpr.2022.09.365

Plastic and Concrete Construction Products: A Comparative Review of Material Performance, Cost, Sustainability, and Market Substitution Pathways

Guangtao Zhang^{1*}, Jiahang Ni¹, Junlong Zou¹, Bo Chen¹, Guangkai Zhang²

1. Zhejiang Normal University, Jinhua, Zhejiang, 321004, China

2. The Hong Kong Polytechnic University, Hong Kong, 999077, China

*Corresponding author: Guangtao Zhang, 19710586755@163.com

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: This review examines how plastic and concrete construction products compete, coexist, and increasingly hybridize under contemporary pressures of cost control, installation efficiency, service-life performance, and low-carbon transition. Reorganized from a broader bachelor thesis, the paper focuses on product categories in which substitution is already visible or commercially plausible, including drainage pipes, grass block pavers, modular road concepts, and recycled-plastic composite systems. The analysis is structured around five dimensions: intrinsic material characteristics, processing and installation routes, cost structure, application scenario, and sustainability performance. Plastics show clear advantages in low density, corrosion resistance, standardized factory processing, and reduced logistics burden, which supports their expansion in lightweight and modular applications. Concrete retains decisive strengths in compressive capacity, rigidity, fire resistance, and long-term reliability under heavy loading. The review argues that market change does not follow a simple replacement model. Instead, it proceeds through scenario-based substitution, selective coexistence, hybrid integration, and circular upgrading. For construction decision-makers, the central implication is that material choice should be based on service-fit and life-cycle value rather than on nominal material price alone.

Keywords: Plastic Products; Concrete Products; Construction Materials; Comparative Review; Life-Cycle Cost; Sustainability; Market Substitution

Published: Apr 19, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1280>

1. Introduction

Material choice in construction has traditionally been guided by technical performance, local practice, and cost. However, growing attention to maintenance expenditure, carbon emissions, transport efficiency, and circular resource use has made the comparison between conventional and alternative materials more strategically important. Among the most visible material pairings in this transition are concrete and plastics. Both are deeply embedded in modern construction markets, but they compete, coexist, and increasingly hybridize in different ways and at different scales^[1-3].

Concrete remains one of the most widely used construction materials because of its compressive strength, rigidity, fire resistance, and suitability for structural and infrastructure systems^[2,4,5]. Plastic products, by contrast, are valued for low density, corrosion resistance, design flexibility, and ease of mass manufacture. These characteristics have enabled plastics to expand rapidly in

pipings, insulation, modular components, pavers, and other non-structural or semi-structural applications^[1,6,7].

The source thesis contained extensive descriptive background on the history, classification, and processing of both materials, followed by product-by-product description. For journal submission, that structure is too broad and insufficiently analytical. The present paper therefore narrows the discussion to construction products in which substitution is already observable or commercially plausible. The central question is not whether plastics can replace concrete in general, but under what conditions plastic, concrete, or hybrid systems become the more rational market choice.

Accordingly, this review pursues three objectives. First, it compares the material logic of plastics and concrete from the perspectives of properties and processing routes. Second, it evaluates cost structure, application fit, and sustainability implications in representative product categories. Third, it proposes a concise framework of market substitution pathways that explains why some segments move toward plastic, others remain concrete-dominant, and others evolve toward hybrid or circular-material solutions.

2. Review Scope and Analytical Framework

This study is a narrative review rather than a systematic meta-analysis. The source base was substantially upgraded from the original bachelor thesis and now emphasizes peer-reviewed journal literature, life-cycle assessment studies, international policy reports, technical association publications, and documented infrastructure case materials relevant to plastic and concrete construction products^[8–12]. The review retains selected product cases only when they clarify substitution, coexistence, hybridization, or circular-economy pathways within construction markets.

To increase analytical focus, the review does not attempt to compare the entire plastics and concrete industries. Instead, it treats the construction product as the unit of analysis and evaluates five decision-relevant dimensions: (1) intrinsic material characteristics, (2) manufacturing and installation logic, (3) cost structure across the product life cycle, (4) application scenario, and (5) sustainability implications. On this basis, a synthesis of market substitution pathways is proposed.

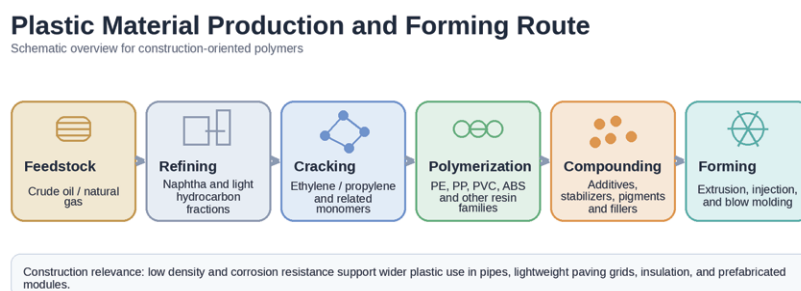
Because the cited sources use different accounting scopes, regions, and performance standards, the paper does not present a unified quantitative cost model. Cost is therefore discussed in qualitative or semi-quantitative terms, with emphasis on transport burden, installation complexity, maintenance demand, replacement risk, and system value rather than on a single nominal market price.

3. Comparative Review and Market Analysis

3.1 Material characteristics and processing routes

Plastics used in construction are mainly polymer-based thermoplastics such as polyethylene (PE), polyvinyl chloride (PVC), polypropylene (PP), and acrylonitrile-butadiene-styrene (ABS). Their common advantages include low density, corrosion resistance, electrical insulation, design adaptability, and ease of standardized factory production^[6,7,13]. Extrusion, injection molding, and blow molding enable plastics to be formed into pipes, fittings, pavers, liners, and modular shells with comparatively low transport burden and consistent geometry^[6,7]. As a result, plastic products often perform best where lightweight handling, chemical stability, and rapid installation are important. As shown in Figure 1, the plastic route is characterized by a refinery-to-polymer-to-forming chain in which product diversity emerges largely through downstream shaping processes.

Figure 1: Author-generated schematic of the plastic production chain from refinery feedstock to polymer pellets and downstream forming routes. Synthesized from^[6,7].

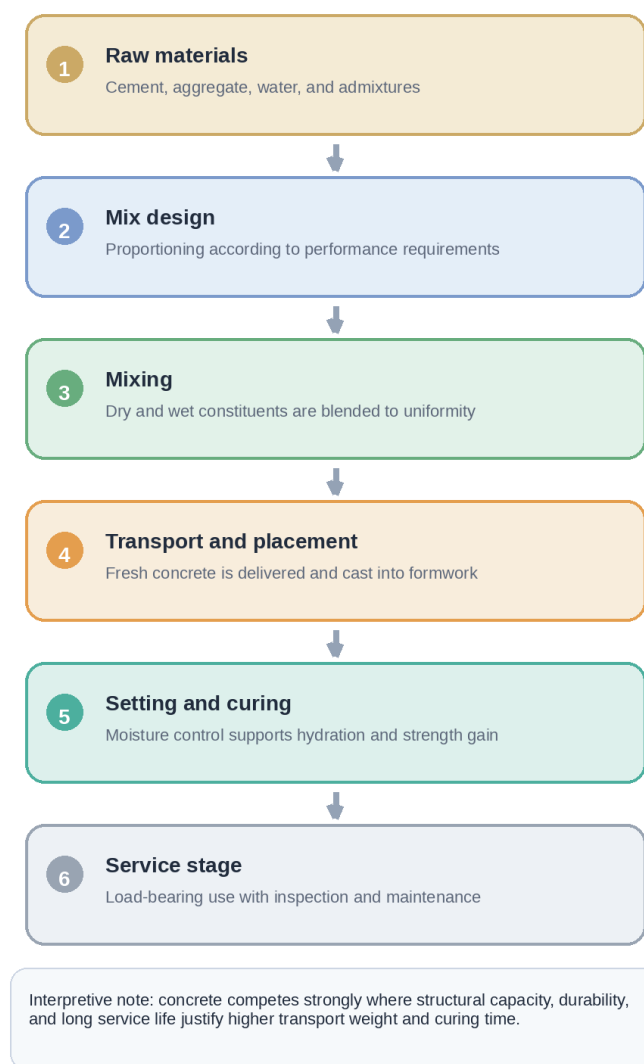


Concrete is a composite system composed primarily of cement, water, fine aggregate, coarse aggregate, and admixtures. Its commercial strength lies in rigidity, compressive capacity, fire resistance, and long-term service in load-bearing environments^[2,4,5]. Concrete is especially well suited to structural members, rigid pipes, heavy-duty pavers, and large-section infrastructure components^[2,4,5]. Unlike plastics, which typically enter the market as finished factory products, concrete can be cast in place or precast in large units, allowing wide dimensional flexibility but usually at higher transport weight and installation effort. By contrast, the concrete route shown in Figure 2 is organized around batching, mixing, placement, compaction, and curing, which ties performance more directly to site execution and structural service conditions.

Figure 2: Author-generated schematic of the concrete production and construction sequence, from material proportioning and mixing to placement and curing. Synthesized from^[2,4,5].

Concrete Production and Construction Workflow

Step-by-step route from raw materials to service stage



These differences in processing route help explain market boundaries. Plastic products are more likely to expand where the product is modular, repetitive, corrosion-sensitive, and installation-intensive. Concrete remains more competitive where high compressive demand, dimensional stability, abrasion resistance, or long-term heavy loading dominate product selection. The comparison therefore begins not only with material properties but also with the manufacturing logic embedded in each material system.

Table 1. Comparative profile of plastic and concrete construction products.

Dimension	Plastic products	Concrete products	Market implication
Density and handling	Low density; easy transport; simple manual handling in many cases	High self-weight; greater transport and lifting demand	Plastic gains advantage in installation-sensitive and distributed projects
Mechanical behavior	Good toughness and flexibility; lower stiffness and lower load-bearing capacity	High compressive strength and rigidity; suitable for heavy-duty service	Concrete remains preferred for structural and heavily loaded segments
Chemical and moisture resistance	Strong corrosion resistance in many service environments	Can deteriorate under aggressive chemical exposure if not properly designed	Plastic is attractive in sewage, drainage, and corrosive environments
Fire and thermal behavior	Generally weaker fire resistance and higher temperature sensitivity	Strong fire resistance and thermal mass	Concrete is favored where fire safety is a primary criterion
Manufacturing route	Extrusion/injection/blow molding support high-volume standardized production	Cast-in-place and precast routes support large-section products	Plastic fits modular mass products; concrete fits rigid large-scale systems
Maintenance and replacement	Often easier to replace and repair because of lower weight	Long service life in stable heavy-duty conditions but replacement can be labor-intensive	Life-cycle cost depends strongly on service environment and access conditions
Environmental issues	Recycling potential and recycled feedstock options, but waste leakage and end-of-life remain concerns	Durable and recyclable as aggregate, but cement production is carbon-intensive	Both materials require life-cycle assessment rather than single-indicator judgment

3.2 Cost structure and life-cycle implications

A meaningful material comparison in construction cannot rely on raw material price alone. Product competitiveness depends on a broader cost structure that includes manufacturing efficiency, transport, installation time, required equipment, maintenance frequency, service life, and replacement risk. In many lightweight product categories, plastics gain a clear economic advantage because low density reduces freight cost and simplifies storage, handling, and on-site assembly^[9,14–16]. This is particularly evident in pipes and modular pavers, where transportation and labor can account for a substantial part of delivered cost^[9,10,14,17].

Concrete, however, should not be treated as universally more expensive. Its raw constituents are abundant, its production system is mature, and its durability under compression and abrasion gives it strong life-cycle value in heavy-duty applications^[2–5]. For large-diameter drainage systems, high-load pavements, and structural settings, concrete frequently retains an advantage where dimensional stability and load-bearing reliability under demanding service conditions outweigh the initial burden of transport and installation^[2,4,9].

The cost comparison therefore leads to a scenario-dependent conclusion. Plastics are often cheaper in total installed cost when handling, labor, corrosion protection, and rapid construction matter more than extreme structural capacity. Concrete remains economically rational where the dominant cost risk comes from failure under load, abrasion, fire, or long-term deformation. In practice, market substitution usually occurs where plastics lower system cost without creating unacceptable performance trade-offs.

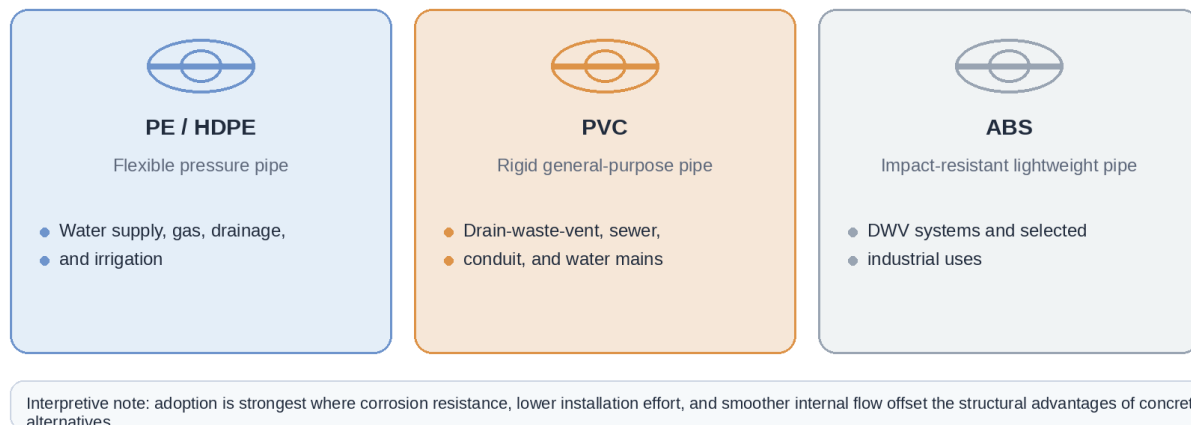
3.3 Application scenarios and functional fit

Drainage pipes provide one of the clearest examples of differentiated substitution. Plastic pipe systems are widely used because of their low weight, corrosion resistance, smooth internal surfaces, and ease of installation^[9,14–16]. They are especially competitive in building drainage, stormwater conveyance, and buried pipelines where joint handling and transport efficiency are important. Concrete pipe, by contrast, remains relevant in some large-diameter and high-load gravity applications, where rigidity and external load performance are central selection criteria^[9,14]. This pattern suggests not blanket replacement, but segmentation within the same product category. Figure 3 illustrates the diversity of plastic pipe families that support this shift, especially in buried and corrosion-sensitive drainage uses.

Figure 3: Author-generated comparison schematic of representative plastic pipe families used in construction applications, including PE/HDPE, PVC, and ABS. Synthesized from ^[9,14–16].

Representative Plastic Pipe Families in Construction

Pipe materials differ in flexibility, rigidity, and target application

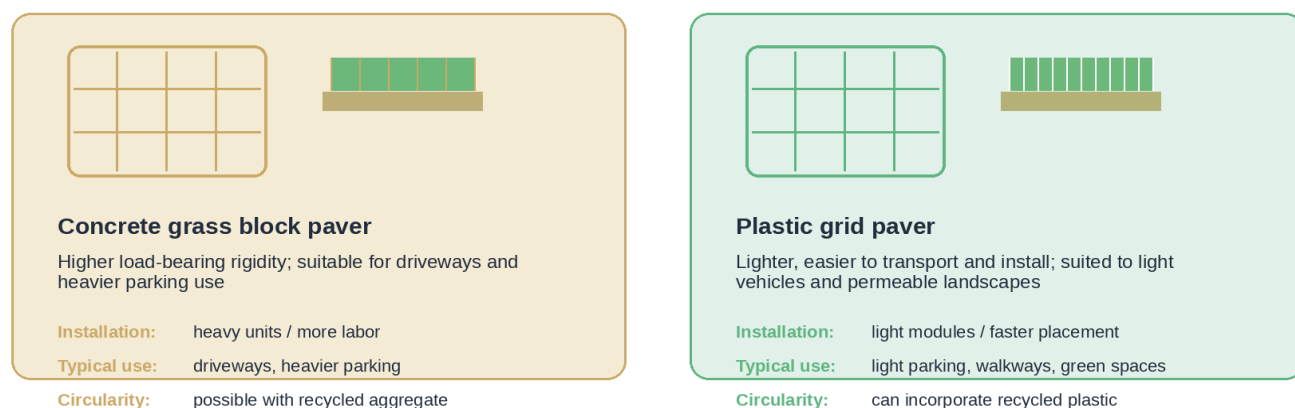


Grass block pavers show a similar but more visibly market-oriented division. Plastic grass pavers are generally favored in light-vehicle parking, pedestrian areas, and stormwater-sensitive landscapes because their lower weight and modular grid configuration can simplify transport and installation^[10,17–19]. Concrete grass pavers, by contrast, are more suitable where higher load resistance and surface robustness are required^[10,18,19]. As a result, the market tends to allocate plastic pavers to lighter and more environmentally themed applications, while concrete pavers retain a stronger position in high-load or high-wear surfaces. This contrast is visualized in Figure 4, where the heavier, rigid concrete format differs clearly from the lightweight grid-type plastic system.

Figure 4: Author-generated comparative schematic of concrete grass block pavers and plastic grid pavers, emphasizing differences in installation logic, load-bearing behavior, and typical use conditions. Synthesized from ^[10,17–19].

Concrete vs. Plastic Grass Block Paver Systems

Conceptual comparison of structure, installation logic, and use intensity



Road applications illustrate an even more important point: innovation often proceeds through hybridization rather than one-to-one replacement. The PlasticRoad concept uses prefabricated modular plastic structures to integrate drainage, cable management, and lightweight installation, thereby expanding the functional value of pavement systems^[12]. At the same time, plastic-modified asphalt and related road-material strategies use waste plastic as an additive or modifier rather than as a full replacement for mineral materials^[11,21–26]. Similarly, recycled plastic particles introduced into concrete can improve selected properties or support lightweight and circular-design objectives, even though such mixes still face performance limitations and require careful interface engineering^[8,27–29]. The modular system logic of PlasticRoad is shown in Figure 5.

Figure 5: Author-generated schematic section of the PlasticRoad modular concept, highlighting the hollow cavity, modular shell, and utility/drainage functions. Synthesized from ^[12].

PlasticRoad Modular Structure

Schematic section of a hollow, climate-adaptive pavement element

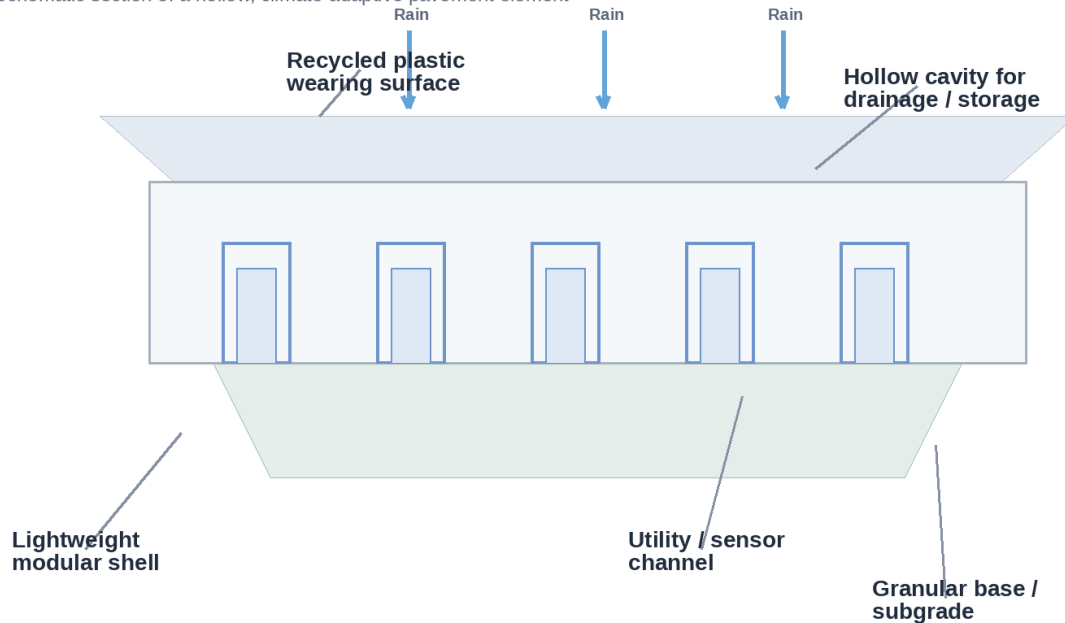


Table 2. Scenario-based comparison of representative construction products.

Product segment	Plastic-dominant conditions	Concrete-dominant conditions	Likely market pattern
Drainage pipes	Corrosion-sensitive environments; small to medium diameter; speed of installation; low handling burden	Large diameter; high external loads; rigid gravity systems; severe structural demand	Segmented coexistence with increasing plastic share in many non-structural applications
Grass block pavers	Light vehicles; pedestrian areas; high infiltration demand; low-cost modular layouts	Repeated heavy loading; higher abrasion; stronger surface stability requirement	Parallel use depending on traffic level and landscape function
Road modules	Need for modularity, integrated drainage, or utility corridors	Conventional heavy-duty pavement with proven long-term performance	Emerging niche for plastic systems; conventional dominance remains
Plastic-modified concrete/asphalt	Waste valorization; targeted performance improvement; circular branding	Where no performance or supply-chain advantage is demonstrated	Hybrid growth pathway rather than direct substitution

3.4 Sustainability trade-offs

Sustainability is one of the main reasons that the plastic–concrete comparison can no longer be reduced to traditional engineering criteria. Plastics contribute to lightweight logistics, corrosion resistance, and long service life in many non-structural applications^[6,7,9,16]. Their production can also incorporate recycled feedstocks, and products such as recycled plastic pavers, as well as modular pavement systems such as PlasticRoad, demonstrate how waste plastics can be redirected into longer-lived construction uses^[1,11,12,30,31]. However, these benefits are accompanied by persistent concerns over fossil-derived raw materials, end-of-life management, low recycling rates in some regions, and environmental leakage, including microplastic generation^[1,13,30–32].

Concrete, on the other hand, benefits from long service life and broad technical familiarity, but its sustainability burden is strongly affected by cement production, which is associated with substantial carbon emissions and resource demand^[2–5,33,34]. The extraction and transport of aggregates also impose environmental costs. These drawbacks have stimulated growing

interest in recycled aggregate concrete, permeable concrete, and the incorporation of recycled plastic particles into cementitious systems^[8,10,17–19,27–29]. Such approaches do not eliminate the environmental burden of concrete, but they help shift the material system toward resource recovery and functional efficiency.

The practical implication is that sustainability should be evaluated on a life-cycle basis rather than through isolated indicators. A lighter plastic product may reduce transport and installation impacts but still perform poorly if disposal is unmanaged. A concrete product may have a high embodied carbon burden but remain environmentally competitive if it delivers long service life in a high-load application with low replacement frequency. For both materials, the most promising path lies in circular design, improved recovery systems, and closer matching between material properties and actual service demand.

3.5 Market substitution pathways

The evidence reviewed above suggests that market change occurs through several distinct pathways rather than through a single replacement logic. The first pathway is direct substitution. This occurs when plastic products satisfy the required function at lower total cost and with fewer service risks, as in many low- to medium-load drainage and modular landscape applications^[9,10,14–17]. Direct substitution is most likely when weight, corrosion resistance, and installation speed are more important than high compressive capacity.

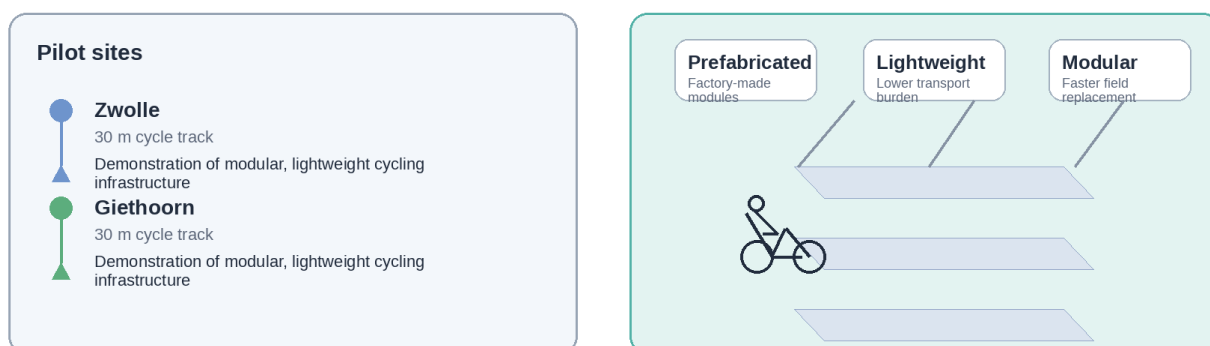
The second pathway is selective coexistence. In this pattern, plastic and concrete remain in the same product family but serve different performance bands. Drainage pipes and grass block pavers are typical examples: plastics expand in lighter-duty, easier-to-install, and sustainability-oriented segments, while concrete retains the heavy-duty and high-rigidity market^[9,10,14–19]. Here, substitution does not eliminate the incumbent material; instead, it redistributes market share according to service conditions.

The third pathway is hybrid integration. PlasticRoad systems, plastic-modified asphalt, and concrete containing recycled plastic particles show that innovation increasingly comes from combining the advantages of both material families^[8,11,12,21–29]. Hybridization is particularly relevant in markets shaped by circular-economy policy, where the commercial value of a product includes not only mechanical performance but also waste reduction, modularity, and environmental branding. Figure 6 shows how this concept has been translated into pilot installation.

Figure 6. Author-generated project summary schematic of the PlasticRoad pilot implementation in the Netherlands, based on publicly reported pilot information and the source synthesis. Synthesized from^[12].

PlasticRoad Pilot Implementation in the Netherlands

Illustrative project summary derived from publicly reported pilot information



The fourth pathway is circular upgrading. In this case, plastic waste or recycled concrete is not simply treated as disposal material but as a feedstock for new products, such as recycled plastic pavers, recycled aggregate paving systems, plastic-modified asphalt, and cementitious materials containing recycled polymer particles^[8,11,21–29]. From a management perspective, this pathway is important because it links material selection with supply-chain design, municipal waste systems, and sustainability reporting^[11,30,31]. It also broadens the competitive basis of construction products beyond price alone. A related pathway is shown in Figure 7, which summarizes the conversion of post-consumer plastic into road-material feedstock.

Figure 7. Author-generated process schematic for incorporating recycled plastic into road materials, from waste collection

and sorting to blending and paving. Synthesized from ^[11,20–26].

Recycled Plastic Road Material Process Chain

Conceptual route for incorporating waste plastic into asphalt-based systems



Taken together, these pathways show that the future competition between plastics and concrete will be determined less by abstract claims of superiority and more by the alignment among product function, life-cycle cost, environmental regulation, and circular-material infrastructure. The most realistic market outlook is therefore a differentiated one: direct replacement in some segments, stable coexistence in others, and hybrid expansion where regulatory and commercial incentives reward multifunctionality and recycled content.

3.6 Implications for material selection and market strategy

For developers, contractors, and public procurement agencies, the decisive question is not which material is universally superior, but which product configuration delivers the best fit under specific service conditions. Plastic products tend to create value where logistics, corrosion exposure, modular installation, and distributed maintenance dominate total system cost^[9,10,14–17]. Concrete remains preferable where stiffness, fire performance, mass, and long-term confidence under heavy loading carry greater economic weight^[2,4,5,9]. In practice, this means that substitution decisions should be made at the product level and supported by life-cycle costing rather than by headline material price alone^[9,10,14,15,17].

A second implication concerns innovation strategy. Firms are more likely to succeed through hybrid and circular propositions than through blanket claims that plastic can replace concrete or vice versa. Products that combine recycled feedstocks, drainage functionality, modular construction, or maintenance savings align more closely with current infrastructure and sustainability pressures. For policy-makers, the enabling conditions are standards, waste-sorting infrastructure, and reporting frameworks that allow recycled-content products to compete on credible performance terms^[8,11,12,20–29].

Table 3. Proposed market substitution pathways for plastic and concrete products in construction.

Pathway	Description	Representative examples	Key decision variables
Direct substitution	Plastic replaces concrete where required performance can be met with lower system cost	Lightweight drainage pipes; plastic landscape pavers	Weight, corrosion resistance, labor intensity, installation speed
Selective coexistence	Both materials remain in use but occupy different performance bands	Plastic vs. concrete grass pavers; pipe size and load segmentation	Traffic load, stiffness demand, abrasion, service environment
Hybrid integration	Products combine plastic and mineral systems or use plastic as a modifier	Plastic-modified roads; concrete with recycled plastic particles	Performance gain, compatibility, supply-chain availability, standards
Circular upgrading	Waste streams become input materials for new products	Recycled plastic pavers; recycled aggregate permeable grass bricks	Waste collection system, policy incentives, environmental reporting

4. Conclusions and Implications

This review develops a focused comparative account of plastic and concrete construction products by centering the analysis on cost structure, application fit, sustainability, and market substitution pathways. The comparison shows that the two materials are not interchangeable in a general sense; rather, each is commercially strongest under a specific set of service conditions^[2,9–11].

Plastics are most competitive in product categories that reward low density, corrosion resistance, modular production, and rapid installation^[6,7,9,13,16]. Concrete remains dominant where rigidity, mass, fire resistance, and high compressive reliability are central to performance^[2,4,5].

The review also indicates that the most promising future direction lies in hybrid and circular solutions. Plastic-modified road materials, recycled plastic pavers, and concrete systems that incorporate recycled plastics or recycled aggregates extend the comparison beyond simple material rivalry and toward integrated resource management^[8,11,21–29]. For researchers and industry decision-makers, the key implication is that material selection should be based on service-fit and total system value rather than on nominal unit price alone^[9,10,14,15,17].

Future work would benefit from standardized life-cycle cost data, region-specific carbon accounting, and more long-term field evidence for emerging products such as modular plastic road systems and recycled-plastic concrete. Such evidence would strengthen the evaluation of substitution not only as a technical decision, but also as a strategic response to environmental policy and changing construction-market demand.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] OECD. (2022). Global Plastics Outlook: Economic Drivers, Environmental Impacts and Policy Options. OECD Publishing. <https://doi.org/10.1787/de747aef-en>
- [2] Habert, G., Miller, S. A., John, V. M., Provis, J. L., Favier, A., Horvath, A., & Scrivener, K. L. (2020). Environmental impacts and decarbonization strategies in the cement and concrete industries. *Nature Reviews Earth & Environment*, 1(11), 559–573. <https://doi.org/10.1038/s43017-020-0093-3>
- [3] Belaïd, F. (2022). How does concrete and cement industry transformation contribute to mitigating climate change challenges? *Resources, Conservation & Recycling Advances*, 15, 200084. <https://doi.org/10.1016/j.rcradv.2022.200084>
- [4] Scrivener, K. L., John, V. M., & Gartner, E. M. (2018). Eco-efficient cements: Potential economically viable solutions

- for a low-CO₂ cement-based materials industry. *Cement and Concrete Research*, 114, 2–26. <https://doi.org/10.1016/j.cemconres.2018.03.015>
- [5] International Energy Agency, & Cement Sustainability Initiative. (2018). *Technology Roadmap: Low-Carbon Transition in the Cement Industry*. International Energy Agency. <https://www.iea.org/reports/technology-roadmap-low-carbon-transition-in-the-cement-industry>
- [6] Andrady, A. L., & Neal, M. A. (2009). Applications and societal benefits of plastics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1526), 1977–1984. <https://doi.org/10.1098/rstb.2008.0304>
- [7] PlasticsEurope. (2024). *Plastics – the Fast Facts 2024*. PlasticsEurope. <https://plasticseurope.org/knowledge-hub/plastics-the-fast-facts-2024/>
- [8] Siddique, R., Khatib, J., & Kaur, I. (2008). Use of recycled plastic in concrete: A review. *Waste Management*, 28(10), 1835–1852. <https://doi.org/10.1016/j.wasman.2007.09.011>
- [9] Vahidi, E., Jin, E., Das, M., Singh, P., & Eckelman, M. J. (2016). Environmental life cycle analysis of pipe materials for sewer systems. *Sustainable Cities and Society*, 27, 167–174. <https://doi.org/10.1016/j.scs.2016.06.028>
- [10] Scholz, M., & Grabowiecki, P. (2007). Review of permeable pavement systems. *Building and Environment*, 42(11), 3830–3836. <https://doi.org/10.1016/j.buildenv.2006.11.016>
- [11] Appiah, J. K., Berko-Boateng, V. N., & Tagbor, T. A. (2017). Use of waste plastic materials for road construction in Ghana. *Case Studies in Construction Materials*, 6, 1–7. <https://doi.org/10.1016/j.cscm.2016.11.001>
- [12] VolkerWessels. (2019). *Sustainability Report 2018*. VolkerWessels. https://www.volkerwessels.com/sites/vw_corporate/files/2024-04/VW%20MVO2018%20ENG.pdf
- [13] Geyer, R., Jambeck, J. R., & Law, K. L. (2017). Production, use, and fate of all plastics ever made. *Science Advances*, 3(7), e1700782. <https://doi.org/10.1126/sciadv.1700782>
- [14] Uni-Bell PVC Pipe Association. (2017). *Life Cycle Assessment of PVC Water and Sewer Pipe and Comparative Sustainability Analysis of Pipe Materials*. Uni-Bell PVC Pipe Association. https://www.uni-bell.org/files/Reports/Life_Cycle_Assessment_of_PVC_Water_and_Sewer_Pipe_and_Comparative_Sustainability_Analysis_of_Pipe_Materials.pdf
- [15] Plastics Pipe Institute. (2020). *Life Cycle Assessment of North American Stormwater Pipe*. Plastics Pipe Institute. <https://plasticpipe.org/common/Uploaded%20files/Technical/PPI-TR-53.pdf>
- [16] TEPPFA. (2024). *100 Years Lifetime of Plastic Pipes: Meta-Study*. TEPPFA. <https://www.teppfa.eu/wp-content/uploads/2024-04-25-Meta-study-100-years-of-lifetime-of-plastic-pipes.pdf>
- [17] Kuruppu, U., Rahman, A., & Rahman, M. A. (2019). Permeable pavement as a stormwater best management practice: A review and discussion. *Environmental Earth Sciences*, 78, 327. <https://doi.org/10.1007/s12665-019-8312-2>
- [18] Muttuvelu, D. V., & Kjems, E. (2021). A systematic review of permeable pavements and their unbound material properties in comparison to traditional subbase materials. *Infrastructures*, 6(12), 179. <https://doi.org/10.3390/infrastructures6120179>
- [19] U.S. Environmental Protection Agency. (2021). *Permeable Pavements: Stormwater Best Management Practice*. U.S. Environmental Protection Agency. <https://www.epa.gov/system/files/documents/2021-11/bmp-permeable-pavements.pdf>
- [20] Patil, P., Stankevich, N., Tsydenova, N., & Ilic, Z. D. (2023). *Plastic Waste in Road Construction: A Path Worth Paving? Application of Dry Process in South Asia*. World Bank. <https://documents1.worldbank.org/curated/en/099003105242322144/pdf/IDU0743eda4a0091c049000acd506d2f0b9ae123.pdf>
- [21] Xu, F., Zhao, Y., & Li, K. (2022). Using waste plastics as asphalt modifier: A review. *Materials*, 15(1), 110. <https://doi.org/10.3390/ma15010110>
- [22] Grady, B. P. (2021). Waste plastics in asphalt concrete: A review. *SPE Polymers*, 2(1), 4–18. <https://doi.org/10.1002/pls2.10034>
- [23] Ma, Y., Zhou, H., Jiang, X., Polaczyk, P., Xiao, R., Zhang, M., & Huang, B. (2021). The utilization of waste plastics in asphalt pavements: A review. *Cleaner Materials*, 2, 100031. <https://doi.org/10.1016/j.clema.2021.100031>

- [24] Duarte, G. M., & Faxina, A. (2021). Asphalt concrete mixtures modified with polymeric waste by the wet and dry processes: A literature review. *Construction and Building Materials*, 312, 125408. <https://doi.org/10.1016/j.conbuildmat.2021.125408>
- [25] Santos, J., Pham, A., Stasinopoulos, P., & Giustozzi, F. (2021). Recycling waste plastics in roads: A life-cycle assessment study using primary data. *Science of the Total Environment*, 751, 141842. <https://doi.org/10.1016/j.scitotenv.2020.141842>
- [26] Hao, G., He, M., Lim, S. M., Chong, D., Ren, H., & Ng, C. A. (2024). Recycling of plastic waste in porous asphalt pavement: Engineering, environmental, and economic implications. *Journal of Cleaner Production*, 440, 140865. <https://doi.org/10.1016/j.jclepro.2024.140865>
- [27] Saikia, N., & de Brito, J. (2012). Use of plastic waste as aggregate in cement mortar and concrete preparation: A review. *Construction and Building Materials*, 34, 385–401. <https://doi.org/10.1016/j.conbuildmat.2012.02.066>
- [28] Park, J., & Kim, M. O. (2020). Mechanical properties of cement-based materials with recycled plastic: A review. *Sustainability*, 12(21), 9060. <https://doi.org/10.3390/su12219060>
- [29] da Silva, T. R., de Azevedo, A. R. G., Cecchin, D., Marvila, M. T., Amran, M., Fediuk, R., ... Monteiro, S. N. (2021). Application of plastic wastes in construction materials: A review using the concept of life-cycle assessment in the context of recent research for future perspectives. *Materials*, 14(13), 3549. <https://doi.org/10.3390/ma14133549>
- [30] OECD. (2022). *Global Plastics Outlook: Policy Scenarios to 2060*. OECD Publishing. <https://doi.org/10.1787/aa1edf33-en>
- [31] European Environment Agency. (2024). *The Role of Plastics in Europe's Circular Economy*. European Environment Agency. <https://www.eea.europa.eu/en/analysis/publications/the-role-of-plastics-in-europes>
- [32] Hopewell, J., Dvorak, R., & Kosior, E. (2009). Plastics recycling: Challenges and opportunities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1526), 2115–2126. <https://doi.org/10.1098/rstb.2008.0311>
- [33] Miller, S. A., Horvath, A., & Monteiro, P. J. M. (2018). Impacts of booming concrete production on water resources worldwide. *Nature Sustainability*, 1(1), 69–76. <https://doi.org/10.1038/s41893-017-0009-5>
- [34] Watari, T., Cao, Z., Hata, S., & Nansai, K. (2022). Efficient use of cement and concrete to reduce reliance on supply-side technologies for net-zero emissions. *Nature Communications*, 13, 4158. <https://doi.org/10.1038/s41467-022-31806-2>

Technology for Preparing Nanocrystalline Diamond Coatings by Hot-Filament CVD Method

Zhihang Xue, Jianfeng Wang, Xing Hu*

Shanghai DianJi University, Shanghai, 201306, China

*Corresponding author: Xing Hu, hux@sdju.edu.cn

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: This paper systematically investigates the core technical system for preparing nanocrystalline diamond coatings (NDC) by hot-filament chemical vapor deposition (HFCVD), so as to provide theoretical and process references for the industrial development and cross-field applications of this technology. Nanocrystalline diamond coatings have a grain size of 5–100 nm, combining ultra-high hardness, exceptional surface finish, excellent thermal conductivity and strong chemical inertness. With the merits of low equipment cost, large deposition area and high controllability of process parameters, HFCVD has become the core method for the industrial preparation of NDC. This paper reviews the development history of HFCVD technology and NDC, elaborates on the core processes in preparation including substrate selection, gas source design, hot-filament matching and pretreatment, analyzes the working principle of the equipment, the preparation flow and the optimization mechanism of key process parameters, summarizes the application status of the coatings in multiple fields, clarifies the industrialization bottlenecks of the current technology, and proposes key research directions for future efforts.

Keywords: Hot-filament Chemical Vapor Deposition (HFCVD); Nanocrystalline Diamond Coating (NDC); Preparation Process; Superhard Coating

Published: Apr 24, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1306>

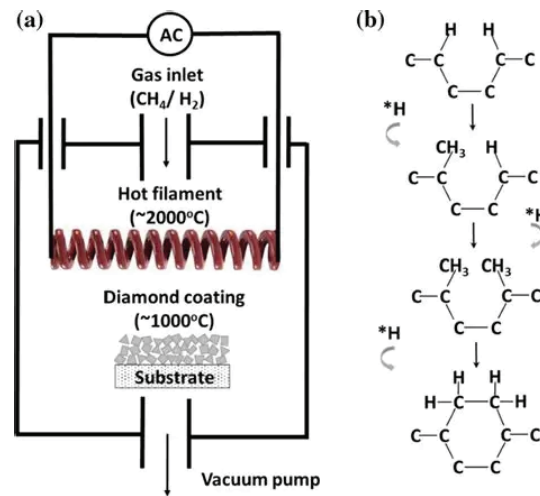
Introduction

With the rapid development of high-end manufacturing towards high precision, long service life and low energy consumption, superhard functional coating technology has become a key breakthrough for improving the service performance of core components. Among them, nanocrystalline diamond coating specifically refers to a diamond-phase thin film with a grain size of 5-100 nm and sp^3 -bonded carbon atoms as the main component. Compared with traditional micro-scale diamond coatings, it achieves both ultra-hard properties and high surface finish by virtue of the high-density grain boundaries formed by nano-grains, and also retains the intrinsic ultra-high thermal conductivity and chemical inertness of diamond materials. It has shown irreplaceable application value in fields such as precision machining, aerospace seals, biomedical implants, and high-end electronic devices.

Hot-Filament Chemical Vapor Deposition (HFCVD) is the mainstream industrial process for preparing nanocrystalline diamond coatings. Its core principle is to crack the mixed reaction gas of CH_4/H_2 through high-melting-point metal wires electrically heated to 2000-2200 °C, initiate gas-phase chemical reactions on the substrate surface several millimeters away from the hot filaments, and ultimately achieve the oriented deposition of carbon atoms in the diamond phase on the substrate.

Compared with other CVD technologies, HFCVD technology has the advantages of simple equipment structure, low investment cost, capability of batch deposition of large-size workpieces, and flexible regulation of process parameters, which can adapt to different substrate materials and coating performance requirements. It is currently the preferred technology for the industrial preparation of nanocrystalline diamond coatings.

Figure 1. Schematic Diagram of HFCVD



Currently, the industrialization process of nanocrystalline diamond coatings still faces multiple technical bottlenecks: on the one hand, HFCVD technology has inherent problems such as weak substrate adhesion, low deposition rate (usually $<1 \mu\text{m/h}$), and poor uniformity in large-area deposition; on the other hand, the standardized evaluation system for coating performance has not yet been improved, and the lack of cross-industry application standards restricts the transformation efficiency of technological achievements. Against this background, major industrial countries around the world have incorporated advanced coating technology into their new material development strategies. China's "14th Five-Year Plan" clearly lists "high-performance coating materials" as a key research field for cutting-edge new materials, and the European Union and the United States have also successively invested special resources to support the innovation of surface engineering technology. Policy orientation has promoted the research and development focus to shift from single process optimization to the whole-chain innovation of "material-equipment-application", and at the same time driven the evolution of HFCVD technology towards green manufacturing with low temperature, low pressure and low carbon.

This paper, by systematically sorting out the technical system for preparing nanocrystalline diamond coatings by HFCVD, integrates the core contents of material selection, process regulation, application scenarios and technical bottlenecks, aiming to provide a systematic reference for the breakthrough in basic research and the promotion of industrial application of this technology.

1. Core Materials and Pretreatment Technology for the Preparation of Nanocrystalline Diamond Coatings

The selection and pretreatment of preparation materials are the foundation for determining the final quality of nanocrystalline diamond coatings, which directly affect the nucleation efficiency, crystalline purity, interface adhesion and service performance of the coatings. It mainly includes three core components: substrate materials, gas source system and hot-filament materials, as well as the supporting substrate pretreatment process.

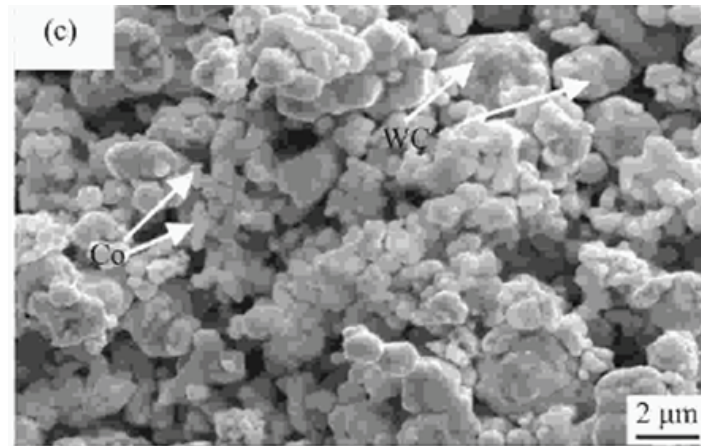
1.1 Selection and Compatibility of Substrate Materials

The substrate is not only a physical support for coating deposition, but its thermal expansion coefficient, crystal structure, and high-temperature chemical stability directly affect the nucleation and growth of the coating, the internal stress level, and the interface bonding effect. In the HFCVD process, the substrate must withstand a deposition temperature of 300-900 °C and form a stable bonding interface with the diamond material to avoid coating cracking and peeling caused by thermal mismatch. The characteristics and suitable scenarios of commonly used substrate material classifications are as follows:

Cemented Carbide: Centered on WC-Co cemented carbide, its hardness can reach HV 1500-1800, with excellent mechanical

strength, which can form performance synergy with diamond coatings, making it the preferred substrate material for cutting tools and wire drawing dies. Its core defect is that the surface Co binder phase will catalyze the graphitization of diamond at high temperatures, damaging the bonding between the coating and the substrate. Therefore, the surface Co phase must be removed through pretreatment. Among them, the WC-Co cemented carbide with a cobalt content ≤ 6 wt% can have a bonding strength with the coating increased to more than 50 MPa after pretreatment.

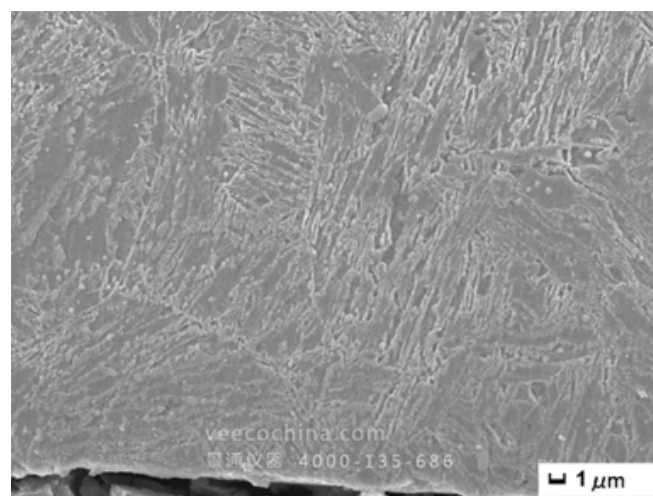
Figure 2. Typical SEM Image of WC-Co Powder



Silicon Materials: Including monocrystalline silicon and polycrystalline silicon, they have a regular lattice structure, a diamond nucleation density of up to 10^9 cm⁻², good compatibility with microelectronic processes, and a thermal expansion coefficient slightly different from that of diamond, which can reduce thermal stress cracks during the deposition process. They are the core substrate materials for coatings of microelectronic devices and micro sensors. Their disadvantages are low mechanical strength and high brittleness, which make them unable to withstand heavy-load working conditions.

Metal Materials: Including titanium alloys, stainless steel, molybdenum, etc. Titanium alloys have strong corrosion resistance, stainless steel has low cost, and molybdenum has a high melting point, which can adapt to different working condition requirements. The core problem of this type of material is that their thermal expansion coefficient is greatly different from that of diamond, which is prone to generate huge internal stress after deposition, leading to coating peeling. It is necessary to prepare a transition layer to alleviate the problem of thermal mismatch, and they are mainly used in the fields of aerospace structural components, corrosion-resistant components, and medical surgical instruments.

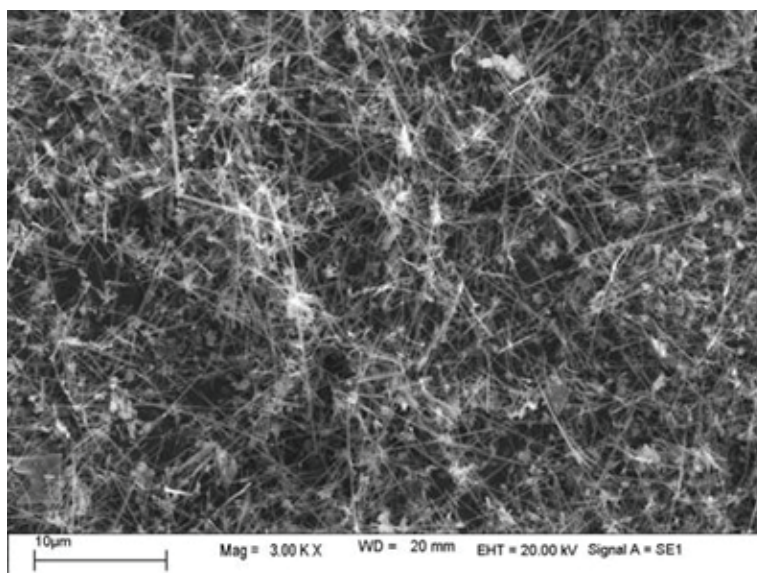
Figure 3. Typical SEM Image of Martensitic Stainless Steel



Ceramic materials: Including SiC, Al₂O₃, Si₃N₄, etc. They have excellent high-temperature resistance. The long-term service temperature of SiC can reach 1600°C, with strong chemical stability, and the thermal expansion coefficient is closer to that of diamond. The internal stress level of the coating is low, which can effectively reduce the risk of coating cracking. The disadvantages of this type of material are high brittleness and high difficulty in processing complex shapes. They are mainly

used in extreme working conditions such as high-temperature wear-resistant components and insulating devices.

Figure 4. Typical SEM Image of SiC Powder



1.2 Design of Gas Source System

The gas source system of the HFCVD process is the material foundation for the growth of diamond coatings, which directly determines the carbon source supply, phase purity and grain structure of the coatings. It is mainly divided into three categories: carbon source gas, hydrogen and doping gas. The functions and ratio design of various gases are as follows:

Carbon source gases: Commonly used gases include methane (CH_4), acetylene (C_2H_2), acetone ($\text{C}_3\text{H}_6\text{O}$), etc. Methane is the most commonly used carbon source, with a simple molecular structure and stable reaction. It decomposes under the high temperature of the hot wire to generate carbon-containing active groups, which is suitable for the controllable growth of nanodiamond coatings. Acetylene has a high carbon content and can achieve a higher deposition rate, but the reaction process is difficult to control; acetone is more suitable for the uniform deposition on substrates with complex shapes.

Hydrogen (H_2): Hydrogen is indispensable in the process of preparing nanodiamond coatings by hot filament CVD. On the one hand, it is used to dilute the carbon source gas, regulate the carbon concentration of the reaction system, and prevent the excessive formation of non-diamond phases; on the other hand, it decomposes at high temperatures to generate hydrogen atoms, which preferentially etch sp^2 -hybridized non-diamond carbon, promote the dominant growth of sp^3 -hybridized diamond phases, and ensure the phase purity of the coating.

Doping Gases: Commonly used ones include nitrogen (N_2) and borane (B_2H_6). Doping gases can specifically regulate the grain size, mechanical properties and electrical properties of the coating. For example, the N atoms formed by the decomposition of nitrogen can accelerate the hydrogen extraction step in the reaction process, increase the secondary nucleation rate, refine the diamond grains, and improve the hardness of the coating; borane doping can realize the p-type semiconductor modification of the diamond coating, expanding its application in the field of electronic devices.

Table 1. Gas Sources and Reaction Gases for HFCVD

Gas Types	Commonly Used Gases	Functions and Roles	Typical Ratios
Carbon Source Gases	Methane (CH_4) ^[8, 10] , Acetylene (C_2H_2), Acetone ($\text{C}_3\text{H}_6\text{O}$)	Provide carbon atoms required for deposition	1-5 vol%
Carrier Gas / Diluent Gas	Hydrogen (H_2) ^[6-9] , Argon (Ar)	Dilute the carbon source and promote the formation of sp^3 bonds	95-99 vol%
Doping Gases	Nitrogen (N_2) ^[7] , Borane (B_2H_6)	Regulate electrical properties (n-type / p-type)	100-5000 ppm

1.3 Hot Filament Material Selection

The hot filament is the core component of an HFCVD system and serves as the energy source for the cracking of reaction gases. Its performance directly determines the process stability and coating quality. It must meet four core requirements: high melting point, good high-temperature mechanical strength, low vapor pressure, resistance to carburization, and stable resistance-temperature characteristics.

In HFCVD processes, the hot filament must be heated to 1800–2400 °C to achieve efficient cracking of hydrogen and carbon-source gases. For this reason, the hot filament material must have a melting point above 3000 °C to avoid melting or deformation at high temperatures. Meanwhile, it should maintain good high-temperature mechanical strength to prevent bending and sagging of the suspended filament under high heat, which would cause uneven filament–substrate spacing and lead to variations in coating thickness and quality. Low vapor pressure and carburization resistance reduce coating contamination caused by high-temperature evaporation of the filament, avoid corrosion of the filament by carbides, and extend its service life. Stable resistance-temperature characteristics enable precise control of filament temperature, ensuring consistency in batch processing.

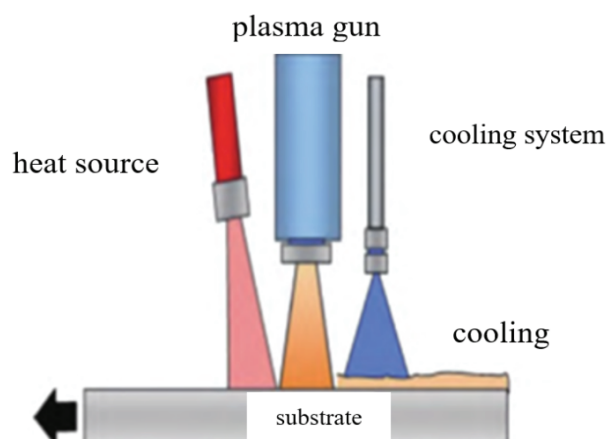
Commonly used hot filament materials at present include tantalum (Ta), tungsten (W), rhenium (Re) and their alloys. The filament diameter is usually 0.1–0.5 mm, mostly arranged in a parallel array with a spacing of 5–15 mm. Among them, tungsten filaments^[7, 8] are the most widely used. With a melting point of 3422 °C, they can operate stably for a long time at 2000–2300 °C, and feature good linearity in resistance-temperature characteristics and high temperature control accuracy. Tantalum filaments exhibit excellent high-temperature strength and corrosion resistance, making them suitable for long-period continuous deposition processes. Rhenium alloys have the strongest carburization resistance, with a service life significantly better than that of tungsten and tantalum, but at a higher cost.

1.4 Key Processes of Substrate Pretreatment

Substrate pretreatment is a key step to improve coating-substrate adhesion and diamond nucleation density. It mainly includes three core processes: surface roughening, chemical etching and interlayer design, together with supporting pre-deposition activation treatments such as hydrogen baking and filament baking.

Surface roughening: Its core function is to form a mechanical anchoring effect by roughening the substrate surface, increase the actual contact area between the coating and the substrate, and improve the interfacial bonding strength. The standard process adopts sandblasting: using 20–50 μm Al₂O₃ abrasives as the medium, the substrate surface is blasted at an incident angle of 60° under a pressure of 0.5 MPa to obtain an optimal roughness of Ra=0.8–1.2 μm, which can increase the coating bonding strength by 40%–60%. Excessive roughening will cause stress concentration in the coating and lead to early spalling, so the upper limit of roughness must be strictly controlled.

Figure 5. Schematic diagram of synchronous laser thermal spraying^[3]

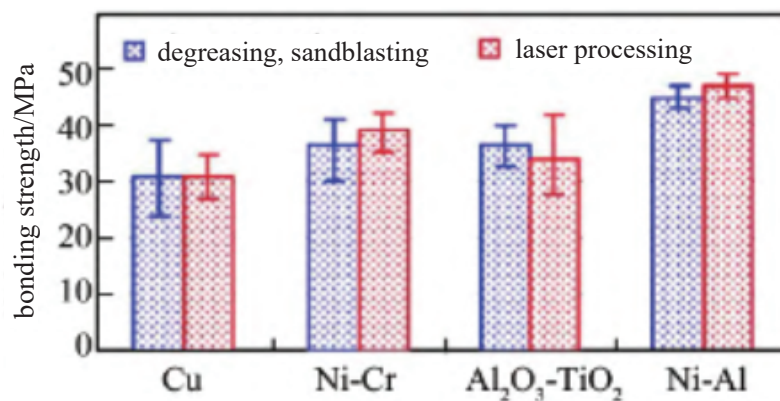


Researchers integrated laser cleaning equipment with thermal spraying equipment (including plasma gun, heat source and cooling system) to develop a synchronous laser thermal spraying process, which realizes simultaneous laser cleaning of the substrate and coating deposition during thermal spraying, as shown in Figure 5. The plasma gun is the key component, which

sprays molten or semi-molten particles such as Ni-Al and Al_2O_3 onto the substrate surface under compressed gas flow. The heat source supplies energy to the plasma gun to ensure the sprayed materials reach the required state for film formation. The cooling system cools the equipment and controls the substrate temperature ($\leq 200^\circ\text{C}$) to avoid high-temperature damage and ensure mechanical bonding between the coating and the substrate. Synchronous laser cleaning removes the micron-scale oxide layer on the substrate to provide a clean deposition interface, and continuous cooling ensures stable operation of the equipment and substrate. This process breaks through the limitation of separate traditional pretreatment and spraying, and provides a technical solution for improving coating bonding performance, extending service life and optimizing production efficiency.

L Qin, L Ping et al. ^[4] compared the effects of “degreasing + sandblasting” and laser pretreatment on the bonding strength of Cu, Ni-Cr, Al_2O_3 -TiO₂ and Ni-Al coatings. The results in Figure 6 show that after laser cleaning, the bonding strength of Cu, Ni-Cr and Ni-Al coatings is equivalent to or higher than that of the sandblasted group (the interfacial toughness of Ni-Al coating increases from $0.39 \text{ MPa}\cdot\text{m}^{1/2}$ to $2 \text{ MPa}\cdot\text{m}^{1/2}$). Only the high-melting-point Al_2O_3 -TiO₂ coating exhibits increased interfacial porosity and lower bonding strength after laser cleaning, due to the semi-molten state of most particles and the smooth surface ($R_a < 1\mu\text{m}$). The results quantitatively compare different pretreatment methods and verify the suitability of laser cleaning for Cu and Ni-based coatings, but do not analyze the regulation of laser parameters. It provides experimental support for the selection of pretreatment methods, and suggests that high-melting-point ceramic coatings need laser etching to optimize surface roughness for improved bonding strength.

Figure 6. Effect of laser cleaning on coating bonding strength^[4]



Chemical etching: It is mainly applied to WCCo cemented carbide substrates to remove the surface Co binder phase and eliminate its catalytic graphitization interference on diamond growth. The mainstream process is the “alcoholalkali twostep method” ^[5]. The substrate is first ultrasonically cleaned in methanol solution for 20 minutes to remove surface oil contamination, followed by ultrasonic etching in a $\text{KOHK}_3\text{Fe}(\text{CN})_6\text{H}_2\text{O}$ mixture. This reduces the surface Co content from 6% to 0.4%–0.83%, increases surface roughness from 94.5 nm to 366 nm, and achieves a diamond nucleation density up to 10^9 cm^{-2} — far higher than the conventional acid pickling and grinding process. It is also suitable for pretreatment of complexshaped substrates such as drills and dies.

Figure 7. Comparison of surface morphologies of the substrate before and after two-step treatment

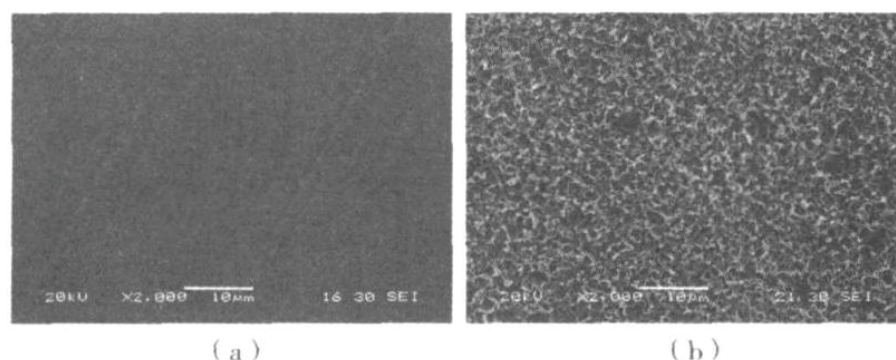
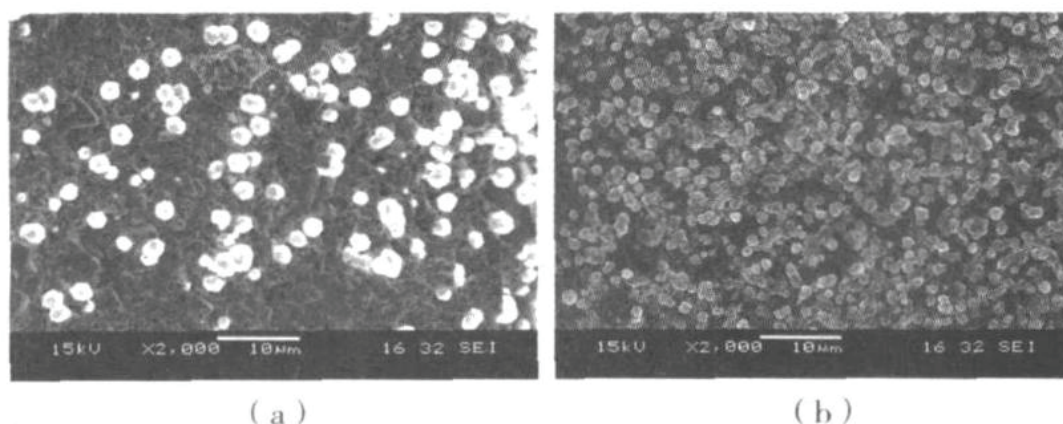


Figure 8. Nucleation density on the substrate surface under different treatment conditions



Interlayer design: Its core function is to alleviate the thermal expansion coefficient mismatch between diamond and the substrate, eliminate interfacial internal stress, and prevent high-temperature diffusion of elements such as Co from the substrate, ensuring high-quality growth of the diamond coating. The mainstream scheme adopts a Ti/TiC/TiN graded interlayer system, following four principles: gradual matching of thermal expansion coefficients, strong interfacial adhesion, stable chemical properties, and suppression of element diffusion. The Ti layer (0.3–0.5 μm) acts as an activation layer to enhance surface activity of the substrate; the TiC layer (1–2 μm) serves as a thermal matching layer with a thermal expansion coefficient between that of the substrate and diamond, smoothly transferring interfacial stress; the TiN layer (0.5–1 μm) functions as a diffusion barrier layer, which can react with Co to form a stable Co_3TiN phase and reduce the Co diffusion depth from 500 nm to below 80 nm. This gradient system reduces the interfacial residual stress from 8 GPa to 3.2 GPa, significantly improving the spalling resistance of the coating.

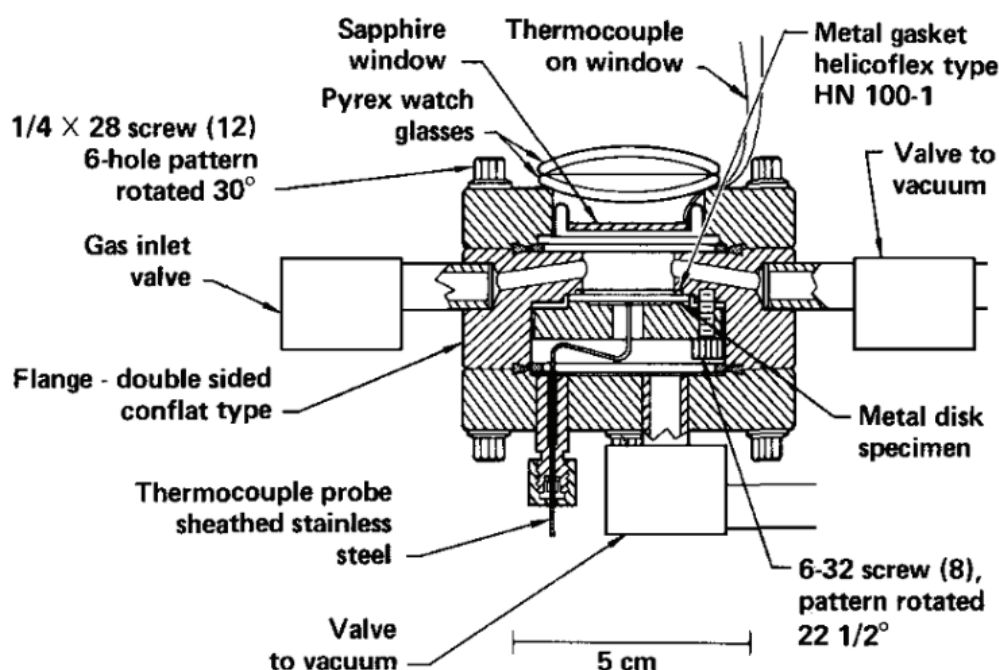
Pre-deposition activation treatment: includes filament carburization treatment (“filament baking”) and substrate hydrogen treatment (“hydrogen baking”). “Filament baking” involves heating tungsten filaments to 2300 $^{\circ}\text{C}$ and holding for 0.5 h to complete carburization and ensure stable working performance of the filaments. “Hydrogen baking” uses bombardment by high-temperature hydrogen atoms and ions on the substrate surface in a hydrogen atmosphere to remove surface impurities and oxide layers, providing a clean and active surface for diamond nucleation and further increasing nucleation density.

2 Process Principle and Parameter Control of Hot Filament CVD Preparation

2.1 Technical Principle and Equipment Structure

The basic principle of HFCVD technology is to heat a high-melting-point metal hot filament to 1500–2300 $^{\circ}\text{C}$ by electric current, so that the CH_4/H_2 mixed reaction gas introduced into the reaction chamber undergoes high-temperature pyrolysis on and near the filament surface, producing various active species such as atomic hydrogen and methyl radicals. Driven by concentration and thermal gradients, these active species migrate to the substrate surface maintained at 700–1000 $^{\circ}\text{C}$. Through a series of complex surface reactions including adsorption, desorption and surface diffusion, diamond nucleation and continuous growth are completed, eventually forming a continuous nanodiamond coating on the substrate surface. Throughout the process, the selective etching effect of atomic hydrogen is critical. It preferentially removes sp^2 -hybridized non-diamond carbon formed on the substrate surface, ensuring the dominant growth of the sp^3 -hybridized diamond phase, which is the key to obtaining high-purity diamond coatings.

Vacuum Reaction Chamber^[11]: It is the core space for coating deposition, usually made of 304L stainless steel. A high vacuum level above 10^{-5} Torr is required to isolate contamination from air impurities during coating growth. It is also equipped with a sapphire observation window and heating and temperature control components to realize real-time monitoring of the deposition process and precise temperature control, eliminating uneven coating thickness and composition caused by edge effects.

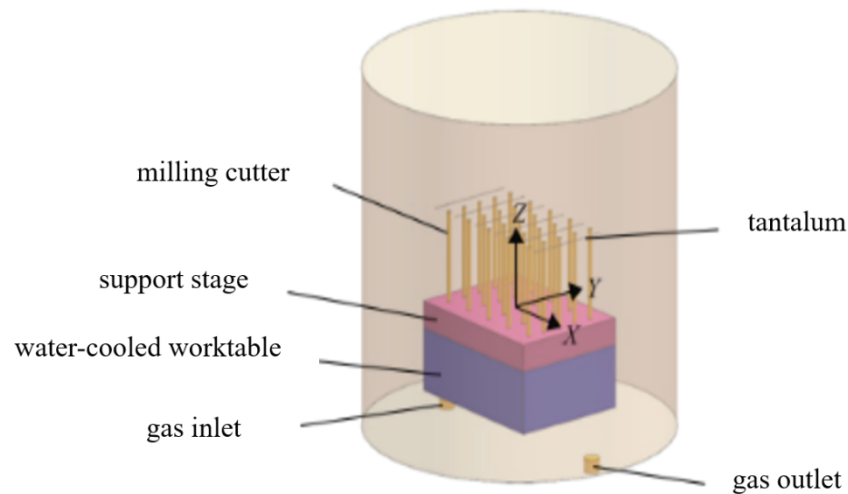
Figure 9. Cross-section of a cylindrical vacuum chamber for edge-effect-free studies of gas-metal reactions ^[11]

Gas Delivery System: It precisely controls the flow ratio of CH_4/H_2 , the core gases for diamond growth (a CH_4 concentration of 3% ensures sp^3 -bond-dominated growth), stabilizes the generation efficiency of carbon-active species, and avoids an increase in graphite phase caused by excessively high concentration or a decrease in nucleation density caused by excessively low concentration. Doping gases such as N_2/Ar allow precise tuning of grain size (e.g., N_2 doping achieves nanodiamond grains smaller than 30 nm). Through quantitative gas control, the system realizes precise regulation of coating composition and structure, ensures batch-to-batch consistency, and avoids performance drift caused by fluctuations in gas proportion.

Hot Filament Power Supply System: It maintains tungsten/tantalum filaments at a high temperature of 2000–2200 °C via stable electric current, serving as the core energy source for H_2 decomposition into H atoms for etching and purification and CH_4 decomposition into carbon-active groups for diamond growth in HFCVD. Current instability can cause temperature fluctuations of ± 50 °C, reducing the H_2 decomposition rate from 90% to 75%, resulting in abnormal carbon source behavior or increased impurity carbon. The system ensures continuous and efficient energy supply from the hot filament through steady-current temperature control, avoiding deposition interruption or degraded coating quality.

Substrate Heating System: It provides precise temperature control for different substrates (600–900 °C for cemented carbide, < 500 °C for polymers), avoiding substrate damage and optimizing coating adhesion. In the cemented carbide temperature range, it suppresses surface diffusion of Co, reduces graphitization interference, and improves interfacial bonding strength. In the low-temperature range for polymers, it preserves the structural integrity of the substrate. Through adaptive temperature control, the system protects the substrate while providing suitable surface energy for diamond nucleation, solving the compatibility problem between the substrate and the diamond coating.

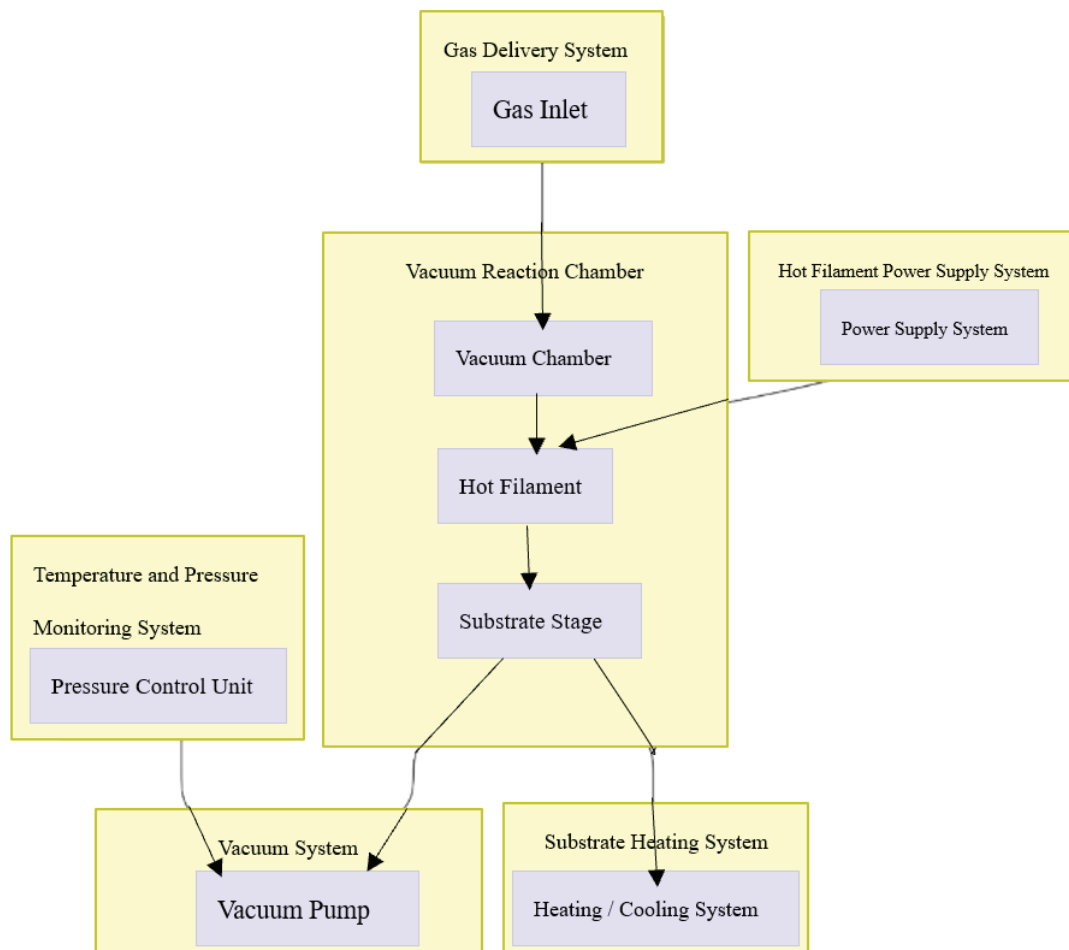
In Figure 10, the components of the HFCVD apparatus work synergistically for precise control of the tool temperature field: tantalum filaments at 2000–2200 °C produce radiant heat to drive the decomposition of CH_4/H_2 into carbon-active species; the milling cutter must be kept within the 600–900 °C process window, whereas conventional structures easily suffer from thermal accumulation and overheating above 950 °C, which induces graphite phase formation. After flow channel optimization of the water-cooled worktable, the temperature gradient is reduced from 25 °C/mm to 8 °C/mm. The coppermolybdenum composite support stage balances thermal conductivity and service life, achieving a threefold lifetime extension over pure copper. The 45° oblique gas inlet nozzles generate a swirling flow, raising the deposition rate by 12%. Through the “heat generation–heat transfer–heat dissipation” chain, all components realize precise temperature control and avoid the risk of thermal runaway.

Figure 10. Simulation model of the HFCVD system^[12]

Vacuum System: Composed of a mechanical pump, molecular pump and other components, it enables dynamic pressure control within 10^{-3} – 10^3 Pa. Before deposition, the chamber is pumped to high vacuum to remove residual impurity gases; during deposition, a stable reaction pressure is maintained to provide a suitable pressure environment for coating growth and ensure the steady diffusion and reaction of process gases.

Temperature and Pressure Monitoring System: It collects key process parameters such as hot filament temperature, substrate temperature and chamber pressure in real time. Automatic adjustment of parameters is realized through a closed-loop control system, eliminating parameter fluctuations caused by equipment aging, gas flow drift and other factors. This ensures the stability of each deposition batch and supports the industrial mass production of coatings.

Figure 11. Schematic Diagram of the Hot Filament CVD (HFCVD) Apparatus



2.2 Standard Fabrication Process Flow of HFCVD Diamond Coatings^[8, 10, 13]

Reaction Chamber Preparation and Evacuation: The cleaned AISI 316 stainless steel substrate is placed at the designated position in the quartz tube reaction chamber, and the coiled tungsten filament is fixed in place. A precise and constant filament-to-substrate distance of 3–10 mm is maintained to avoid uneven heating. The chamber is then evacuated with a vacuum pump to a high vacuum of 10^{-5} – 10^{-7} Torr to remove impurities such as oxygen and water vapor, preventing filament oxidation and non-diamond carbon formation, thus ensuring coating purity and interfacial adhesion strength.^[7]

Heating of Hot Filaments and Introduction of Process Gases: The hot filament power supply is activated to heat the filaments to the set temperature (typical value: 2300 °C).^[8] The reactive gases such as H₂ and CH₄ are decomposed by high-temperature radiation to produce carbon active species required for diamond nucleation. Meanwhile, deposition gases are introduced in proportion, commonly hydrogen, methane and nitrogen, with the flow ratio of methane to hydrogen being 1:30-100 and the flow ratio of nitrogen to methane being 1:0.1-3.^[1] Precise control of carbon source concentration can avoid graphitization or slow growth, and nitrogen doping can regulate grain size, providing core support for the regulation of coating structure and properties.

Precise control of the substrate temperature: is maintained within the suitable growth range of 700°C–1000°C for diamond coatings using a heating device. An appropriate temperature promotes the adsorption and diffusion of active species, avoids loose coatings at low temperatures, suppresses graphitization at high temperatures, and reduces thermal mismatch internal stress. Accurate temperature control is critical to the quality and growth rate of nanocrystalline diamond coatings.

Coating Growth Stage: Under suitable temperature and gas conditions, active species adsorb and react on the substrate surface to form nanodiamond nuclei and grow continuously. The nuclei gradually connect and merge into a continuous coating. The deposition time ranges from several hours to tens of hours, and the coating thickness can be adjusted on demand by controlling the time (e.g., 3–8 μm for cutting tool coatings, thinner for microelectronic coatings) to adapt to different application scenarios. A continuous, pore-free coating can improve wear resistance and protection, providing structural guarantee for its functional performance.

Cooling and Post-Treatment: After deposition, turn off the hot filament power supply and gas passages, and allow the reaction chamber to cool naturally to room temperature to avoid a sudden increase in thermal stress between the coating and the substrate caused by rapid cooling, prevent coating cracking and spalling, and ensure structural integrity. Targeted post-treatment can be performed according to application requirements^[2]: The coating on medical scalpels forms micro-nano structures through high-temperature oxidation to achieve super hydrophobicity and anti-blood adhesion; coatings for cutting tools and microelectronic devices are ground and polished to reduce surface roughness (e.g., Ra < 5 nm)^[11], to meet high-precision and flatness requirements and make coating properties more suitable for practical applications.

Table 2. Typical Process Steps of HFCVD

Step	Purpose	Parameter Range	Notes
Substrate Pretreatment	Enhance coating adhesion	Diamond grinding (0.5-1μm), acid etching (H ₂ SO ₄ /H ₂ O ₂)	Thoroughly clean to remove residues
Chamber Evacuation	Remove impurity gases	Base vacuum ≤10 ⁻³ Pa	Ensure no leaks
Filament Activation	Stabilize the filament	2000-2200°C, 10-20min	Avoid uneven initial carburization
Deposition Process	Grow nanocrystalline diamond	CH ₄ /H ₂ =1-5%, 1-5kPa, 600-900°C	Control filament-substrate distance at 3-10 mm
Cooling and Sampling	Prevent thermal stress cracking	Cool to <150°C under inert atmosphere	Slow cooling rate <5°C/min

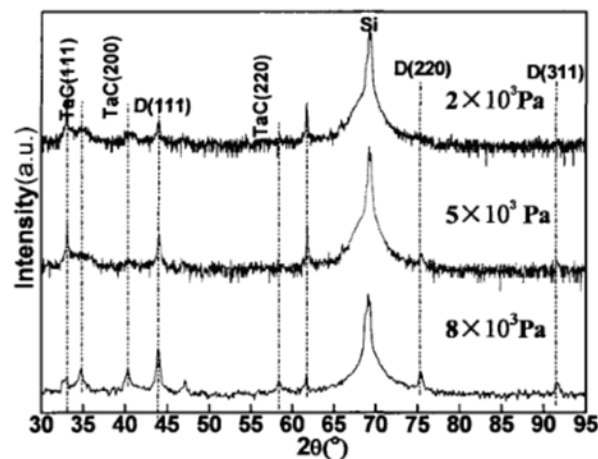
2.3 Core Process Parameter Optimization Mechanism

Effect of Carbon-to-Hydrogen Flow Ratio: It directly determines the carbon concentration in the reaction system, affecting the growth rate and quality of nanocrystalline diamond coatings. An excessively high carbon source proportion tends to

form non-diamond phases and degrade coating quality, while an excessively low proportion leads to a slow growth rate; For example, in “Nanocrystalline Diamond Coating and Its Preparation Method and Process”^[1], the methane-to-hydrogen flow ratio for microcrystalline diamond coatings is 1:30–80, and the optimal coating quality can be obtained in the range of 1:40–60.

Effect of Reaction Pressure: It has a significant impact on coating growth. Low pressure facilitates the diffusion and adsorption of active species, promoting the growth of the diamond phase. However, an excessively low pressure may result in an overly slow growth rate, while high pressure tends to increase the probability of non-diamond phase formation. For example, in “Nanocrystalline Diamond Coating and Its Preparation Method and Process”^[1], the deposition pressure for nanocrystalline diamond coatings is 3–8.5 kPa. This range balances diffusion and reaction efficiency, maximizes active group concentration, and achieves a good balance between coating quality and growth rate.

Figure 12. X-ray diffraction patterns of diamond films deposited at different reaction pressures^[14]



Effect of Filament and Substrate Temperatures: The filament temperature (around 2000°C) determines the decomposition efficiency of the reactive gases, which is critical for generating H atoms and carbon active species. Insufficiently low temperature leads to incomplete gas decomposition and slow coating growth, while excessively high temperature accelerates filament degradation and causes reaction runaway. The substrate temperature (600–900°C) regulates the surface behavior of active species, avoiding slow growth at low temperatures and the graphitization of diamond above 1200°C. The combination of 2000°C filament temperature and 850°C substrate temperature is a mature preparation parameter.^[7]

Effect of Deposition Time: Deposition time directly determines the coating thickness; thickness increases with prolonged time, but excessively long deposition tends to increase internal stress and induce cracks. It is necessary to control thickness by adjusting time and manage stress through coordinated parameters to balance thickness and structural integrity. For example, 3–8 μm is commonly used for tool coatings, balancing wear resistance and crack resistance.^[7]

Effect of Auxiliary and Diluent Gases: Auxiliary gases such as N_2 and Ar enable precise regulation of coating properties. N_2 can improve hardness and control nucleation density as well as grain size.

Substrate Temperature: Synergistically regulated by tungsten filament thermal radiation, gas convection and workbench cooling water, it directly affects microscopic reaction processes such as atomic migration and nucleation on the substrate surface. The optimal range is 600–900°C. Too low a temperature results in a slow deposition rate, while too high a temperature tends to form graphite phases. Since diamond graphitizes above 1200°C, the industry commonly uses a deposition temperature of 600–1100°C.

Filament Substrate Distance: Must be strictly controlled within 3–10 mm. A distance $< 3 \text{ mm}$ easily causes overheating of the substrate and subsequent graphitization; a distance $> 10 \text{ mm}$ reduces the active species concentration by 60%, causes the nucleation rate to drop sharply by three orders of magnitude, and halts growth. This range balances thermal radiation and active species transport, achieving coordinated matching of temperature zone and carbon source concentration.

Pressure and Gas Flow Rate: Pressure determines gas density and molecular collision probability; an appropriate pressure

maximizes active group concentration, with a core range of 15 kPa. Flow rate determines gas residence time, typically 50200 sccm, which avoids defects and reduced compactness caused by abnormal residence and ensures coating uniformity. Accurate regulation of the above parameters controls coating nucleation, grain size, phase purity and stress state, determines core properties such as coating hardness, adhesion and wear resistance, and serves as the key regulation basis for the industrial application of HFCVD technology.

3. Applications

With the maturity of HFCVD preparation technology, nanocrystalline diamond coatings have been industrially applied in various fields such as mechanical processing, biomedicine, electronics and optics, and special working conditions by virtue of their excellent comprehensive properties, solving performance bottlenecks that are difficult to break through with traditional materials.

3.1 Mechanical Processing and Cutting Tools

Mechanical processing is the most core application field of nanocrystalline diamond coatings. Its ultra-high hardness, low friction and high chemical stability can significantly improve tool wear resistance and service life, solving the pain points in machining difficult-to-process materials.

For example, the nanocrystalline diamond-coated cemented carbide tools prepared in “Nanocrystalline Diamond Coating and Its Preparation Method and Process”^[1] exhibit greatly extended service life compared with uncoated tools when cutting difficult-to-machine materials such as fiber composites and hot-bent graphite molds. When machining aerospace aluminum alloys, titanium alloys, CFRP and other materials, the tool life is increased by 8–20 times, while cutting friction and built-up edge are reduced. This coating enables high-precision machining of aerospace high-strength aluminum alloys with a hole position tolerance of $\pm 5\text{ }\mu\text{m}$ and surface roughness $R_a < 0.4\text{ }\mu\text{m}$. It reduces the machining surface roughness of CFRP from $R_a\text{ }1.2\text{ }\mu\text{m}$ to below $0.8\text{ }\mu\text{m}$, minimizing burrs and delamination. It also increases the number of tool regrinds and lowers the full-life-cycle cost. When applied to wire drawing and stamping dies, the service life can be prolonged by more than 10 times.

3.2 Biomedical Devices

Nanocrystalline diamond coatings exhibit great application value in the field of biomedical devices due to their good biocompatibility, strong corrosion resistance, excellent wear resistance and tunable surface properties. They are mainly used in surgical knives, dental instruments, orthopedic implants and other scenarios.

In the field of surgical blades, depositing a nanocrystalline diamond coating on the surface of stainless steel scalpels forms a dense physical protective barrier. Its self-corrosion current density is only 1/50 that of uncoated scalpels, effectively resisting corrosion from blood, tissue fluid and disinfectants. Meanwhile, the ultra-high hardness of the coating enables the scalpel to maintain long-term cutting edge sharpness and avoid edge rolling and wear^[2]. The micro-nano composite structure coating prepared by high-temperature oxidation post-treatment can achieve a water contact angle of more than 150° , realizing a superhydrophobic state. The contact area between blood and the blade surface is only 1/5 that of traditional stainless steel blades, and blood residue is reduced by more than 90%, significantly decreasing visual obstruction during surgery and lowering the risk of cross-infection. This makes it ideally suitable for high-precision surgical scenarios such as minimally invasive surgery and cardiovascular surgery.

In the field of dental instruments, diamond-coated dental burs can reduce cutting edge wear by more than 60% during machining, improving the accuracy and efficiency of dental tissue processing. In the field of orthopedic implants, nanocrystalline diamond coatings on titanium alloy artificial joints and bone screws reduce the wear rate by more than 80% compared with traditional CoCr alloys, with annual wear volume controlled below 0.1 mm^3 , avoiding tissue inflammation caused by metal wear ions. Meanwhile, the micro-nano rough structure of the coating promotes the adhesion and proliferation of osteoblasts, enhancing the bonding effect between implants and bone tissue. It strictly complies with the ISO 10993 biocompatibility standard, showing no cytotoxicity or immune rejection.

3.3 Electronics and Optical Devices

Nanocrystalline diamond coatings have excellent optical transparency, ultra-high thermal conductivity, good dielectric properties and chemical stability, and their applications in the field of electronic and optical devices continue to expand. In the

field of optical devices, ultra-thin nanocrystalline diamond films with a thickness of 50-100nm have a transmittance of more than 80% in the 200-800nm visible light band, and have excellent wear resistance, anti-fogging and self-cleaning properties. They can be used as transparent protective coatings for optical components such as mobile phone screens, camera lenses and laser windows. While ensuring optical performance, they significantly improve the scratch resistance and environmental adaptability of components.

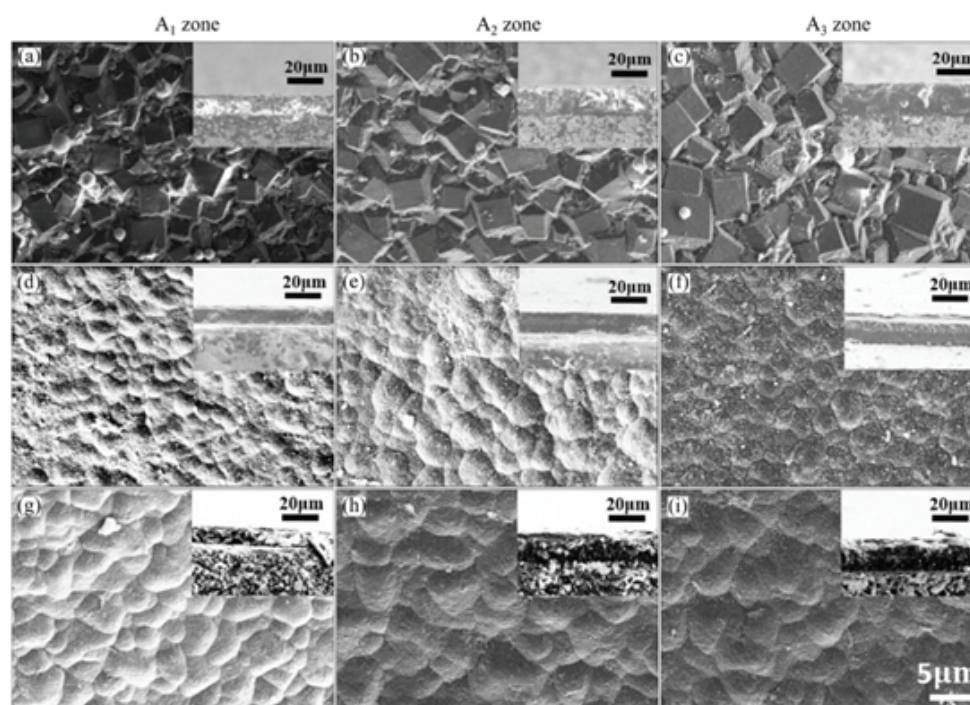
In the field of electronic devices, the wide band gap, high carrier mobility and ultra-high thermal conductivity of nanocrystalline diamond make it a core material for next-generation power semiconductor devices and high-frequency devices. The nanocrystalline diamond coating deposited on the silicon-based surface can significantly improve the heat dissipation capability of chips and solve the thermal management bottleneck of high-power devices. At the same time, nanocrystalline diamond coatings can be used as protective and functional layers for micro-sensors and MEMS devices, improving the sensitivity, stability and service life of devices to meet the application requirements of high-precision and high-reliability electronic devices.

3.4 Core Components Under Special Working Conditions

Nanocrystalline diamond coatings possess a stable sp^3 covalent bond structure, strong corrosion resistance, excellent wear resistance and extreme environmental adaptability, making them a core protective material for special working condition components in aerospace, deep-sea engineering, nuclear energy and other fields. For example, the nanocrystalline diamond composite coating drawing die mentioned in “Preparation and Application Research of Diamond Coated Tools^[15]” has broad application prospects in related industries.

In the field of wear-resistant components such as drawing dies and mechanical seals, the friction coefficient of nanocrystalline diamond coatings can be as low as 0.05-0.1, and the wear rate is reduced by more than 80% compared with traditional WC-Co coatings. The maintenance cycle of mechanical seals for centrifugal pumps and compressors can be extended by 2-3 times, and the service life of drawing dies can be increased by more than 10 times. At the same time, the forming accuracy and surface quality of wires and tubes are significantly improved.

Figure 13. Surface and cross-sectional morphologies of three diamond films on wire drawing dies: microcrystalline diamond (MCD) film; nanocrystalline diamond (NCD) film; composite diamond film^[16]



In extreme environment applications, for the 600-800°C high-temperature and high-speed service conditions of aero-engine bearing seal rings, nanocrystalline diamond coatings can still maintain more than 80% of room-temperature hardness above 600°C without oxidative weight loss, effectively resisting adhesive wear and ensuring sealing performance. In the field of

deep-sea engineering, the coating can form a dense physical barrier, reducing seawater corrosion rate by more than 90% while inhibiting marine biofouling, ensuring the long-term reliable operation of deep-sea equipment in high-pressure and highly corrosive environments. In the nuclear industry, the sp^3 bonds of diamond coatings exhibit excellent radiation stability, resisting erosion by γ -rays and neutron radiation, avoiding substrate embrittlement, and significantly improving the service life of nuclear reactor components.

4. Challenges and Future Research Directions

4.1 Current Technical Bottlenecks

Uniformity of complex-shaped substrates: Complex structures such as deep holes and threads lead to uneven hot-wire radiation and gas convection, resulting in significant differences in coating thickness and composition. The coating thickness difference in deep holes of cylinder blocks reaches 50%, affecting the service life of components.

Internal stress control: During the deposition of thick films $>20\text{ }\mu\text{m}$, the internal stress caused by thermal expansion mismatch between diamond and the substrate (12 GPa for a $25\text{ }\mu\text{m}$ coating) exceeds the bonding strength, making the coating prone to cracking and spalling, which limits the application of thick films.

Surface roughness: Nanocrystalline diamond coatings have $R_a > 30\text{ nm}$ due to grain agglomeration, which cannot meet the ultra-smooth requirement of $R_a < 10\text{ nm}$ for optical components, restricting penetration in high-end fields.

Large-area deposition cost: The investment of a single HFCVD equipment exceeds ten million yuan, with a small number of parts processed per batch and low efficiency. The monthly production capacity of coated tools is only 1,000 pieces, limiting industrial promotion.

4.2 Advanced Research and Development Directions

Low-temperature deposition process ($<400^\circ\text{C}$): Aiming at the thermal degradation of polymer substrates such as PEEK, develop plasma-enhanced/catalyst-assisted CVD technology to realize low-temperature diamond growth and expand application scenarios in biomedicine and flexible electronics.

Nanocomposite coatings (ND-SiC, ND-TiN): Modified by SiC and TiN phases, the fracture toughness of the coating is increased from $2\text{ MPa}\cdot\text{m}^{1/2}$ to $5\text{ MPa}\cdot\text{m}^{1/2}$, meeting the requirements of aerospace impact resistance and tool chipping resistance.

Quantum sensing applications (NV color center regulation): Precisely regulate NV color centers, develop high-sensitivity magnetic sensors, break through the limits of traditional sensing accuracy, and achieve a magnetic field resolution of $1\text{ nT}/\text{Hz}^{0.5}$, which can be used for magnetocardiography detection.

Smart coating systems: Embed shape memory alloys/piezoelectric materials, adaptively adjust the coating structure through temperature/pressure response, realize self-repair and dynamic performance optimization, and improve service reliability in extreme environments.

5. Conclusion

As an efficient method for preparing nanocrystalline diamond coatings, Hot Filament CVD Technology can achieve controllable growth of high-performance coatings by optimizing material selection, process parameters and post-treatment technologies. With the improvement of process maturity and the decrease of cost, the application of nanocrystalline diamond coatings in machinery, biomedicine, electronics and other fields will continue to expand.

Future research should focus on solving key challenges such as deposition on complex-shaped substrates, reducing surface roughness, and developing low-temperature processes, while exploring the innovative applications of diamond coatings in frontier fields such as quantum technology and intelligent materials. Interdisciplinary integration will promote the development of nanocrystalline diamond coating technology towards higher performance and wider applications.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Xu, Y. C., Tang, B. Y., Li, Y. Y., et al. (2024). Nano-diamond coating and its preparation method.
- [2] Zhang, Y. M., Zhou, Z., & Li, Z. H. (n.d.). An anti-blood adhesion scalpel and its preparation method.
- [3] Wang, Z., Feng, Y., Huang, L., et al. (2022). Research progress on application and influence of laser pretreatment in thermal spraying. *Materials Development and Application*, 37(04), 86–92.
- [4] Qin, L., Lin, P., Zhang, Y., et al. (2013). Influence of surface wettability on the tribological properties of laser textured Co-Cr-Mo alloy in aqueous bovine serum albumin solution. 268(Mar.1), 79–86.
- [5] Ma, Y. P., Chen, M., Sun, F. H. (2009). Study on new pretreatment methods for CVD diamond coated cemented carbide substrates. *Diamond & Abrasives Engineering*, (01), 1–4+12.
- [6] Ristic, G., Bogdanov, Z., Trtica, M., et al. (2011). Diamond deposition on thin cylindrical substrates. 76(3).
- [7] Kamble, M. M., Waman, V. S., Ghosh, S. S., et al. (2013). High growth rate of a-SiC:H films using ethane carbon source by HW-CVD method. 36(7), 1177–1185.
- [8] Motahari, H., Bellah, S. M., Malekfar, R. J. A. P. A. (2017). A new tubular hot-wire CVD for diamond coating. 123(6), 411.
- [9] Kudarenko, I. P., Malykhin, S. A., Orekhov, A. S., et al. (2018). Detonation nanodiamond-assisted carbon nanotube growth by hot filament chemical vapor deposition. (1).
- [10] Baek, J., Kim, J., Kim, J., et al. (2019). Growth of graphene on non-catalytic substrate by controlling the vapor pressure of catalytic nickel. 143, 294–299.
- [11] Patterson, R. G. J. J. o. V. S., Surfaces T. A. V., et al. (1984). Vacuum chamber for edgeless studies of gas–metal reactions. 2(1), 82–84.
- [12] Qian, Y. Z., Zhang, T., Wang, S., et al. (2020). Simulation optimization of temperature field of diamond coated tools deposited by HFCVD based on tool bottom heat dissipation mode. *Diamond & Abrasives Engineering*, 40(03), 40–45.
- [13] Komura, Y., Tabata, A., Narita, T., et al. (2008). Influence of gas pressure on low-temperature preparation and film properties of nanocrystalline 3C-SiC thin films by HW-CVD using SiH₄/CH₄/H₂ system. (5), 516.
- [14] Ren, Y. (2009). Preparation of nanocrystalline diamond films by hot filament chemical vapor deposition. [Doctoral dissertation]
- [15] Ma, Y. P. (2007). Research on preparation and application of diamond coated tools. Shanghai Jiao Tong University.
- [16] Wang, C. C., Wang, X. C., Sun, F. H. (2018). Preparation and application of micro-nano composite diamond coated drawing dies based on alternating carbon sources (in English). *Transactions of Nonferrous Metals Society of China*, 28(08), 1602–1610.

Optimization of Key Process Parameters for Air-Jet Vortex Spun Yarn

Kexiang Yang*

School of Management, Xi'an Polytechnic University, Xi'an, Shaanxi, 710600, China

*Corresponding author: Kexiang Yang, 17363642907@163.com

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: As the latest spinning technology in the world at present, air-jet vortex spinning features a short process and high spinning speed by eliminating the two processes of roving and winding. Nevertheless, many problems related to this technology have not been thoroughly solved. To address the difficulty in identifying the key process parameters during yarn formation in air-jet vortex spinning, the input-output relationship of process parameters in the yarn formation process is first analyzed. Based on the collected production process data, a method for screening and optimizing key process parameters is proposed by integrating recursive feature elimination and deep neural network. A deep neural network (DNN) surrogate model is constructed. On the basis of orthogonal experimental design data ($L_{27}(3^{13})$), a high-precision nonlinear mapping between process parameters and quality indices is established. Experimental results show that the average relative error between the predicted values and the measured values of the model is below 3.85%. On this basis, the optimal combination of process parameters (A2B1C3D1E3F1G2H3) that takes into account four quality indices is determined.

Keywords: Process Parameters; Yarn Formation Process; Air-Jet Vortex Spinning; Quality Control

Published: May 11, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1424>

1. Introduction

1.1 Research Background and Importance

As the most advanced spinning machine in the world today, air-jet vortex spinning features a short process (omitting roving and winding processes), high efficiency (spinning speed of 550–600 m/min, about 20–25 times that of ring spinning and 2–3 times that of rotor spinning), and low energy consumption. Due to the vortex bundling principle, few floating fibers are exposed on the yarn surface, and the yarn smoothness is comparable to compact spinning. However, it suffers from core quality problems such as low tenacity and poor evenness^[1]. The fundamental reason is that the air-jet vortex spinning process is an extremely complex nonlinear dynamic coupling system of four bodies: fiber assembly, rigid body, fluid, and elastomer, accompanied by complicated physical and chemical changes. Traditional mechanism modeling methods based on differential equations and empirical trial and error become inadequate when facing such high-dimensional, strongly coupled, and multi-disturbed black-box or gray-box processes, making it difficult to identify key quality-influencing factors and hindering process optimization.

Therefore, the fundamental reasons restricting the accurate prediction of yarn quality are as follows. First, air-jet vortex spinning involves numerous process parameters, such as spinning speed, nozzle pressure, and machine component

parameters, for which yarn quality control feature selection algorithms may not be universally applicable^[2]. Second, existing models have weak anti-interference ability against outliers in small-sample data, poor generalization ability, and low adaptability to small samples.

Accordingly, this paper holds that data-driven methods are the key to breaking the current deadlock^[3]. Instead of completely relying on difficult-to-obtain precise mechanisms, massive data generated in production can be regarded as a new knowledge resource. Advanced learning algorithms can directly mine process rules from data, construct prediction models, and reversely guide optimization.

Taking the difficulty in accurately predicting the yarn quality of air-jet vortex spinning as the entry point^[4], a complete data-driven research framework is proposed. Based on recursive feature elimination, deep neural networks, and broad learning system theories, and aiming at improving yarn quality, this paper systematically studies the key process parameter selection and optimization method integrating recursive feature elimination and deep neural networks^[5].

1.2 Research Status

Process parameters of air-jet spinning include pre-spinning parameters and post-spinning parameters^[6]. Pre-spinning parameters cover raw material specifications, carding speed, drawing speed, number of doublings, sliver linear density, etc. Raw material specifications mainly include fiber length and fineness, raw material moisture regain, and nep content. Post-spinning parameters include directly adjustable parameters such as spinning speed and nozzle pressure, as well as machine component parameters including nozzle orifice diameter, nozzle angle, hollow spindle diameter, and distance from front roller to hollow spindle.

Current research on air-jet vortex spinning parameters is divided into two main aspects. Early studies focused on pre-spinning processes; with technological advancement, the focus shifted to post-spinning processes of air-jet vortex spinning. Rabisankar Chattopadhyaya et al. discussed the structural characteristics of yarn and their influence on yarn performance, finding that yarns produced by different spinning processes have unique and distinct structural features^[7]. Erdumlu et al. showed that the draft ratio has no significant effect on yarn breaking strength, while yarn breaking elongation and breaking work reach the maximum at a middle zone draft ratio of 2.5. When the total draft ratio is maximized, a smaller middle zone draft ratio leads to better evenness^[8].

For post-spinning processes, Basal found that lower spinning speed prolongs fiber passage time and strengthens airflow twisting. With increasing pressure, fiber transfer rate gradually rises. A smaller hollow spindle diameter results in a more compact yarn structure, fewer hairiness, and fewer yarn faults^[9]. Ortlek et al. conducted spinning experiments with pure cotton as raw material to study the effects of yarn fineness, spinning speed, and nozzle pressure on air-jet vortex yarn quality^[10]. Khurshid Alam et al. explored the interactions among fiber composition, spinning process, and yarn performance, and conducted a comparative analysis with conventional ring-spun yarn^[11].

Other scholars have focused on key component parameters of air-jet vortex spinning, especially nozzle structure parameter optimization. Pei et al. performed three-dimensional numerical simulations of flow characteristics in a modified vortex spinning nozzle for core-spun yarn, investigating the effects of process and nozzle parameters including inlet, spindle, and filament feed tube protrusion length^[12]. Shang et al. established a three-dimensional nozzle model and corresponding airflow domain model for numerical calculation, studying airflow behavior with fiber bundles inside the vortex spinning nozzle^[13]. Fang et al. introduced the double-card structure from rotor spinning research into air-jet spinning machines, evaluated the performance of single/double-split jet spinning and nozzle characteristic numbers via simulation, verified the technical advantages of double-split jet spinning, and assessed the influence of nozzle characteristic numbers on the flow field^[14]. Chuan et al. numerically simulated airflow inside the nozzle to design an air-saving vortex spinning nozzle, studying the effects of conical chamber length in the middle section of the vortex tube and inlet diameter on internal airflow^[15].

Current main limitations lie in two aspects. First, the systematization of post-spinning parameter optimization is insufficient. Most studies on the correlation between process parameters and yarn quality rely on time-consuming and discrete manual test data. Such data acquisition not only has limited sample size and high cost but also fails to dynamically and continuously capture the complete correlation between parameter changes and quality fluctuations, leading to information blind spots in

data. Therefore, existing studies are mostly limited to isolated analysis of single or a few post-spinning parameters (e.g., nozzle pressure, draft ratio), lacking the ability to systematically model complex nonlinear interactions among multiple parameters using high-frequency online data. As a result, process optimization schemes are often based on incomplete information, making it difficult to achieve accurate and stable regulation in the global parameter space.

1.3 Research Objectives

Therefore, it is necessary to clarify the yarn formation mechanism of air-jet vortex spinning, investigate and obtain data, screen key process parameters, identify critical factors affecting yarn quality, design orthogonal experiments, construct nonlinear mapping models between process parameters and yarn breaking strength, hairiness H value, and solve the optimal process parameter combination.

2. Material Sources and Quality Index Selection

2.1 Experimental Materials

In this paper, viscose staple fiber is used as the outer fiber, and DTY filament as the core yarn raw material^[16]. Viscose fiber exhibits excellent moisture absorption, dyeability and soft hand feel, which can endow the yarn with good serviceability. DTY filament provides necessary structural support and mechanical properties for core-spun yarn due to its high strength and dimensional stability. The detailed specifications and performance indicators of the raw materials are shown in Table 1.

Table 1 Parameters of Raw Materials

Raw Material Type	Raw Material Specification	Breaking Strength / (cN·dtex ⁻¹)	Breaking Elongation / %	Modulus of Elasticity / (cN·dtex ⁻¹)
Viscose Staple Fiber	1.44 dtex (38 mm)	1.91	18.09	41.75
DTY Filament	55.5 dtex (36 mm)	3.82	16.69	16.01

2.2 Process Flow

The air-jet vortex spinning process flow and corresponding machines adopted in this paper are as follows^[17]:

FA003 Automatic Bale Plucker → MX-6 Multi-bin Blender → CL-C1 Opener → TC10 Carding Machine → TMFD81S Drawing Frame → TMFD81S Drawing Frame → TD8-600 Drawing Frame → VORTEX 870 Air-jet Vortex Spinning Machine. The yarn is packaged according to the winding requirements.

2.3 Experimental Platform and Parameter Selection

The experiments were conducted on a Windows 11 64-bit operating system, with the involved libraries including sklearn (1.3.0), matplotlib (3.3.4), tensorflow (2.10.0), numpy (1.19.5), pandas (0.23.0), pytorch (2.1.2), and keras (2.10.0), and PyCharm was used as the development tool. The data in the research were derived from real production data of multiple branches of a large domestic textile enterprise. The dataset contains a total of 500 samples, each consisting of 16 feature dimensions, including 14 input features (independent variables): {carded sliver linear density, drawn sliver linear density, total draft ratio, main drafting zone gauge, drafting roller pressure, filament feeding tension, nozzle air pressure, yarn drawing speed, winding tension, winding angle, distance from front roller nipper to nozzle inlet, roller gauge, feeding ratio, spinning speed}, and 2 output labels (dependent variables): {breaking strength, hairiness H value}. Data with large-dimensional missing values were eliminated, and those with small-dimensional missing values were filled by the mean value method. The structure of the small-scale and high-dimensional dataset balances multidimensionality and pertinence, aiming to comprehensively capture the key factors affecting the output results^[18].

3. Experiments

3.1 Comparison of Feature Selection Algorithms and Experimental Results

In this section, the stability of three feature selection algorithms is evaluated based on the Jaccard^[19] similarity coefficient and consistency index^[20]. Through multiple random samplings, the Jaccard similarity coefficient and consistency index corresponding to the results of each algorithm are calculated^[21], and finally the average values are obtained for comparison. The stability comparison results are shown in Table 2.

Table 2 Stability Comparison of Feature Selection Algorithms

Feature Selection Algorithm	Quality Indicator	Jaccard Similarity Coefficient	Consistency Index
LR-RFE	Breaking Strength	0.6622	0.7900
	Hairiness	0.7597	0.7807
RF-RFE	Breaking Strength	0.8636	0.9220
	Hairiness	0.7714	0.8157
SVR-RFE	Breaking Strength	0.9067	0.9480
	Hairiness	0.8576	0.8868

Figure 1 Feature Subset Scores of Yarn Breaking Strength Analyzed by SVR-RFE

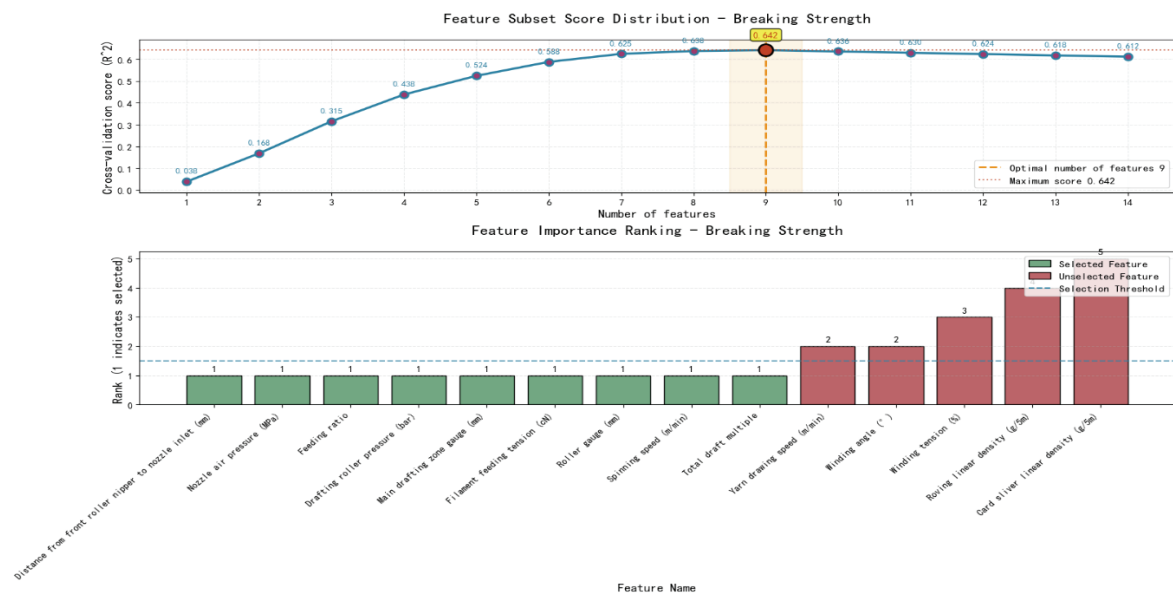
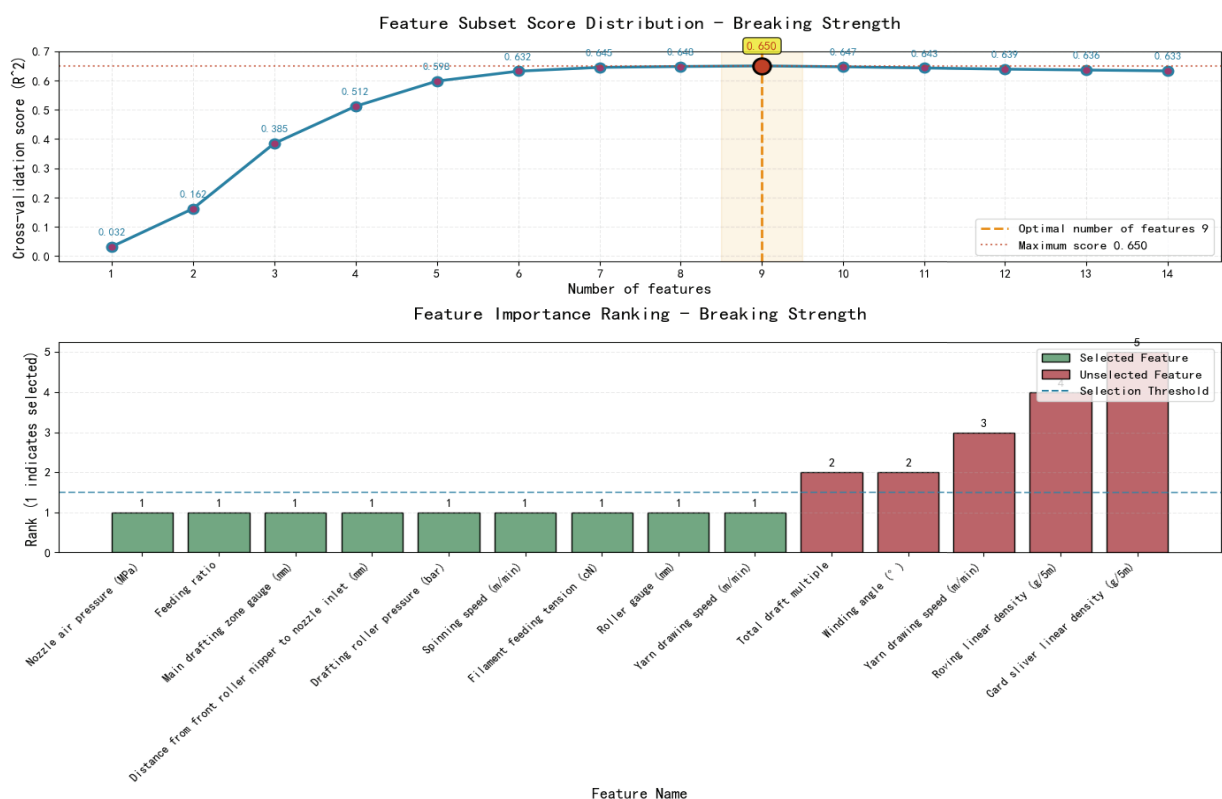


Figure 2 Feature Subset Scores of Hairiness H Value Analyzed by SVR-RFE



It can be seen from Table 2 that under the two yarn quality indicators, the Jaccard similarity coefficient and consistency index of RF-RFE are higher than those of LR-RFE, indicating stronger stability. Meanwhile, the Jaccard similarity coefficient and consistency index of SVR-RFE are both higher than those of the other two algorithms, which means that the feature selection based on the SVR-RFE algorithm is the most stable.

Furthermore, the scores of each feature subset based on yarn breaking strength and hairiness H value were analyzed by the SVR-RFE algorithm, as shown in Figure 1 and Figure 2.

As can be seen from the figures, 9 process parameters affect the two quality indicators during the spinning process, and different process parameters have certain differences in their impacts on different quality indicators. For example, the winding angle parameter of air-jet vortex spinning has a greater impact on yarn breaking strength than on hairiness. The same process parameter has a considerable impact on both indicators, and this study focuses on the process parameters that affect both indicators. Therefore, for subsequent optimization, in terms of selecting key parameters, optimization will be performed on the parameters that have a common impact on the indicators.

3.2 Process Parameter Optimization

Based on the ranking of feature contribution degrees of yarn breaking strength and hairiness H value by the SVR-RFE algorithm, the key process parameters that mainly affect the two indicators can be obtained, namely 8 controllable factors: spinning speed (A), roller gauge (B), nozzle air pressure (C), distance from front roller nipper to nozzle inlet (D), feeding ratio (E), filament feeding tension (F), main drafting zone gauge (G), and drafting roller pressure (H). The specific parameters are shown in Table 3.

Table 3 Process Parameters of Air-Jet Vortex Spun Yarn

Process Parameter	Level 1	Level 2	Level 3
A / (m/min)	360	390	420
B / (mm)	43	44	45
C / (MPa)	0.5	0.6	0.7
D / (mm)	22	26	30
E	1.01	1.02	1.03
F / (cN)	8	14	20
G / (mm)	0.54	0.55	0.56
H / (bar)	0.3	0.35	0.4

Based on the key process parameter information of air-jet vortex spinning obtained by the algorithm, various factors affecting yarn quality in the air-jet vortex spinning process were systematically analyzed. To efficiently study the influence of these factors and determine their significance, the orthogonal experimental design method was adopted. Eight key process parameters were selected as factors, and an $L_{27}(3^{13})$ orthogonal array was designed, which significantly reduced the number of experiments while ensuring the effectiveness of analysis.

A Deep Neural Network (DNN) ^[22] model was constructed by taking the eight process parameters (A, B, C, D, E, F, G, H) in the orthogonal experiment as input variables and four quality indicators as output variables, so as to search for and verify the optimal configuration of process parameters. The DNN model took the eight process parameters as the input layer, adopted multiple hidden layers, and used the sigmoid activation function in each hidden layer to enhance the nonlinear mapping ability and complex feature extraction capability of the model^[23], so as to better capture the complex relationship between process parameters and yarn quality. Finally, the model took yarn breaking strength (maximum), hairiness H value (minimum), breaking elongation (19%), and evenness CV value (minimum) as the output layer, and was trained with optimization algorithms (such as Adam or RMSprop) and loss function (such as Mean Squared Error, MSE) to accurately predict and

verify the optimal process combination^[24].

After setting the DNN structure and hyperparameters, it was necessary to verify the accuracy and generalization ability of the model. Two groups of experimental data not involved in training were randomly selected to compare the measured values and DNN predicted values of yarn quality indicators. The results are shown in Table 4. It can be seen from Table 4 that the average relative errors between the predicted values and the actual measured values of the DNN model are all less than 3.85%, indicating that the model has high reliability and accuracy, and its prediction results can be used as an effective basis for subsequent process optimization. It can be seen from Figure 3 that the DNN model has gradually learned the relationship between process parameters and quality indicators, and can predict the four quality indicators relatively accurately. The model is fully trained and can be used for parameter optimization^[25].

Figure 3 Degree of Difference Between Predicted Values and True Values

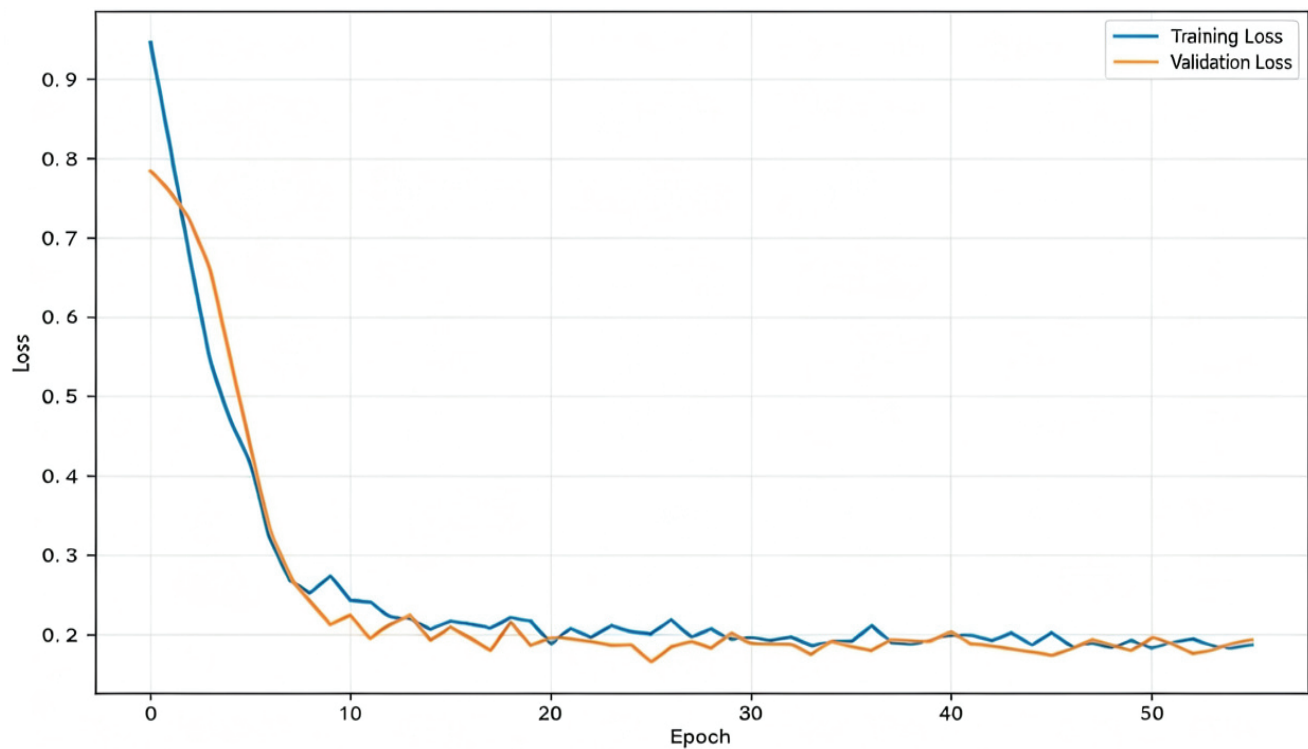


Table 4 Comparative Analysis Results of Measured and Predicted Quality Control Indicators

Quality Control Indicator	Test Batch	Measured Value	Predicted Value	Error / %	Regression Error / %
Breaking Strength	1	20.15	19.72	2.13	2.15
	2	20.45	21.08	3.08	3.10
Hairiness	1	2.45	2.38	2.86	2.90
	2	2.32	2.40	3.45	3.42
Breaking Elongation	1	18.5	19.10	3.24	3.35
	2	19.2	18.60	3.13	3.28
Evenness CV	1	11.35	10.92	3.79	3.82
	2	10.82	11.25	3.97	3.85

To obtain the optimal process parameters and conduct a comprehensive analysis and comparison with the DNN prediction results, an orthogonal experiment was designed to optimize the yarn quality process parameters. The parameter design table is shown in Table 5. The quality indicators of each experiment were measured to evaluate the effects of different yarn forming process parameters on the two quality indicators, and each group of experiments was repeated 10 times.

Table 5 Orthogonal Experiment Parameter Design Table

Test No.	A/(m/min)	B/(mm)	C/(MPa)	D/(mm)	E	F/(cN)	G/(mm)
1	360	43	0.5	22	1.01	8	0.54
2	390	43	0.6	26	1.02	14	0.56
3	420	43	0.7	30	1.03	20	0.55
...	...	...	...	...	...	...	...
26	390	45	0.6	26	1.02	8	0.55
27	420	45	0.7	30	1.03	14	0.54

Based on the validated DNN model, the breaking strength, hairiness H value, breaking elongation and evenness CV value under different yarn forming process parameter combinations were predicted. The optimal process parameter combination was obtained as: spinning speed 390 m/min (A2), roller gauge 43 mm (B1), nozzle air pressure 0.7 MPa (C3), distance from front roller nipper to nozzle inlet 22 mm (D1), feeding ratio 1.03 (E3), filament feeding tension 2 cN (F1), main drafting zone gauge 0.55 mm (G2), drafting roller pressure 0.4 bar (H3), i.e., A2B1C3D1E3F1G2H3.

The visual analysis of the effects of process parameters on breaking strength and hairiness H value is shown in Table 6, where k_1 , k_2 and k_3 represent the mean values of quality indicators at level 1, 2 and 3 in the experiments respectively, R is the range, and σ^2 is the variance.

It can be seen from Table 6 that the influence degrees of different process parameters on breaking strength are in descending order: $E > H > A > F > B > G > D > C$. The most significant factors are E (feeding ratio) and H (drafting roller pressure), with ranges of 1.153 and 0.897, and variances of 0.2218 and 0.1552, respectively.

The influence degrees on hairiness H value are in descending order: $C > H > A > F > G > E > D > B$. The most significant factors are C (nozzle air pressure) and H (drafting roller pressure), with ranges of 0.397 and 0.343, and variances of 0.0279 and 0.0197, respectively.

The influence degrees on breaking elongation are in descending order: $F > E > B > A > D > G > H > C$. The more significant factors are F (filament feeding tension) and E (feeding ratio), with ranges of 1.134 and 0.808, and variances of 0.2183 and 0.1212, respectively.

The influence degrees on evenness CV value are in descending order: $C > F > H > E > G > B > A > D$. The most significant factors are C (nozzle air pressure) and F (filament feeding tension), with ranges of 0.781 and 0.654, and variances of 0.105 and 0.0726, respectively.

The larger the K_n values in Table 6, the larger the breaking strength, hairiness H value, breaking elongation and evenness CV value. It can be further seen from Table 6 that different process parameters have different effects on the four quality indicators. For example, the feeding ratio has the greatest influence on breaking strength but less influence on hairiness and evenness. To obtain the optimal process parameters, combined with the chart analysis, the optimal spinning process parameter configuration is A2B1C3D1E3F1G2H3, which is consistent with the DNN prediction result.

Table 6 Visual Analysis of the Effects of Yarn Forming Process Parameters on Four Quality Indicators

Quality Indicators	Indicator	A	B	C	D	E	F	G	H
Breaking Strength	k_1	19.245	20.057	19.619	19.985	19.172	20.087	19.645	19.051
	k_2	20.084	19.487	19.856	19.522	19.784	19.392	19.887	19.756
	k_3	19.785	19.375	20.071	19.567	20.325	19.484	19.327	19.974
	R	0.839	0.682	0.452	0.463	1.153	0.695	0.56	0.923
	top	3	5	8	7	1	4	6	2
	σ^2	0.1205	0.0892	0.0341	0.0435	0.2218	0.0950	0.0526	0.1552
Hairiness H Value	k_1	2.305	2.448	2.715	2.515	2.472	2.428	2.395	2.618
	k_2	2.438	2.442	2.432	2.51	2.511	2.552	2.342	2.432
	k_3	2.632	2.485	2.318	2.53	2.432	2.585	2.428	2.275
	R	0.327	0.043	0.397	0.02	0.079	0.157	0.086	0.343
	top	3	8	1	7	6	4	5	2
	σ^2	0.0180	0.0004	0.0279	0.0001	0.0010	0.0046	0.0013	0.0197
Breaking Elongation	k_1	18.615	18.211	18.292	18.419	17.827	18.861	18.218	18.279
	k_2	18.422	18.517	18.527	18.227	18.467	18.427	18.156	18.424
	k_3	18.125	18.024	18.528	18.022	18.635	17.727	17.825	18.527
	R	0.49	0.493	0.236	0.397	0.808	1.134	0.393	0.248
	top	4	3	8	5	2	1	6	7
	σ^2	0.0406	0.0413	0.0123	0.0263	0.1212	0.2183	0.0298	0.0103
Evenness	k_1	10.971	11.159	10.573	11.064	10.952	10.902	10.852	11.063
	k_2	11.252	11.052	11.087	10.926	10.825	10.498	11.172	10.722
	k_3	11.093	10.857	11.354	10.974	11.202	11.152	10.932	11.259
	R	0.281	0.302	0.781	0.138	0.377	0.654	0.32	0.537
	top	7	6	1	8	4	2	5	3
	σ^2	0.0132	0.0156	0.1050	0.0033	0.0245	0.0726	0.0185	0.0492

4. Conclusion

Aiming at the problem that yarn quality is affected by multiple process parameters and there are complex interactions between features in the air-jet vortex spinning process, this chapter compared the performance of three estimators—Linear Regression (LR), Random Forest (RF), and Support Vector Regression (SVR)—under the Recursive Feature Elimination (RFE) framework. It was found that SVR-RFE performed the best in both Jaccard similarity coefficient and consistency index (yarn breaking strength: 0.9067 and 0.9480; hairiness H value: 0.8576 and 0.8868), showing the strongest stability, and successfully selected 8 key process parameters from 14 original parameters.

On this basis, the study adopted orthogonal experimental design ($L_{27}(3^{12})$ array) combined with Deep Neural Network (DNN) to construct a nonlinear mapping model between process parameters and yarn breaking strength, hairiness H value, breaking elongation, and evenness CV value. Experimental verification shows that the prediction error of the DNN model is less than

3.85%, demonstrating excellent accuracy and generalization ability.

Through DNN prediction and orthogonal experiment analysis, the optimal process parameter combination considering the four quality indicators was determined as A2 B1 C3 D1 E3 F1 G2 H3, namely spinning speed 390 m/min (A2), roller gauge 43 mm (B1), nozzle air pressure 0.7 MPa (C3), distance from front roller nipper to nozzle inlet 22 mm (D1), feeding ratio 1.03 (E3), filament feeding tension 2 cN (F1), main drafting zone gauge 0.55 mm (G2), and drafting roller pressure 0.4 bar (H3).

Through the systematic method of "feature selection - parameter optimization - model verification", this study provides directly applicable optimal process configuration for textile enterprises, effectively improving yarn quality and production efficiency. It demonstrates the effectiveness and practicality of data-driven algorithms in solving complex industrial optimization problems, and has important theoretical significance for promoting the intelligence of textile production.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Alam, K., Rahman, M., Li, M., et al. (2025). Development and characterization of air jet vortex yarn with novel fiber blends. *SPE Polymers*, 6(1).
- [2] Ingle, N., & Jasper, W. J. (2025). A review of deep learning within the framework of artificial intelligence for enhanced fiber and yarn quality. *Textile Research Journal*, 95(7–8), 904–933.
- [3] Doran, E. C., & Sahin, C. (2020). The prediction of quality characteristics of cotton/elastane core yarn using artificial neural networks and support vector machines. *Textile Research Journal*, 90(13–14), 1558–1580.
- [4] Wang, M., Wang, J., Gao, W., et al. (2023). One-dimensional convolutional neural network with data characterization measurement for cotton yarn quality prediction. *Cellulose*, 30(6), 4025–4039.
- [5] Bhargava, D. (n.d.). Revolutionizing textile quality control with AI: A path to zero defects.
- [6] Liu, Z. Y., Chen, N. L., & Wang, J. (2023). Yarn quality prediction for small samples based on AdaBoost algorithm. *Journal of Donghua University (English Edition)*, 40(3).
- [7] Chattopadhyay, R. (2025). Yarn structure properties relationship. In *Developments in yarn spinning technologies* (pp. 159–186). Elsevier Science Ltd.
- [8] Erdumlu, N., Ozipek, B., Oztuna, A. S., et al. (2009). Investigation of vortex spun yarn properties in comparison with conventional ring and open-end rotor spun yarns. *Textile Research Journal*, 79(7), 585–595.
- [9] Basal, G. (2003). The structure and properties of vortex and compact spun yarns. North Carolina State University.
- [10] Ortlek, H. G. (2006). Influence of selected process variables on the mechanical properties of core-spun vortex yarns containing elastane. *Fibres and Textiles in Eastern Europe*, 14(3), 42.
- [11] Alam, K., Rahman, M., Li, M., et al. (2025). Development and characterization of air jet vortex yarn with novel fiber blends. *SPE Polymers*, 6(1), e10158.
- [12] Pei, Z., Wang, X., Li, Z., et al. (2021). Effect of process and nozzle structural parameters on the wrapping quality of core-spun yarns produced on a modified vortex spinning system. *Textile Research Journal*, 91(15–16), 1841–1856.
- [13] Shang, S., Yu, Z., Sun, G., et al. (2022). Modeling the airflow field of vortex spinning. *Textile Research Journal*, 92(9–10), 1466–1483.
- [14] Haifeng, F., Zhang, J., Sun, H., et al. (2024). Structure analysis of double comb about jet spinning machine based on CFD. *International Journal of Clothing Science and Technology*, 36(4), 582–598.
- [15] Xi, C., Wang, J., Wang, Y., et al. (2024). Effects of structural parameters on air consumption and mechanical energy characteristics of airflow in the vortex spinning nozzle. *Textile Research Journal*, 94(11–12), 1245–1262.
- [16] Liu, G., & Yu, W. (2008). The BP neural network modeling on worsted spinning with grey superior theory and correlation analysis. In *2008 Fourth International Conference on Natural Computation* (Vol. 3, pp. 374–378). IEEE.

- [17] Stephan, et al. (1996). The nature of statistical learning theory by V.N. Vapnik. *Technometrics*, 38(4), 409.
- [18] Wang, Y., Song, J., Fan, T., et al. (2022). Study of yarn quality prediction model based on fuzzy comprehensive evaluation. *Journal of The Institution of Engineers (India): Series E*, 103(2), 305–313.
- [19] Kuo, C. F. J., Tien, C. P., & Chiu, C. H. (2007). Analytical research on intellectual control of yarning characteristics for cotton collocation and rotor spinning. *The International Journal of Advanced Manufacturing Technology*, 32(7), 764–773.
- [20] Majumdar, A. (2010). Modeling of cotton yarn hairiness using adaptive neuro-fuzzy inference system. *Indian Journal of Fibre & Textile Research*, 35(2), 121.
- [21] Ethridge, D., & Zhu, R. (1996). Prediction of rotor spun cotton yarn quality: A comparison of neural network and regression algorithms. In *Proceedings of the Beltwide Cotton Conference* (Vol. 2, pp. 1314–1317).
- [22] Sette, S., Boullart, L., & Van Langenhove, L. (1998). Using genetic algorithms to design a control strategy of an industrial process. *Control Engineering Practice*, 6(4), 523–527.
- [23] Amin, A. E. (2013). A novel classification model for cotton yarn quality based on trained neural network using genetic algorithm. *Knowledge-Based Systems*, 39, 124–132.
- [24] Hu, Z., Zhao, Q., & Wang, J. (2019). The prediction model of cotton yarn quality based on artificial recurrent neural network. In *International Conference on Applications and Techniques in Cyber Security and Intelligence* (pp. 857–866). Springer International Publishing.
- [25] Liu, Z. Y., Li, X. X., Li, L. Q., et al. (2024). Applicability analyses of machine learning algorithm for yarn quality prediction under small samples conditions. *Cotton Textile Technology*, 52(634).

Analysis of Cross-Border Enterprise Cooperation Cases under the Belt and Road Initiative——Take the Padma Bridge of China Railway Group as an Example

Luyao Liang*

School of International Business, Yunnan University of Finance and Economics, Kunming, Yunnan, 650221, China

**Corresponding author: Luyao Liang*

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: During the high-quality development process of the “Belt and Road Initiative”, Chinese multinational enterprises, represented by central enterprises, have become the core carriers for infrastructure connectivity and international production capacity cooperation. The Padma Bridge, as a major national project of Bangladesh and a landmark project of the “Belt and Road” in South Asia, is a typical sample of China Railway Group’s cross-border infrastructure cooperation carried out under complex institutional environments and harsh geological and climatic conditions. This paper, based on the resource-based theory and the scale economy theory, adopts the single-case in-depth Research method and the PEST analysis framework to systematically analyze the business environment identification, risk response, cost control, technological innovation and local integration strategies of China Railway Group throughout the project’s entire life cycle. The research finds that China Railway Group effectively overcomes the institutional distance, resource constraints and environmental constraints of the host country through dynamic political risk prevention and control, economic scale effect release, deep social and cultural integration, and independent innovation of the technical system, achieving the collaborative creation of project commercial value, social value and strategic value. The project not only directly drives the GDP growth of Bangladesh, reconstructs the national transportation network and creates large-scale employment, but also promotes the upgrading of China Railway Group’s internationalization capabilities, the breakthrough of technical barriers and the shaping of a global brand. This paper constructs a business environment identification-cooperation strategy adaptation-scale economy realization-multidimensional value creation” analysis framework for cross-border infrastructure cooperation, extracts replicable practical experiences, provides theoretical references and practical guidance for Chinese enterprises “going global”, and enriches the research system of cross-border enterprise cooperation cases of the “Belt and Road Initiative”.

Keywords: The Belt and Road Initiative; Padma Bridge; China Railway Group; Cross-border Enterprise Cooperation

Published: May 11, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1425>

1. Introduction

Since the launch of the “Belt and Road” initiative, infrastructure connectivity has served as the core driving force, facilitating the rapid development of China’s foreign engineering contracting industry. Central enterprises, leveraging their advantages in funds, technology, and the entire industrial chain, have become the main force in overseas infrastructure investment and construction. However, the countries along the route generally face practical constraints such as insufficient political stability,

weak economic foundation, significant social and cultural differences, and backward technical support. Coupled with the anti-globalization trend and geopolitical competition, Chinese enterprises' cross-border cooperation is confronted with multiple challenges such as institutional distance, cultural conflicts, cost out-of-control, and expanded risk exposure. Politically, the political situation in the host countries fluctuates, policies are changeable, and external interference is strong, making projects prone to interruption; institutionally, standards are not mutually recognized, the approval process is cumbersome, compliance costs are high, and institutional transparency is low; financing-wise, Western multilateral institutions withdraw loans, commercial financing is expensive, and exchange rate risks are high; in the market, local support is weak, labor skills are low, and supply chains are unstable; socially, there are significant cultural differences, insufficient community trust, and high risk of public opinion. Existing studies mostly focus on the macro policy effects or a single risk dimension, lacking systematic analysis of the full-cycle business environment, scale economy realization, and local collaboration of specific projects, and are unable to address the real dilemmas faced by Chinese enterprises in their internationalization practices.

In this context, the Padma Bridge project, which was constructed by China Railway Group, successfully came into fruition under multiple constraints such as the withdrawal of World Bank investment, extremely complex geology, harsh climate, and weak institutional environment. It has become a landmark example for resolving the difficulties in overseas infrastructure operations. Its experience holds extremely strong demonstration value. The Padma Bridge is the largest infrastructure project in Bangladesh since its independence, with a total investment of 3 billion US dollars and a construction period of 8 years. It faced multiple constraints such as the withdrawal of international financial institutions, extremely poor geological conditions, harsh monsoon climate, and shortages of local technology and labor. Ultimately, China Railway completed the construction with its own technology and full-chain services, becoming a landmark achievement of the "Belt and Road Initiative" implementation.

Based on this, this paper raises the core research question: What are the typical institutional and market challenges faced by Chinese infrastructure enterprises in their overseas operations under the "Belt and Road Initiative"? How did the Padma Bridge project overcome the challenges through strategic adjustments and unlock the institutional benefits? What replicable institutional arbitrage and risk management strategies did this project provide for the overseas operations of Chinese enterprises?

This research is based on the "Padma Bridge" project of China Railway Group, a benchmark case of overseas infrastructure construction under the "Belt and Road Initiative". It systematically reviews the entire process of the project's practical experience, including environmental identification, risk management, institutional arbitrage, and value creation. Through in-depth analysis, it explores the internal mechanisms by which overseas infrastructure operations can leverage institutional benefits to break through the institutional constraints of the host country, rely on the risk closed-loop control to solve the overseas operation problems, and achieve sustainable operation by integrating standards with localization. This research not only provides theoretical references for the high-quality development of overseas infrastructure construction under the "Belt and Road Initiative", but also attempts to offer replicable practical models for Chinese enterprises facing institutional distance, political risks, and operational difficulties. It can provide practical guidance for Chinese enterprises to "go global" in response to overseas institutional risks, and also offer a typical sample for the international community to understand China's infrastructure cooperation model. Ultimately, this research will help solve the dual challenges of institutional and risk issues in the overseas operations of Chinese enterprises, provide solid support for the high-quality internationalization of enterprises, and contribute practical paths for cross-border infrastructure cooperation under the "Belt and Road Initiative".

2. Basic Concepts and Theoretical Foundations

2.1 Review of Core Literature

2.1.1 Research on Cooperation between the Belt and Road Initiative and Multinational Enterprises

The "Belt and Road Initiative" reduces the institutional and transactional costs for enterprises' cross-border cooperation through policy coordination, infrastructure connectivity, trade facilitation, financial integration, and people-to-people connectivity. State-owned enterprises, with policy support, financial advantages, and full industrial chain capabilities, occupy a dominant position in overseas infrastructure construction. However, they face risks such as institutional distance, cultural

conflicts, and geopolitical factors. The resource-based theory suggests that heterogeneous resources of enterprises are the core of international competitive advantages. Emerging market multinational enterprises can make up for their shortcomings by local adaptation, risk management, and technological innovation.^[1]

2.1.2 Research on the Business Environment of Cross-Border Infrastructure Projects

Infrastructure projects have characteristics such as large investment, long duration, irreversibility, and dependence on host country policies. The PEST environment directly determines the success or failure of the project. The political environment affects project approval and policy continuity;^[2] the economic environment determines project costs, financing feasibility, and market returns; the social environment involves labor, community, and religious-cultural adaptation; and the technical environment concerns construction feasibility and efficiency. Current research lacks detailed scenario analysis for the South Asian market and super-large bridge projects.^[3]

2.1.3 Research on Economic Cooperation and Infrastructure Projects between China and Bangladesh

Bangladesh is an important node country along the “Belt and Road” initiative. The cooperation between China and Bangladesh centers around infrastructure construction.^[4] The Padma Bridge is a milestone in bilateral cooperation. Most existing studies focus on the macro significance of the project, but lack in-depth exploration of strategies at the enterprise level, cost scale effects, and details of risk management. As a result, it is difficult to form replicable experiences.^[5]

2.2 Theoretical Foundations

2.2.1 Resource-Based Theory (RBV)

The technical, financial, management and policy resources of China Railway Group provide the core support for the project. Through resource integration and local transformation, a sustainable competitive advantage is formed. Economies of scale theory: Large infrastructure projects achieve cost reduction through capacity expansion, standardization of technologies, local procurement, and reuse of facilities, thereby enhancing the economic and sustainable nature of the projects.^[6]

3. Overview of China Railway Engineering Group and the Padma Bridge Project

3.1 Introduction to China Railway Engineering Group

Founded in 1950, China Railway Engineering Group Co., Ltd. (CREC) has evolved into a world-leading comprehensive infrastructure construction enterprise. Its business covers railways, highways, urban rail transit, housing construction, water conservancy and hydropower, airports, ports and other sectors. CREC boasts rich experience in the construction of complex projects such as bridges and tunnels, with world-class construction technologies and management capabilities.

CREC holds an important position in the domestic market and enjoys a high reputation internationally. As shown in Table 1, the Group actively responds to the national “Going Global” strategy and expands overseas business under the Belt and Road Initiative (BRI). Its overseas projects involve railways, highways, bridges, tunnels, urban rail transit and other fields, with institutions and projects distributed in nearly 100 countries across Asia, Africa, Europe, South America and Oceania.

In terms of technological R&D, CREC continuously increases investment, establishes multiple national-level R&D centers and laboratories, and forms a complete scientific and technological innovation system. The Group has won 127 national science and technology progress and invention awards, holds 24,973 patents and 166 national-level construction methods, and has made remarkable breakthroughs in high-speed railways, shield technology, bridge structural health monitoring and other fields.

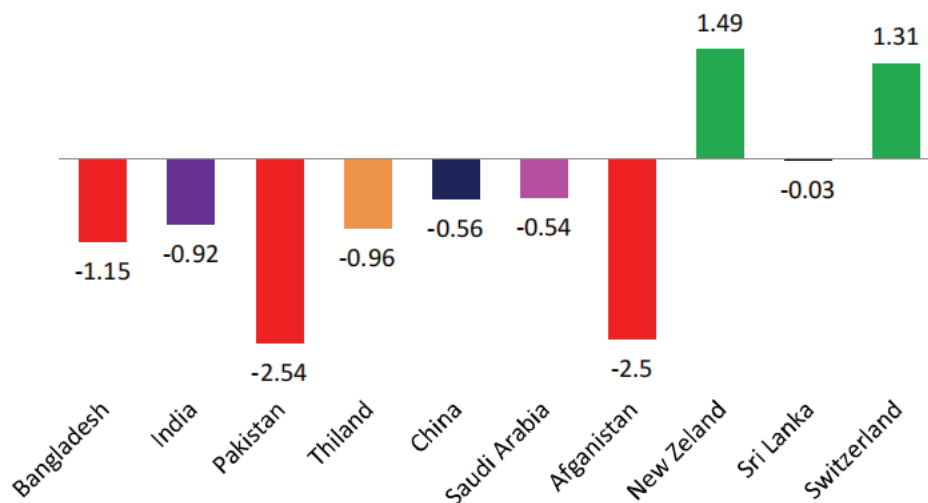
3.2 Background and Significance of the Padma Bridge Project

Known as the “water of life” of Bangladesh, the Padma River divides the country into southwestern and northeastern parts. Previously, no bridge was built downstream of the Padma River due to its wide channel, rapid current and frequent floods, resulting in extremely inconvenient cross-river transportation and severely restricting the economic and social development of the southwestern region. The Padma Bridge project has thus become a key step in the modernization of Bangladesh’s transportation infrastructure.

As shown in Figure 1, the Padma Bridge Railway Link Project starts from the capital Dhaka, passes through the Padma Bridge to Jessore, with a total length of approximately 7.7 km, connecting the capital and 21 southern districts. The project shortens the cross-river travel time from seven to eight hours to just 10 minutes, significantly improving regional

transportation efficiency, promoting the rapid flow of people and materials, enhancing regional connectivity and driving economic development.

Figure 1 Political Stability Index of Some Countries in 2015



As a key infrastructure project promoted by the Bangladeshi government, the Padma Bridge not only resolves transportation bottlenecks but also boosts economic and social development in the southwestern region, improves people's livelihood and enhances the country's international image. As an exemplary project under the BRI framework in South Asia, it has earned international reputation for Chinese enterprises and provided valuable experience for other countries.

In conclusion, the Padma Bridge is not only a milestone in Bangladesh's transportation development but also a model for the successful implementation of the BRI in South Asia, facilitating Bangladesh's better integration into regional economic integration.

3.3 Project Planning and Implementation

Located about 40 km southwest of Dhaka, the capital of Bangladesh, the Padma Bridge crosses the Padma River and serves as a critical transportation hub linking Dhaka with 21 southern districts. It is a rail-cum-road bridge with a total length of 6.15 km, adopting a double-deck steel truss structure: the upper deck is a two-way four-lane highway, and the lower deck is a single-track railway, both designed for a speed of 100 km/h. With a total investment of approximately 3 billion US dollars, it is one of the largest infrastructure projects in Bangladesh's history. Construction started in 2014 and lasted nearly eight years, with official opening to traffic in 2022. At the peak of construction, the project employed 10,000 workers, including a large number of local laborers, effectively boosting local employment and economic development.

4. PEST Analysis of the Project Environment

4.1 Political Environment

Bangladesh has numerous political parties, with the Awami League and the Bangladesh Nationalist Party as the main ones. The current ruling party is the Awami League, which has maintained relatively stable governance by winning multiple elections in recent years. However, fierce political competition between the two parties has occasionally led to social unrest and political uncertainty. Opposition parties frequently organize protests and strikes to express dissatisfaction, highlighting political risks. Furthermore, although the government has intensified crackdowns on Islamic extremist terrorist groups, activities of the East Bengal Communist Party have increased, becoming a new source of political instability.

According to the 2015 World Bank World Political Stability Index, Bangladesh's political stability is relatively weak, which affected the smooth implementation of the Padma Bridge during construction. Fortunately, the bridge was a flagship project of the Awami League government, and Prime Minister Sheikh Hasina's administration regarded it as a symbol of governance capacity and development achievements. Thus, the project received strong government support with streamlined approval procedures. The government also provided necessary land acquisition support and introduced a series of preferential policies

to attract foreign investors in infrastructure construction.

4.2 Political Environment

The opening of the bridge has promoted progress in transportation, logistics, agriculture, industry and other sectors in the southwestern region, created favorable conditions for employment and real estate development, and improved people's living standards. Notably, before the bridge opened, over 90% of Bangladesh's international trade relied on Chittagong Port. The bridge has significantly shortened the distance between Dhaka-Mongla Port and Chittagong Port, elevated Mongla Port's role in cargo transit, accelerated the export of jute, fishery products and other agricultural products in the Khulna-Barisal region, and increased farmers' income and quality of life. According to the Bangladesh Bridge Authority, in the first year after opening, the bridge recorded 5.65 million vehicle crossings and toll revenue exceeding 7.9 billion taka (about 530 million RMB). On June 15, 2024, the day before the Eid al-Adha holiday, 44,033 vehicles crossed the bridge, generating tolls of over 480 million taka (about 30 million RMB), reflecting huge economic benefits.

4.3 Political Environment

The Padma Bridge has been widely anticipated by Bangladeshis, especially residents of the southwestern region, where about 27% of the population long suffered cross-river transportation difficulties. Before the bridge, residents relied on ferries to Dhaka, taking 2 hours one-way (up to 3 hours in rainy seasons) with high safety risks in bad weather.

After opening, the bridge connects 21 southwestern districts with Dhaka, cutting cross-river time to 10 minutes and shortening the distance by at least 100 km, reducing travel costs and boosting economic development.

Economic activities around the bridge have grown significantly, with increases in shops, restaurants, markets and water tourism services. Factories and gas stations have emerged on the north bank, improving living standards. The bridge has also strengthened people-to-people and cultural exchanges between China and Bangladesh, laying a solid foundation for bilateral friendship.

4.4 Political Environment

In recent years, global bridge construction technology has seen continuous innovation in new materials, intelligent monitoring systems and green construction. High-performance concrete and carbon fiber-reinforced composites have improved bearing capacity, corrosion resistance and durability. Digital technologies such as intelligent monitoring and big data analysis have enhanced operation, management and maintenance.

Prefabricated assembly and adaptive construction have become mainstream trends, improving efficiency and reducing safety risks, especially critical for complex environments like Bangladesh. The Padma Bridge adopted advanced technologies including double-deck steel truss structure, high-performance steel and anti-corrosion coatings, representing international advanced levels at the time. However, it still faced challenges from complex geology and extreme climates, and Bangladesh's weak technical foundation and talent shortage affected construction quality and progress.

5. Impacts of the Project on CREC and Bangladesh

5.1 Impacts on CREC

As a landmark project in South Asia, the Padma Bridge marks CREC's further expansion in international infrastructure construction. Its successful implementation enhanced CREC's global brand image and international competitiveness. Cooperation with the Bangladeshi government and enterprises accumulated cross-border project management experience, laying a foundation for exploring Southeast Asian, South Asian and broader markets. Hailed as the "Dream Road," the project showcased CREC's technical and management strengths, attracting global attention.

5.2 Impacts on Bangladesh

The World Bank estimates the bridge will directly boost Bangladesh's GDP by 1.2% and drive related industries. It improves north-south transportation, promotes regional trade and coordinated development of agriculture, manufacturing and services.

The project created over 40,000 direct and indirect jobs during construction, providing vocational training and income growth. Post-opening, logistics, tourism, catering and other industries will flourish, further boosting local economy.

The bridge improves logistics efficiency, promotes import-export growth, ends ferry dependence, shortens transport time and cuts costs. It strengthens economic ties with neighboring countries, advances regional economic integration and elevates

Bangladesh's status in South Asian cooperation.

6. Conclusions and Recommendations

CREC accumulated rich experience from the project, providing guidance for other large international projects. Complex geological and climatic conditions compelled technological innovation, enhancing the Group's technical capabilities. The Padma Bridge positively impacts Bangladesh's economic and social welfare and is crucial to BRI regional development. CREC overcame numerous difficulties, and its successful experience offers valuable references for international infrastructure construction.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Abdal, S. Mansour, W., Agwa, I., Nasr, M., Abadel, A., Onuralp Ozkılıç, Y., & Akeed, M. H. (2023). Application of ultra-high-performance concrete in bridge engineering: Current status, limitations, challenges, and future prospects. *Buildings*, 13(1), 185.
- [2] Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management* 17(1), 99-120.
- [3] Buckley, P. J., Devinney, T. M., & Louviere, J. (2007). Do managers behave the way theory suggests? A choice-theoretic examination of foreign direct investment location decision-making. *Journal of international business studies*, (38), 1069-1094.
- [4] Diego, Q., & Rosario, A. (2023). International strategies of Chinese tourism MNEs: the impact of the belt and Road Initiative and state ownership %J *International Journal of Emerging Markets*. 18(9), 2934-2953.
- [5] Dunning, J. H. (2006). Towards a new paradigm of development: implications for the determinants of international business Transnational corporations. 15(1) 173-227.
- [6] Haiyan, Z., & Filip, D. B. (2024). The impact of contextual distance on the investment locations of Chinese multinationals in countries along the Belt and Road Initiative %J *Asia Pacific Business Review*. 30(2), 220-250.

Design and Implementation of a Cross-Cultural Pattern Classification System Based on ResNet18

Alina Andreeva¹, Zhenning Li², Wenjie Tang², Chaofan Yue³, Xinlei Cao⁴, Xiang Zhang³, Xiangyu Li^{5*}

1. Faculty of Information Technology and Software Engineering, The Bonch-Bruевич St Petersburg State University of Telecommunications, Saint-Petersburg, 193232, Russia

2. Faculty of Engineering and Quality Management, D.F. Ustinov Baltic State Technical University "VOENMEH", Saint Petersburg, 190005, Russia

3. Faculty of Control Systems and Robotics, ITMO University, St. Petersburg, 197101, Russia

4. College of Furnishings and Industrial Design, Nanjing Forestry University, Nanjing, Jiangsu, 210037, China

5. Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

*Corresponding author: Xiangyu Li, xiangyuli@sjtu.edu.cn

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: The automatic recognition of traditional cultural patterns from China, Japan, and Korea poses a challenging cross-cultural image classification problem, compounded by high intra-class stylistic diversity, inter-cultural iconographic overlap, and the scarcity of annotated training data. We propose a lightweight transfer learning framework in which a ResNet18 backbone pre-trained on ImageNet is fully frozen and a compact three-class fully connected head comprising only 2,565 trainable parameters is appended and optimized for the target task, together with a stochastic data augmentation pipeline combining random resized cropping, horizontal flipping, rotation, and Colour jitter. A purposebuilt benchmark of 750 balanced images spanning Chinese, Japanese, and Korean traditional patterns is constructed and partitioned following a 70:15:15 ratio. Comprehensive evaluation against five competing architectures (VGG16, MobileNetV2, EfficientNet-B0, DenseNet121, and a from-scratch SimpleCNN baseline) under identical training conditions demonstrates that the proposed method achieves 87.61% test accuracy and an F1-score of 0.8746, outperforming SimpleCNN by 8.85 percentage points while requiring over four orders of magnitude fewer trainable parameters than VGG16. Controlled ablation experiments reveal that data augmentation contributes the largest single performance gain of 22.00 percentage points, followed by ImageNet pre-training at 11.61 percentage points. Per-class analysis shows that Korean patterns are most reliably discriminated ($F1 = 0.900$), Japanese patterns achieve the highest recall (0.949), and Chinese patterns pose the greatest challenge ($\text{recall} = 0.765$) owing to their broad stylistic range and historical iconographic diffusion into neighboring traditions. These findings establish that a frozen-backbone transfer learning strategy with targeted augmentation provides an efficient and practically deployable solution for cross-cultural heritage pattern recognition under realistic computational constraints

Keywords: Transfer Learning; ResNet18; Cross-cultural Pattern Classification; Cultural Heritage Digitization; Convolutional Neural Network; Data Augmentation; Image Classification

Published: May 11, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1367>

1. Introduction

Traditional patterns encode rich historical information and aesthetic values that have been transmitted across generations. Within the East Asian cultural sphere, China, Japan, and Korea have independently developed distinctive pattern systems: Chinese patterns, exemplified by dragon-and-phoenix motifs and cloud scrolls, embody the ideal of harmony between humanity and nature ^[1]; Japanese patterns, including ukiyo-e compositions and family crests (kamon), priorities minimalist elegance ^[2]; and Korean patterns, characterized by dancheong polychromy and lotus motifs, synthesise Buddhist aesthetics with indigenous sensibilities ^[3]. Despite their cultural significance, accelerating modernisation has placed these traditions at risk, rendering automated recognition an urgent priority ^[4, 5].

Convolutional Neural Networks (CNNs) have been successfully applied to artwork classification, historical document recognition, and style identification ^[6-8]. However, the scarcity of annotated heritage data poses a fundamental challenge: deep learning models require large labelled datasets, yet curating specialised cultural imagery is labourintensive and costly ^[9, 10]. Transfer learning addresses this bottleneck by leveraging representations pre-trained on large-scale datasets ^[11, 12]. Prior work has individually advanced single-culture pattern recognition (e.g. bronze decorations ^[13], ethnic garments ^[4], Chinese painting styles ^[6]) and lightweight CNN design ^[14, 15], but no study has systematically addressed cross-cultural discrimination of Chinese, Japanese, and Korean patterns, and no dedicated benchmark dataset exists for this task.

To address these gaps, this paper makes the following contributions:

1. Benchmark dataset. We construct the first publicly available CJK traditional pattern dataset, comprising 750 curated images equally distributed across three cultural categories.
2. Lightweight transfer-learning classifier. We propose a ResNet18-based ^[16] frozenbackbone strategy yielding only 2,565 trainable parameters while achieving 87.61% test accuracy and an F1-score of 0.8746, representing a parameter reduction of over four orders of magnitude compared with VGG16 ^[17].
3. Systematic multi-model comparison. We benchmark six architectures under identical training conditions (Adam optimiser ^[18], batch size 32, 10 epochs), providing the first empirical accuracy-efficiency analysis on cross-cultural pattern recognition.
4. Component ablation. We quantify the individual contributions of pre-trained weights ^[12] and data augmentation ^[9], revealing that augmentation alone accounts for a 22.00 pp accuracy gain, which is the single largest performance factor in this low-data setting.

2. Methods

2.1 Dataset and Overall Framework

Let $D = \{(x_i, y_i)\}_{i=1}^N$ be the complete dataset ($N = 750$), where $x_i \in R^{H \times W \times 3}$ is the i -th RGB image and $y_i \in \{0, 1, 2\}$ denotes the Chinese, Japanese, or Korean class label. The dataset is partitioned into disjoint training, validation, and test subsets following a 70:15:15 ratio:

$$D = D_{tr} \cup D_{val} \cup D_{te}, \quad |D_{tr}| = 525, \quad |D_{val}| = 112, \quad |D_{te}| = 113. \quad (1)$$

Each cultural category contributes exactly 250 images (Table 1), yielding a perfectly balanced distribution. The predicted class label is

$$\hat{y}_i = \arg \max_{k \in \{0, 1, 2\}} f_{\theta}(x_i)_k, \quad (2)$$

where f_{θ} comprises a frozen convolutional backbone backbone $f_{backbone}: R^{224 \times 224 \times 3} \rightarrow R^{512}$ and a lightweight classification head $f_{head}: R^{512} \rightarrow R^3$. The complete pipeline is illustrated in Figure 1.

Table 1: Dataset Statistics

Category	Number of Images	Percentage
Chinese	250	33.3%
Japanese	250	33.3%
Korean	250	33.3%
Total	750	100%

Figure 1: Overall framework of the proposed CJK pattern classification system

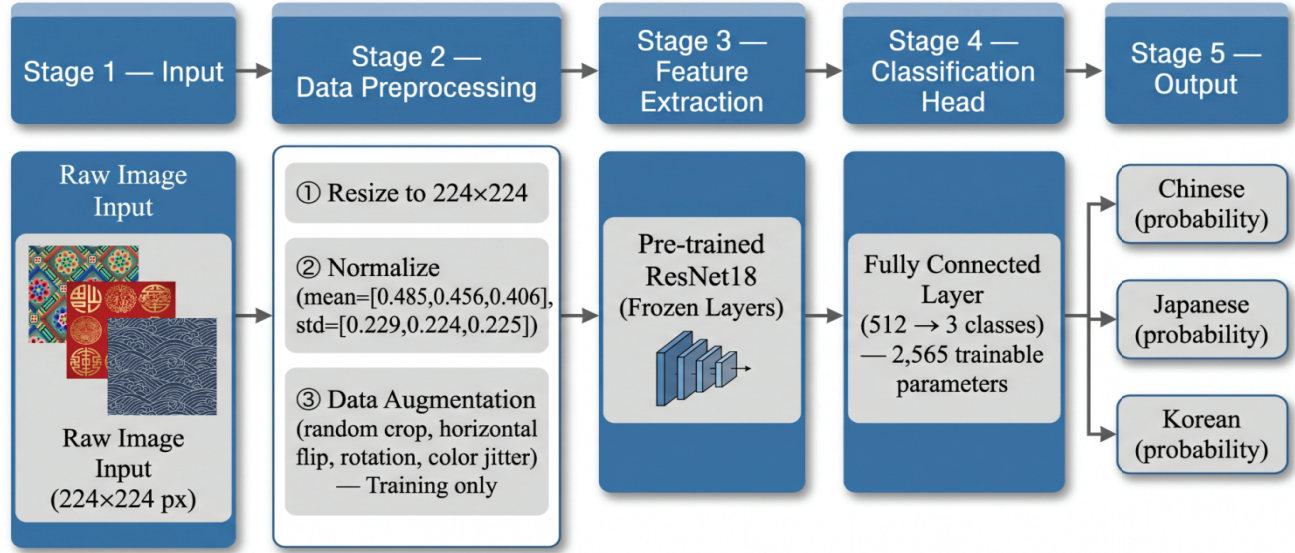


Figure 1 is a data preprocessing (resize, normalize, augment), feature extraction via a frozen ResNet18 backbone pre-trained on ImageNet, and a lightweight fully connected head with only 2,565 trainable parameters.

2.2 Data Preprocessing and Augmentation

All images are resized to 224×224 pixels to match the ResNet18 input requirement^[12, 16]. Channel-wise normalisation uses ImageNet statistics:

$$\widehat{x^{(c)}} = \frac{x^{(c)} - \mu^{(c)}}{\sigma^{(c)}}, \quad c \in R, G, B, \quad (3)$$

where $\mu = (0.485, 0.456, 0.406)$ and $\sigma = (0.229, 0.224, 0.225)$. This preserves the validity of frozen backbone's learned batch normalisation statistics^[19].

To address the limited dataset size, a stochastic augmentation pipeline A is applied exclusively to training images:

$$\tilde{x}_i = A(x_i) = T_{color} \circ T_{flip} \circ T_{rotate} \circ T_{crop}(x_i), \quad (4)$$

Specifically:

1. Random resized cropping (T_{crop}) samples a scale factor $s \in [0.8, 1.0]$, simulating framing variation across museum digitisation protocols^[5];
2. Random rotation (T_{rotate}) samples $\alpha \in [-15^\circ, 15^\circ]$ to introduce misalignment invariance without distorting canonical pattern orientations^[3];
3. Horizontal flipping (T_{flip}) is applied with probability $p = 0.5$, doubling the effective training set without label noise^[9]; Colour jitter (T_{color}) perturbs brightness, contrast, saturation, and hue by $\delta = 0.2$, compensating for illumination and white-balance variability^[6].

Validation and test images receive only resizing and normalisation (Eq. 3).

2.3 Model Architecture: ResNet18 with Frozen Backbone

ResNet18^[16] is selected as the feature extraction backbone for its favourable balance between depth and parameter count ($\approx 11M$ backbone parameters), its residual connections that mitigate the vanishing-gradient problem, and its wide validation in transfer learning studies^[11, 20].

Each residual block computes

$$h^{(l+1)} = F(h^{(l)}, W^{(l)}) + h^{(l)}, \quad (4)$$

where F represents two successive 3×3 convolutions each followed by Batch Normalisation (BN)^[19] and ReLU. When spatial or channel dimensions change, a 1×1 projection shortcut W_s is used. Four sequential stages produce feature maps with 64, 128, 256, 512 channels, progressively capturing low-level textures through to high-level semantic representations.

After the final stage, Global Average Pooling (GAP)^[21] reduces the $7 \times 7 \times 512$ feature map to a 512-dimensional descriptor:

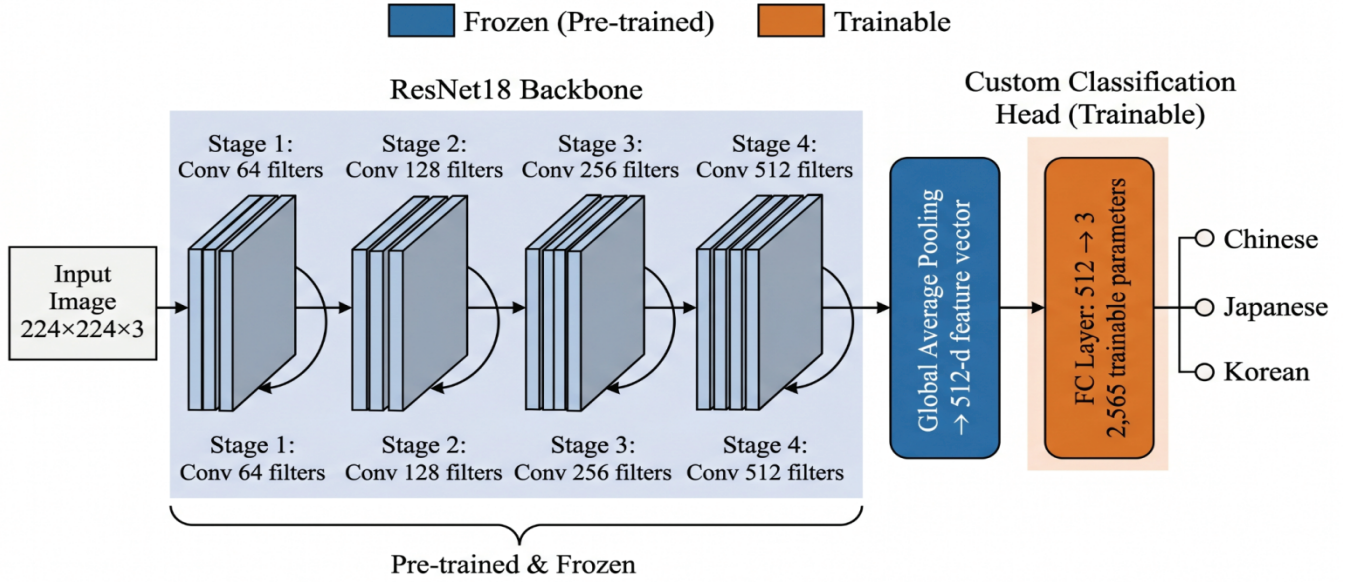
$$z = \frac{1}{H'W'} \sum_{h=1}^{H'} \sum_{w=1}^{W'} F_{h,w}, \quad z \in \mathbb{R}^{512}. \quad (6)$$

The original 1,000-class ImageNet head is replaced by a single fully connected layer:

$$\hat{p}_i = \text{Softmax}(W_c z_i + b_c), \quad (7)$$

where $W_c \in \mathbb{R}^{3 \times 512}$ and $b_c \in \mathbb{R}^3$ yield $512 \times 3 + 3 = 2,565$ trainable parameters. The architecture is illustrated in Figure 2.

Figure2: Architecture of the modified ResNet18 classifier



In Figure 2 the convolutional backbone (blue) is frozen during CJK training; only the fully connected head (orange, 2,565 parameters) is updated via backpropagation.

All backbone parameters are frozen during training. The limited dataset size ($N = 750$, 525 training images) renders full fine-tuning of 11M backbone parameters highly prone to overfitting^[9, 11]. Prior work confirms that ImageNet features transfer constructively to medical imaging^[20], agricultural vision^[22], and artistic classification^[6], even across substantial domain gaps. Freezing also reduces training time by approximately an order of magnitude.

2.4 Training Strategy and Performance Metrics

The model minimises the categorical cross-entropy loss:

$$\mathcal{L}_c = -\frac{1}{|\mathcal{D}_t|} \sum_{i \in \mathcal{D}_t} \sum_{k=0}^2 \mathbb{1}[y_i = k] \log \hat{p}_{i,k} \quad (8)$$

Parameters are updated using the Adam optimiser^[18], with $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. Training runs for $T = 10$ epochs with batch size $B = 32$; the checkpoint maximising validation accuracy is retained:

$$\theta^* \arg \max_{\theta_t, t \leq T} \text{Acc}(f_{\theta_t}, \mathcal{D}_{vl}). \quad (9)$$

Performance is assessed using four complementary metrics. Overall accuracy is

$$\text{Acc} = \frac{1}{|\mathcal{D}_t|} \sum_{i \in \mathcal{D}_t} \mathbb{1}[\hat{y}_i = y_i] \quad (10)$$

Per-class precision and recall are

$$P_k = \frac{TP_k}{TP_k + FP_k}, \quad R_k = \frac{TP_k}{TP_k + FN_k}, \quad (11)$$

and the macro-averaged F1 score is

$$F_1 = \frac{1}{3} \sum_{k=0}^2 \frac{2P_k R_k}{P_k + R_k}. \quad (12)$$

Because the dataset is perfectly balanced (250 images per class), accuracy is a valid primary metric. All experiments are implemented in PyTorch 1.12.1 on a standard CPU platform, with all random seeds fixed at 42 for reproducibility^[8, 22]. The complete training procedure is specified in Algorithm 1.

Algorithm 1. Transfer Learning-Based CJK Pattern Classification**Input:** Dataset D , pretrained ResNet18 weights θ_{pre} , $\eta = 0.001$, $B = 32$, $T = 10$ **Output:** Optimal parameters θ^*

```

1: Split  $D$  per Eq. (1); apply A (Eq. 4) to  $D_{tr}$ ; normalize all splits (Eq. 3)
2: Load  $\theta_{pre}$ ; freeze backbone; replace head with  $(W_c, b_c)$  (2,565 parameters); initialize Adam
3:  $best\_acc \leftarrow 0$ ,  $\theta^* \leftarrow \emptyset$ 
4: for  $t = 1$  to  $T$  do
5:   for each mini-batch  $\subset D_{tr}$  do
6:     Forward pass via frozen backbone and Eq. (7)
7:     Compute  $L_{CE}$  (Eq. 8); backpropagate through head only; update  $(W_c, b_c)$ 
8:   end for
9:   Evaluate  $val\_acc_t$  (Eq. 10)
10:  if  $val\_acc_t > best\_acc$  then
11:     $best\_acc \leftarrow val\_acc_t$  ;  $\theta^* \leftarrow \theta_t$ 
12:  end if
13: end for
14: Evaluate  $P_k, R_k, F_1$  on  $D_{te}$  (Eqs. 11-12)
15: return  $\theta^*$ , evaluation metrics

```

3. Experimental Results

3.1 Model Comparison

Six architectures are benchmarked under strictly identical configurations to ensure a fair comparison.

Table 2 reports the full set of performance metrics alongside the number of trainable parameters for each model.

Table2: Model Comparison Results

Model	Params	Val Acc (%)	Test Acc (%)	Precision	Recall	F1 Score
ResNet18 (Ours)	2,565	90.18	87.61	0.8750	0.8761	0.8746
VGG16	119,566,341	93.75	92.92	0.9362	0.9292	0.9297
MobileNetV2	3,843	87.50	91.15	0.9207	0.9115	0.9133
EfficientNet-B0	3,843	84.82	88.50	0.8867	0.8850	0.8845
DenseNet121	3,075	91.96	92.92	0.9292	0.9292	0.9289
SimpleCNN	51,476,549	70.54	78.76	0.7927	0.7876	0.7848

Several key observations emerge. VGG16 and DenseNet121 jointly achieve the highest test accuracy of 92.92%, yet VGG16 requires $\approx 46,000\times$ more trainable parameters than DenseNet121 (3,075), illustrating that dense connectivity achieves equivalent accuracy at dramatically lower cost. The proposed ResNet18 achieves 87.61% test accuracy with only 2,565 trainable parameters, the smallest footprint across all models, requiring $46,625\times$ fewer parameters than VGG16 while closing only 5.31 percentage points of the accuracy gap. SimpleCNN, despite 51.5M parameters, attains only 78.76%, confirming that raw parameter count is not a reliable proxy for performance under scarce labelled data. Learning dynamics for all six models are shown in Figure 3: all pretrained models converge within three epochs, while SimpleCNN exhibits slow and oscillatory convergence throughout. The parameter-accuracy trade-off is visualised in Figure 4.

Figure 3: Training loss, training accuracy, validation loss, and validation accuracy curves for all six models over 10 epochs.

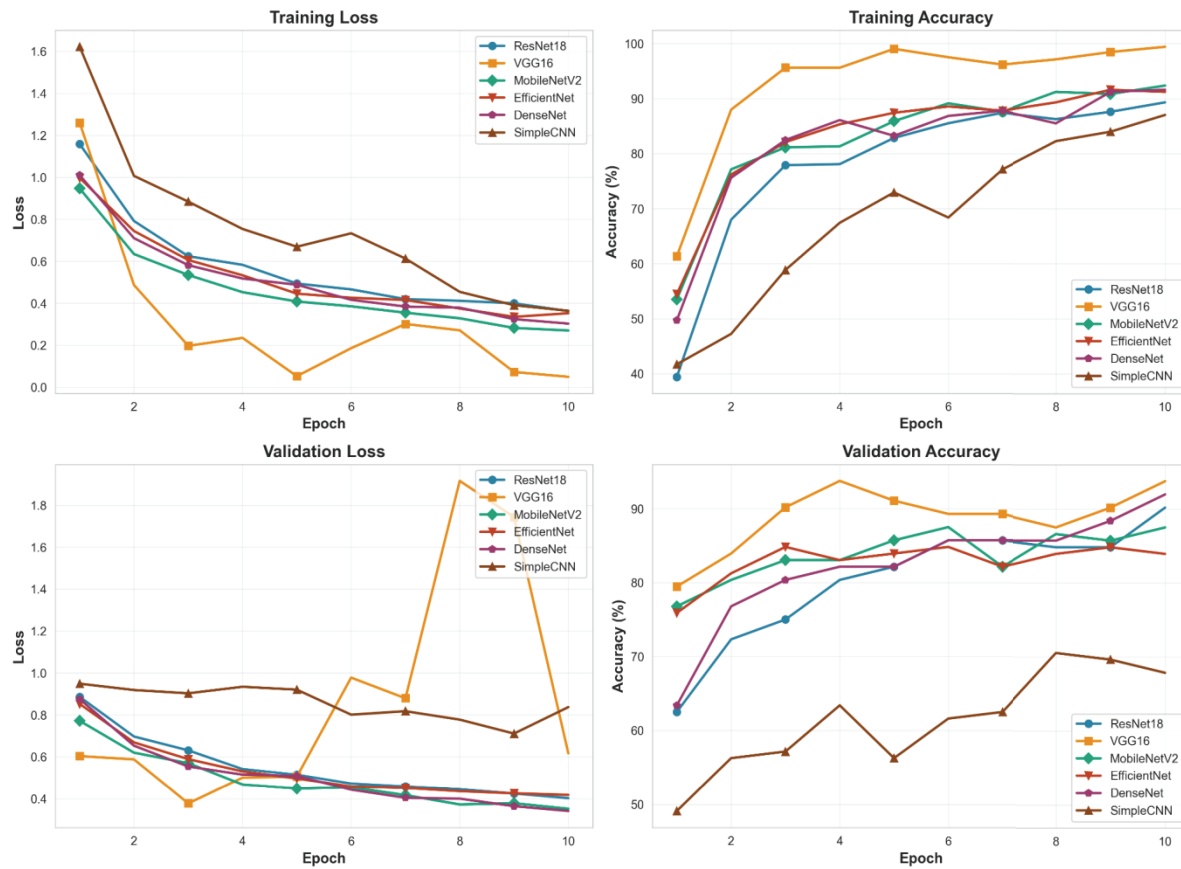
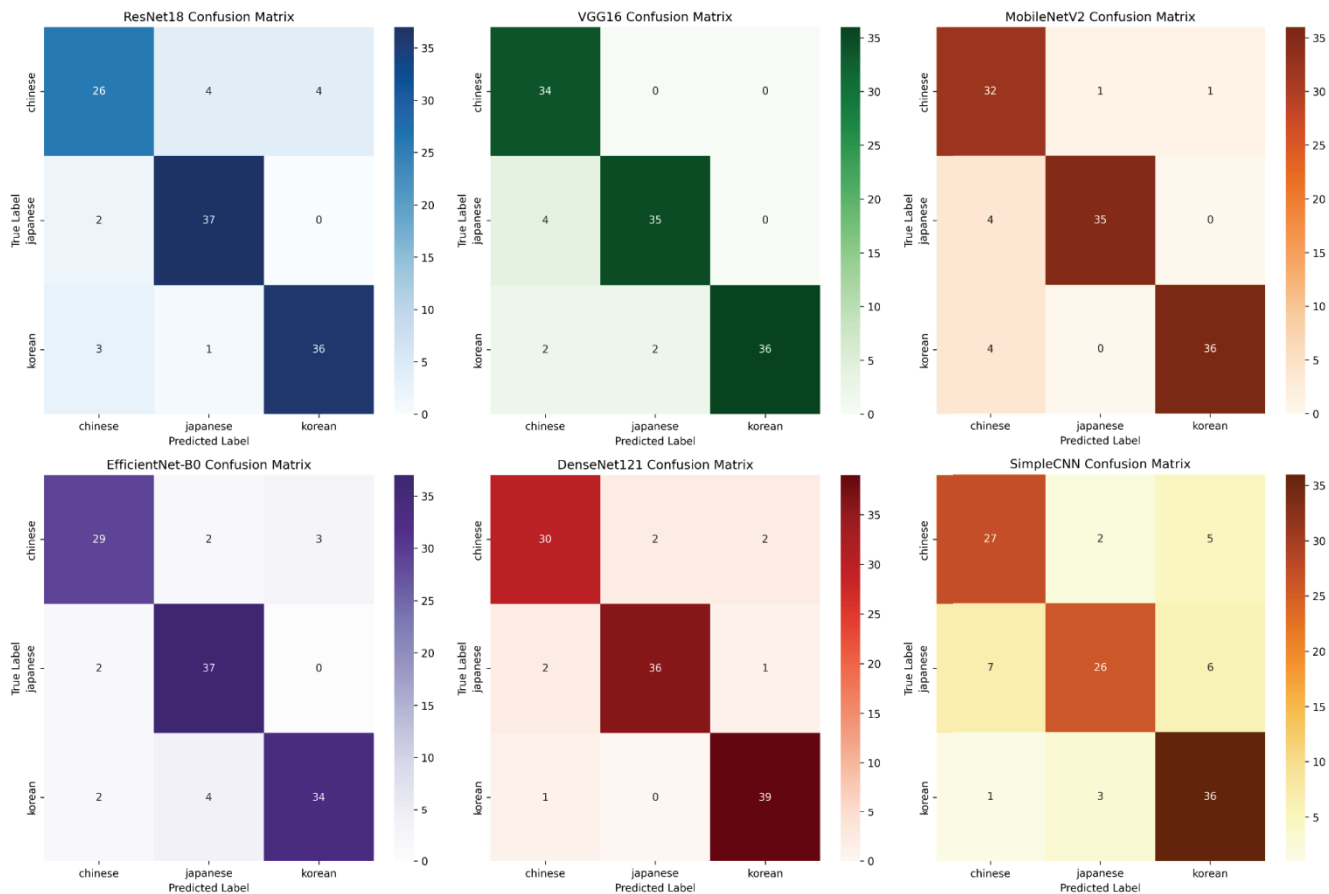


Figure 4: Test accuracy (left) and trainable parameters count on a logarithmic scale (right) for all six models



3.2 Per-Class Performance Analysis

Table 3 disaggregates the precision, recall, and F1 score of the proposed ResNet18 model across the three cultural categories.

Table3: Per-Class Performance Metrics for ResNet18

Category	Precision	Recall	F1 Score
Chinese	0.8387	0.7647	0.8000
Japanese	0.8810	0.9487	0.9136
Korean	0.9000	0.9000	0.9000
Macro avg	0.8732	0.8711	0.8712

Korean patterns achieve the most balanced per-class performance (Precision = Recall = F1 = 0.900), attributable to their distinctive dancheong colouring and symmetric geometric lattice structures that are less common in the other two traditions. Japanese patterns attain the highest recall (0.949), consistent with the strong intra-class consistency of canonical mon crests and ukiyo-e-derived motifs ^[6]. Chinese patterns yield the lowest recall (0.765) and F1-score (0.800): confusion matrices (Figure 5) reveal that the dominant error mode is mutual confusion with Korean patterns, reflecting the historical diffusion of Chinese dragon and cloud scroll motifs into Korean decorative traditions during the Goryeo and Joseon periods ^[3, 13]. The stylistic breadth of Chinese patterns spanning Han embroidery, Tang ceramics, Ming porcelain, and Qing lacquerware creates high intra-class variance that challenges the compact classifier. A radar chart (Figure 6) summarises these per-class asymmetries.

Figure 5: Confusion matrices for all six models on the test set. ResNet 18 concentrates errors in the Chinese-Korean pair

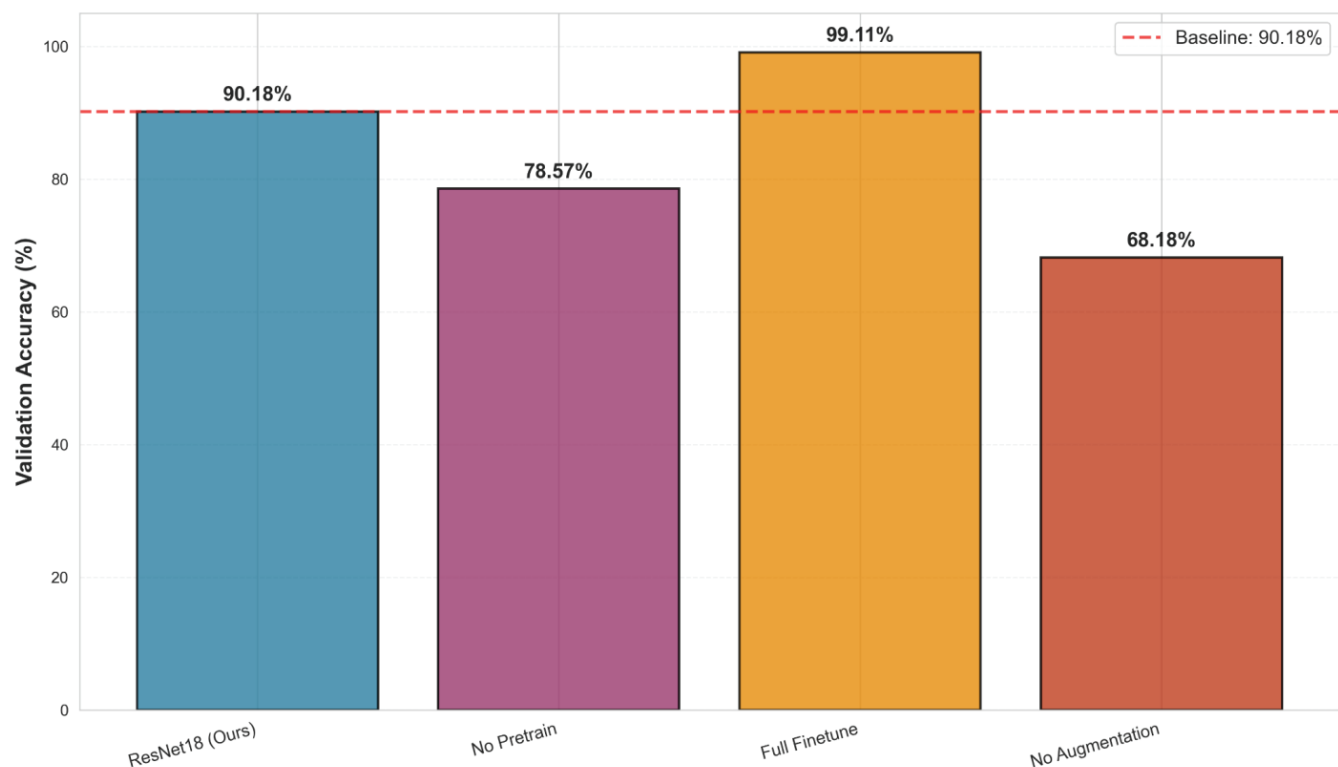
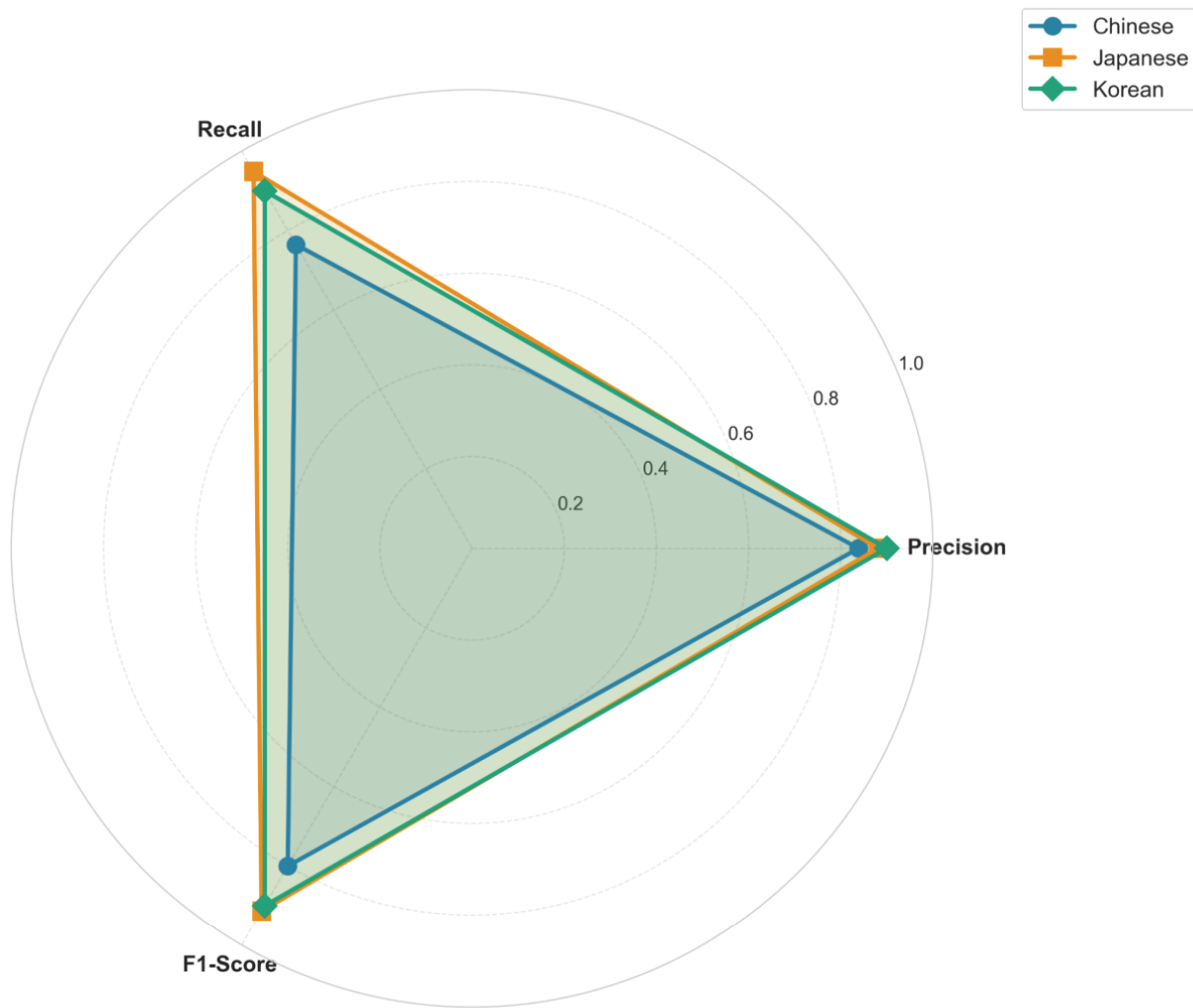


Figure 6: Radar chart of per-class precision, recall, and F1 score for ResNet 18. Korean and Japanese patterns span larger areas than Chinese patterns, reflecting higher per-class discriminability



3.3 Ablation Study

Three controlled ablation experiments isolate the contribution of each pipeline component. Table 4 reports the validation accuracy under each configuration.

Table 4: Ablation Study Results

Configuration	Val Acc (%)	Change (pp)
ResNet18 (Complete Method)	90.18	baseline
w/o Pretrained Weights	78.57	−11.61
Full Model Fine-tuning	99.11	+8.93
w/o Data Augmentation	68.18	−22.00

Data augmentation is the single most critical factor: its removal causes a 22.00 pp drop, demonstrating that geometric and photometric diversity suppresses overfitting more decisively than any other design choice in this low-data regime^[9, 10]. Removing pre-trained weights incurs an 11.61 pp drop, confirming that ImageNet features transfer constructively to traditional pattern imagery despite the substantial domain gap^[11, 12]. Full finetuning raises validation accuracy to 99.11%, but preliminary experiments showed a substantially larger validation-test gap, confirming overfitting to the 525-image training set. The frozen-backbone strategy thus provides a more robust and deployable solution^[20]. Figure 7 visualises the relative magnitude of each component's contribution.

Figure 7: Ablation study validation accuracy

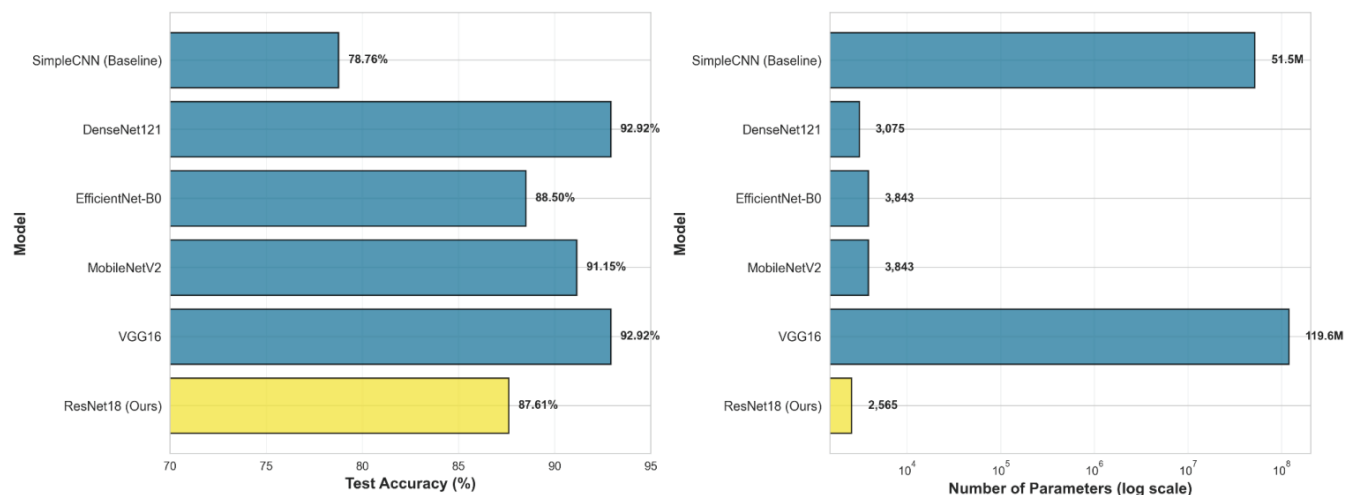


Figure 7 shows the dashed red line marks the complete ResNet18 baseline (90.18%). Data augmentation removal causes the largest drop (−22.00 pp), followed by removing pre-trained weights (−11.61 pp).

4. Discussion

The comparative results reveal a nuanced relationship between model capacity and classification performance. VGG16 and DenseNet121 jointly achieve 92.92% test accuracy, yet their parameter profiles diverge by nearly five orders of magnitude (119.6M vs. 3,075), underscoring that once the backbone is frozen, compact inductive biases can match far larger networks^[11]. Although the proposed ResNet18 ranks fourth in raw accuracy, its accuracy-to-parameter ratio is unmatched: it requires $46,625\times$ fewer trainable parameters than VGG16 while closing only 5.31 percentage points of the accuracy gap, a trade-off highly favourable for mobile cultural tourism applications and museum kiosk systems^[8, 22].

The per-class analysis exposes an inherent asymmetry reflecting real cultural and historical factors. The dominant Chinese-Korean confusion is consistent with the well-documented historical diffusion of artistic motifs during the Goryeo and Joseon periods^[3, 13], and the stylistic breadth of Chinese patterns across five millennia of dynastic production creates high intra-class variance. The ablation results provide concrete design guidance: augmentation and pre-trained weights are complementary rather than redundant, and any extension to additional cultural categories should prioritise collecting and augmenting training data before investing in more complex architectural modifications. The demonstrated ability of a 2,565-parameter classifier to achieve 87.61% accuracy lowers the barrier to entry for heritage institutions operating under budget constraints, consistent with outcomes observed in adjacent fields including cardiovascular risk prediction^[23], software defect detection^[24], IoT anomaly defence^[25], supply chain scheduling^[26], mobile robot path planning^[27, 29], cancer drug synergy prediction^[28], spatio-temporal traffic forecasting^[30], automated scientific writing^[31], and autonomous vehicle motion planning^[32]. Future work should explore attention-based localisation to focus the model on diagnostically distinctive sub-regions, expand the dataset to additional East Asian cultural categories, and investigate self-supervised pre-training on unlabelled heritage imagery to reduce dependence on costly manual annotation.

Conclusion

This paper proposes a lightweight transfer learning framework for the automated classification of Chinese, Japanese, and Korean traditional patterns. By freezing the ImageNet pre-trained ResNet18 backbone and appending a compact three-class head with only 2,565 trainable parameters, the proposed method achieves 87.61% test accuracy and an F1-score of 0.8746, representing a parameter reduction of over four orders of magnitude relative to VGG16 while closing merely 5.31 percentage points of the accuracy gap. Ablation experiments established that data augmentation and ImageNet pre-training are the two most critical design components, contributing 22.00 pp and 11.61 pp gains, respectively. Per-class analysis revealed that Chinese patterns pose the greatest recognition challenge due to high intra-class stylistic diversity and historical iconographic diffusion, while Korean and Japanese patterns are more reliably discriminated. These findings confirm that a frozen-backbone

transfer learning strategy with targeted augmentation provides an efficient and practically deployable solution for cross-cultural heritage pattern recognition under realistic data and computational constraints.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Zhao Liu. (2024). The construction of a digital dissemination platform for the intangible cultural heritage using convolutional neural network models. *Heliyon*, 10(24), e40854. <https://doi.org/10.1016/j.heliyon.2024.e40854>
- [2] Liu, Y., Cheng, P., & Li, J. (2023). Application interface design of Chongqing intangible cultural heritage based on deep learning. *Heliyon*, 9(11), e09450. <https://doi.org/10.1016/j.heliyon.2023.e22166>
- [3] Fu, X. (2021). Research and application of ancient Chinese pattern restoration based on deep convolutional neural network. *Computational Intelligence and Neuroscience*, 2021, 2691346. <https://doi.org/10.1155/2021/2691346>
- [4] Ji, J., Lao, Y., & Huo, L. (2024). Convolutional neural network application for supply-demand matching in Zhuang ethnic clothing image classification. *Scientific Reports*, 14, 4082. <https://doi.org/10.1038/s41598-024-64082-9>
- [5] Belhi, A., Ahmed, H. O., Alfaqheri, T., Bouras, A., Sadka, A. H., & Foufou, S. (2021). Study and evaluation of pre-trained CNN networks for cultural heritage image classification. In *Advances in Intelligent Systems and Computing* (pp. 51-67). Springer. https://doi.org/10.1007/978-3-030-66777-1_3
- [6] Chen, B. (2022). Classification of artistic styles of Chinese art paintings based on the CNN model. *Computational Intelligence and Neuroscience*, 2022, 4520913. <https://doi.org/10.1155/2022/4520913>
- [7] Tan, Y. (2022). Feature recognition and style transfer of painting image using lightweight deep learning. *Computational Intelligence and Neuroscience*, 2022, 1478371. <https://doi.org/10.1155/2022/1478371>
- [8] Lou, J., Shavetov, S., Wen, X., Li, Z., Zhang, X., & Yuan, C. (2026). Deep learning-based instance segmentation: A comprehensive review of algorithms, challenges, and future directions. *The Visual Computer*, 42(6), 235.
- [9] Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2022). Image data augmentation for deep learning: A survey. *arXiv preprint arXiv:2204.08610*.
- [10] Alomar, K., Aysel, H. I., & Cai, X. (2023). Data augmentation in classification and segmentation: A survey and new strategies. *Journal of Imaging*, 9(2), 46. <https://doi.org/10.3390/jimaging9020046>
- [11] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345-1359. <https://doi.org/10.1109/TKDE.2009.191>
- [12] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), 211-252. <https://doi.org/10.1007/s11263-015-0816-y>
- [13] Qi, X.-Z., He, X.-M., Chen, S.-W., & Hai, T. (2024). A framework of evolutionary optimized convolutional neural network for classification of Shang and Chow dynasties bronze decorative patterns. *PLOS ONE*, 19(10), e0293517. <https://doi.org/10.1371/journal.pone.0293517>
- [14] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*(pp. 4510-4520). <https://doi.org/10.1109/CVPR.2018.00474>
- [15] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)* (pp. 6105-6114).
- [16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). <https://doi.org/10.1109/CVPR.2016.90>
- [17] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv*

preprint arXiv:1409.1556.

- [18] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- [19] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)* (pp. 448-456).
- [20] Gu, C., & Lee, M. (2024). Deep transfer learning using real-world image features for medical image classification, with a case study on pneumonia X-ray images. *Applied Sciences*, 14(11), 4083. <https://doi.org/10.3390/app14114083>
- [21] Lin, M., Chen, Q., & Yan, S. (2013). Network in network. arXiv preprint arXiv:1312.4400.
- [22] Cheng, W., Liu, Z., & Li, X. (2026). Enhanced lightweight architecture for real-time detection of agricultural pests and diseases. *Computers, Materials & Continua*, 87(2), 2459-2497. <https://doi.org/10.32604/cmc.2025.074250>
- [23] Yuan, C., Liu, Z., Li, X., Zhou, X., Wang, D., Fan, Y., Sun, X., & Tian, Z. (2025). A dynamic weighted ensemble learning framework for cardiovascular risk prediction in type 2 diabetes: A comparative study with SHAP-based interpretability. *Scientific Reports*.
- [24] Liu, Z., Yuan, C., Liu, H., Li, X., Liu, S., Zhou, X., & Tian, Z. (2026). MSTDP: A multi-scale temporal deep learning framework for just-in-time software defect prediction with cross-attention fusion. *Journal of King Saud University Computer and Information Sciences*.
- [25] Yang, J., Govindarajan, V., Arif, S., Xu, X., Kallel, M., Shaikh, Z. A., Liu, Z., Yuan, C., & Por, L. Y. (2025). SwarmSense-DNN: A trustworthy and decentralized neural framework for proactive anomaly defense in consumer IoT. *IEEE Transactions on Consumer Electronics*.
- [26] Wang, Y., Zhang, H., Yuan, C., Li, X., & Jiang, Z. (2025). An efficient scheduling method in supply chain logistics based on network flow. *Processes*, 13(4), 969.
- [27] Xu, L., Yuan, C., & Jiang, Z. (2025). Multi-strategy enhanced secret bird optimization algorithm for solving obstacle avoidance path planning for mobile robots. *Mathematics*, 13(5), 717.
- [28] Zhang, Y., Yuan, C., Wang, L., Chen, Y., Xing, Y., & Sun, Y. (2025). The structure-preserving spectral graph neural network for dual kinase inhibitors and synergy scoring in gastric cancer. *npj Digital Medicine*.
- [29] Yuan, C., Cai, Y., Que, H., Pei, Y., Zhang, X., Xie, J., Zhang, Q., Mu, L., & Qiao, F. (2026). KA-IHO: A kinematic-aware improved hippo optimization algorithm for collision-free mobile robot path planning in complex grid environments. *Sensors*, 26(8), 2416.
- [30] Yi, X., Yuan, C., Tian, Z., Wu, X., Wu, H., & Liu, L. (2026). BiCoMT: Bidirectional coupling modulation transformer with multi-scale for spatio-temporal traffic flow prediction. *Journal of King Saud University Computer and Information Sciences*.
- [31] Yuan, C., Wei, T., Li, C., Yi, X., Liu, S., Zhang, Z., Cai, Y., & Du, X. (2026). PaperOrchestrator: An LLM-orchestrated multi-agent pipeline for automated end-to-end scientific paper writing. *Journal of King Saud University Computer and Information Sciences*.
- [32] Tian, Z., Lin, Z., Yuan, C., Prajapat, S., Yang, J., & Yee, L. (2026). Consumer-electronics-oriented safe interaction-aware motion planning for ramp driving with perception uncertainty quantification. *IEEE Transactions on Consumer Electronics*.

From Data Possession to Operations Reconfiguration: Transforming Firm Operations under Big Data - An Integrative Literature Review

Mengfei Xiao, Weixiang Gan*

Lincoln University College, Petaling Jaya, Selangor Darul Ehsan-47301, Malaysia

**Corresponding author: Weixiang Gan, Gabriel1995@qq.com*

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: This integrative review develops a theory-oriented account of how big data transforms firm operations. Instead of treating big data analytics capability as a direct antecedent of firm performance, it examines how data is institutionalised, translated into decision routines, embedded in workflows, and orchestrated through platforms and ecosystems. Drawing on recent literature published mainly between 2017 and 2026 across management, information systems, operations, supply-chain, governance, and data-platform research, the review shows that operational transformation is not produced by analytics investment alone. Value emerges when firms combine data governance, privacy control, data quality, analytical culture, decision-right redesign, process instrumentation, automation, supply-chain integration, and platformised data architectures. The paper proposes a Data-to-Operations Reconfiguration Framework in which process visibility, decision-latency compression, exception detectability, and cross-boundary coordination explain how data becomes operational value. By moving beyond the conventional capability–performance view, this review positions big data as a socio-technical and institutional mechanism that reconstructs the operating model of the firm. It further identifies propositions around governance-mediated value creation, curvilinear centralisation effects, platform-enabled SME value capture, and resilience-oriented performance under environmental dynamism.

Keywords: Firm Operations; Operating Model; Big Data Analytics; Data Governance; Data-Driven Decision Making; Process Transformation; Supply-Chain Analytics; SMEs; Data Platforms

Published: May 21, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1384>

1. Introduction

Big data has moved from a specialist information-technology issue to a structural condition of contemporary firm operations. Firms no longer rely only on transactional systems, enterprise resource planning modules and periodic managerial reports. They receive continuous signals from digital platforms, connected products, sensors, customer interfaces, logistics partners, payment systems, service channels and external data markets. This shift alters the grammar of operations. Traditional operating models were often organised around stable process designs, periodic reporting cycles, functional accountability and managerial judgement. Data-intensive operations make inventory, production, procurement, service encounters, demand fluctuations, supplier risks, quality deviations and customer complaints observable at a much finer temporal and spatial granularity. The managerial problem is no longer whether firms possess more data. It is whether data can be converted into an

operating model that is more visible, faster, more adaptive and more accountable than the model it replaces.

The existing literature has established that big data analytics capability is associated with firm performance, innovation, decision quality and supply-chain improvement (Ansari & Ghasemaghahi, 2023; Mikalef et al., 2019; Oesterreich et al., 2022). This evidence is valuable because it demonstrates that analytics is not merely a technical investment. Yet the dominant causal vocabulary remains narrow. Many studies imply that firms acquire data resources, build analytical capability and then achieve superior performance. This formulation compresses the organisational transformation into a single causal arrow. A firm may have large data stores, modern cloud tools and data scientists while still making slow decisions, duplicating data across functions, confining insights to specialist teams or failing to embed predictive outputs into frontline work. Analytics adoption is not the same as operations transformation. The latter requires the redesign of governance arrangements, decision rights, process controls, data architectures and inter-organisational coordination mechanisms.

This distinction matters because recent research shows that performance effects depend on social, organisational and contextual conditions. Meta-analytic evidence suggests that technical factors alone do not explain the business value of analytics; social factors, culture and organisational routines shape value realisation (Oesterreich et al., 2022). Research on data-driven decision making shows that top management support, perceived data quality, analytical culture and centralisation in data use reshape the structure of decision routines rather than simply increasing the amount of information available to managers (Szukits & Móricz, 2024). Data governance and enterprise data marketplace research further indicates that data use depends on rules, metadata, quality assurance, access mechanisms and trust in data sharing (Abraham et al., 2023; Eichler et al., 2023; Nadal et al., 2022). Evidence from SMEs also shows that smaller firms often capture value through platform dashboards, ecosystem access and network capability rather than through a scaled-down version of the large-firm analytics stack (Bar-Gill et al., 2024; Bhatti et al., 2025).

This paper therefore reviews big data as a mechanism of operating-model reconfiguration. It synthesises literature on data governance, privacy, decision-making, process redesign, automation, supply-chain operations, performance realisation, data platforms and SME heterogeneity. The argument is that big data reshapes firm operations through three linked translations. Data must first be institutionalised through governance, privacy, quality, access and architecture. It must then enter decision-system redesign through analytical culture, decision-right redistribution, evidence-based escalation and accountability. It must finally be embedded into workflows through instrumentation, automation, supply-chain integration and platform orchestration. When these translations occur, big data creates operational value through four mechanisms: process visibility, decision-latency compression, exception detectability and cross-boundary coordination.

The contribution of the review is threefold. First, it reframes big data transformation as a multi-level operating-model problem rather than a generic technology-performance relationship. Second, it proposes a Data-to-Operations Reconfiguration Framework that integrates governance, decision redesign, operational embedding, mechanisms and boundary conditions. Third, it develops a research agenda around issues that remain under-theorised: governance-mediated operational value, curvilinear effects of centralisation, platform-based substitution for SME capability constraints and resilience-oriented value creation under environmental dynamism. The review is designed as a theory-building synthesis rather than a descriptive catalogue of digital technologies.

2. Review Approach and Scope

This study adopts a protocol-informed integrative literature review. It does not aim to estimate a pooled effect size or to cover every publication using the phrase big data. Its purpose is to synthesise high-relevance evidence across neighbouring fields that are often reviewed separately: information systems, strategic management, operations management, supply-chain management, data governance, data engineering, platform studies and SME digitalisation. This approach is appropriate because the transformation of firm operations under big data cannot be captured through one disciplinary lens. Governance research explains data access, quality, privacy and trust. Decision-making research explains how analytical evidence changes managerial authority and accountability. Operations and supply-chain research explains forecasting, visibility, integration and responsiveness. Architecture research explains how data lakes, lakehouses, marketplaces and data products affect the cost, timeliness and usability of analytics. SME research shows how resource constraints and platform dependence alter the path of

value realisation.

The review concentrates on studies published between 2017 and 2026, with stronger emphasis on research published from 2021 onward. Earlier studies are retained only where they provide concepts still necessary for interpreting recent evidence, including the strategic business value framework of Grover et al. (2018), Vial's (2019) review of digital transformation, the econometric analysis of BDA and firm performance by Mueller et al. (2018) and the service-innovation analysis by Lehrer et al. (2018). The principal search channels informing the review are Web of Science, Scopus, Google Scholar, IEEE Xplore and ProQuest. Search terms combined concepts related to big data, big data analytics, business analytics, data-driven decision making, data governance, privacy, data products, lakehouse, data marketplace, firm operations, operating models, organisational structure, business processes, automation, supply chains, production operations, SMEs and performance.

Inclusion was guided by conceptual and empirical relevance to firm-level, organisational-level or operational-level transformation. Studies were treated as relevant if they explained mechanisms, boundary conditions or outcomes linking big data with operating models, decision systems, processes, supply chains, platforms or SMEs. Purely technical algorithm papers were considered only when their operational implications were clear. General commentary, marketing material, non-transparent consulting claims and studies without firm-level relevance were excluded from the core evidence base. Official policy materials were used only as institutional context where they clarify privacy, data sharing or industrial data access conditions. They were not treated as substitutes for firm-level evidence.

The synthesis was organised around six themes: data governance and privacy; data-driven decision making and organisational structure; process reconfiguration and automation; supply-chain and production operations; performance measurement and value realisation; and technology stacks, platform ecosystems and SME heterogeneity. The interpretive focus was operational mechanism rather than simple adoption. A study showing that top management support predicts analytics adoption is useful, but its theoretical value for this review depends on whether it explains how analytics changes decision routines, visibility, coordination or performance. This mechanism-oriented reading allows the review to integrate diverse studies into a coherent theory-building framework.

3. Big Data as Operating-Model Reconfiguration

The capability-performance view has provided an important foundation for big data research. Big data analytics capability is usually defined as a configuration of technology, data, human expertise, managerial capability and organisational routines that enables firms to acquire, process, analyse and act on data. Empirical studies and meta-analyses have linked such capability to firm performance, operational performance, innovation and decision quality (Ansari & Ghasemaghahi, 2023; Mikalef et al., 2019; Oesterreich et al., 2022). The field has therefore moved beyond the early question of whether big data matters. The more demanding question is how analytics capability is translated into changes in how firms operate.

The conventional capability-performance view has four limitations when used as the dominant framework. The first is causal compression. A claim that BDA capability improves performance combines many intervening organisational changes into one arrow. It does not reveal whether value comes from better forecasts, faster responses, more standardised processes, earlier detection of exceptions, stronger supplier coordination or more individualised service. The second is organisational abstraction. Capability measures often aggregate technology, skills, management support and culture into broad latent constructs. Such constructs are useful for modelling but can obscure where the transformation occurs in the operating model. The third is insufficient attention to institutionalisation. Data becomes actionable only when it is governed, trusted, discoverable, legally usable, semantically interpretable and linked to workflow authority. The fourth is weak treatment of heterogeneity. The same analytical architecture will not create identical value in retail, manufacturing, logistics, financial services or SMEs.

A more precise perspective treats big data as a force that reconfigures the operating model of the firm. An operating model refers to the way a company organises processes, roles, decision rights, technologies, data flows, controls, performance measures and interdependencies to execute its strategy. Big data changes this configuration in several ways. It improves observability by turning events that were previously invisible or late into analysable signals. It changes response speed by reducing the time between event occurrence, detection, interpretation and action. It alters authority relations by shifting

some decisions from experience-based judgement to model-supported or automated routines. It changes process design by favouring standardised, instrumented and modular workflows. It also transforms cross-boundary coordination through shared dashboards, platform interfaces and data-driven supply-chain integration.

This operating-model perspective connects big data research with digital transformation theory. Digital transformation concerns organisational change induced by digital technologies that alters value creation and value capture (Vial, 2019). Big data is not one more tool within this transformation; it is an infrastructural condition for intelligent workflows, predictive operations, personalised services, algorithmic coordination and ecosystem data exchange. It also creates governance complexity, privacy responsibilities, model risk, data-quality vulnerabilities, skill gaps and tensions between central control and local responsiveness. These challenges are not peripheral. They are part of the transformation process itself.

The conceptual shift is from data possession to data operationalisation. Data possession means that firms hold datasets, platforms, tools and specialists. Data operationalisation means that data is converted into routine action. The former increases analytical potential; the latter changes the mechanics of work. A firm can perform analytics without redesigning decisions, and it can own large data assets without embedding them in processes. Conversely, a firm with limited internal analytics capacity may operationalise data effectively through platform dashboards, supplier portals, external analytics services or ecosystem tools. This distinction helps explain why big data produces asymmetric returns across firms and sectors.

4. Data Governance and Privacy as Operational Infrastructure

Data governance is often treated as a compliance or information-management concern, but the recent literature suggests that it should be understood as operational infrastructure. Big data without governance is fragmented, inconsistent, legally uncertain and difficult to trust. Nadal et al. (2022) show that governance across the data lifecycle is frequently manual or ad hoc, especially where organisations integrate heterogeneous and evolving data sources. Their work on operationalising and automating data governance is relevant for management research because it links data ingestion, transformation, integration, metadata and availability to the practical need to prepare reliable data for business users. Operationally, governance shapes whether a production manager trusts a quality signal, whether a logistics team can combine internal and external data, whether a service team can use customer information lawfully and whether a predictive model receives timely and standardised inputs.

The link between governance and process visibility is direct. Operational visibility is not achieved by collecting more data alone. It requires agreement on definitions, ownership, lineage, access rights, quality thresholds, timeliness and responsibility for correction. A warehouse delay may appear in transport management, inventory control, supplier communication, customer promise and financial penalty systems. If definitions, timestamps and identifiers differ across those systems, the firm may appear to see more while actually creating more confusion. Governance provides the institutional layer that makes data interpretable across functions.

Privacy research strengthens this argument. Gotsch and Schoegel (2023) show that the privacy paradox has an organisational dimension. Firms must balance value extraction from data with legitimacy, trust and regulatory conformance. These tensions directly affect operational use cases involving customers, employees, vendors or connected devices. Consent, purpose limitation, minimisation, access control and reuse rules are embedded in predictive maintenance, personalised service, employee scheduling, behavioural modelling and connected-product analytics. When privacy is treated as an afterthought, firms may build operational systems that later become legally fragile or reputationally damaging. When privacy is built into governance, firms can design workflows that are analytically useful and institutionally sustainable.

Data marketplaces and enterprise data marketplaces deepen the governance issue. Abraham et al. (2023) identify governance decision domains in data marketplaces, including quality, security, architecture, metadata, lifecycle, storage and pricing. Eichler et al. (2023) extend the logic to enterprise data marketplaces, whose goal is to democratise company data. From an operational standpoint, enterprise data marketplaces can reduce search costs, duplication and functional silos by making data products discoverable and reusable. Yet they require rules specifying who may publish data, who may consume it, how quality is certified, how usage is monitored and how incentives for sharing are created. Data democratisation without governance invites misuse; governance without democratisation preserves bottlenecks.

Policy developments reinforce the same logic. The OECD (2024) highlights the links among AI, data governance and privacy as firms face questions around lawful data use, risk management and international cooperation. The European Commission's Data Governance Act and Data Act stress trusted data sharing, data intermediaries, industrial data access and interoperability. For firms, these developments determine what operational data may be exchanged, reused, pooled or monetised across organisational boundaries. They also change the cost and design of analytics. Firms building operations around connected products, industrial machinery, customer data or external data markets must design their operating models around data-rights clarity and trust-based sharing.

The theoretical implication is that data governance should not be modelled only as an antecedent of analytics capability. It should be treated as a mediator between data availability and operational value because it turns raw data abundance into reliable operational visibility. Governance automation is therefore an important capability for data-intensive operations. Manual governance may work when data sources are limited and stable, but it struggles when firms manage many fast-changing and heterogeneous streams. Automated metadata management, lineage tracking, access control and data-quality monitoring are technical practices with direct operational consequences.

5. Data-Driven Decision Making and Organisational Structure

Big data transforms operations partly by changing how decisions are made. Data-driven decision making is not merely the use of dashboards or predictive models. It changes who has authority to decide, what counts as legitimate evidence, how exceptions are escalated, how uncertainty is interpreted and how responsibility is allocated. Janssen et al. (2017) argued that decision-making quality under big data depends on mechanisms that manage data veracity, variety and velocity. Li et al. (2022) later showed that analytics usage improves organisational decision quality through analytics capabilities. Together, these studies suggest that decision quality emerges from the interaction of data, capability, governance and routine.

Szukits and Móricz (2024) provide a useful bridge between analytics and organisation design. Their study links strategic emphasis on analytics, top management support, perceived data quality, analytical culture, data-driven decision making and centralisation in data use. The contribution is the recognition that analytics does not merely make decisions more rational; it changes information asymmetry and authority relations. Centrally managed and analytically standardised data can increase senior managers' visibility into local operations. At the same time, frontline managers may gain faster access to evidence that previously resided in specialised analytical departments. Big data can centralise oversight while decentralising evidence-based action.

Analytical culture deepens this issue. Karaboga et al. (2023) find that data-driven culture mediates the relationship between analytics management capability and operational and financial performance. This implies that capability is insufficient unless managers and employees accept data as a legitimate basis for action. Culture matters because operational decisions are often made under time pressure, incomplete information and competing incentives. A purchasing manager may resist model-based supplier recommendations because of long-standing relationships. A production supervisor may ignore predictive maintenance alerts when output targets are tight. A data-driven culture does not eliminate judgement, but it changes the burden of justification. Decisions that depart from evidence must be explained rather than assumed.

The centralisation question requires more nuance than much of the literature provides. Some centralisation is beneficial because it reduces duplicated datasets, standardises definitions, improves quality and aligns decisions across functions. Excessive centralisation may slow responsiveness, create analytical bottlenecks, detach models from local context and weaken learning at the operational edge. The relationship between centralisation and operational responsiveness may therefore be curvilinear. Low centralisation leaves fragmented data and inconsistent metrics. Moderate centralisation provides shared infrastructure and coherent governance while preserving local action. High centralisation can increase decision latency if operational teams must wait for central approval or specialist interpretation.

Data-driven decision making also changes escalation logic. In conventional operations, escalation often depends on hierarchy, personal experience or threshold-based reports. Under big data, escalation may be triggered by real-time anomalies, predicted risks, statistical control limits or confidence scores. This creates the possibility of decision-latency compression: the reduction in time between operational signal and managerial response. Many operational benefits of big data depend on this mechanism,

including the prevention of stockouts, logistics rerouting, quality control, fraud detection, service-resource reallocation and customer churn intervention. Insight has limited value if it arrives after the operational window has closed.

A further issue is accountability. Big data can make deviations more visible, but visibility does not automatically clarify responsibility. Firms must decide whether accountability belongs to the data owner, model developer, process owner, decision maker or algorithmic system. When algorithm-supported decisions lead to poor outcomes, ambiguity arises: did the model fail, did the manager misinterpret the output, did the data pipeline deteriorate or did the context change? Future research should examine this accountability problem directly. The transformation of firm operations under big data is not only a question of better decisions; it is a redistribution of evidence, authority, responsibility and learning.

6. Process Reconfiguration and Automation

The operational significance of big data becomes most visible when it is embedded in processes. A process is not transformed because managers can analyse it after the event. Transformation occurs when data changes task sequencing, exception detection, resource allocation, quality control, customer service and feedback into redesign. Lehrer et al. (2018) show that BDA enables service innovation through affordances such as sourcing, storage, event recognition and prediction, behaviour recognition and prediction, rule-based action and visualisation. Their insight is that data does not merely inform service; it can become part of the service-production architecture.

This logic extends to production, logistics, procurement and customer operations. In production, data supports predictive maintenance, dynamic scheduling, defect detection, energy optimisation and capacity planning. In logistics, it supports routing, delivery prediction, fleet utilisation, warehouse slotting and risk alerts. In procurement, it supports supplier risk scoring, spend analysis, fraud detection and contract monitoring. In customer operations, it supports triage, churn prediction, personalisation and complaint escalation. In each domain, process transformation depends on instrumentation. A sensor reading, transaction record or customer click becomes operationally meaningful only when it is connected to a workflow that can act on it.

Automation is the most visible form of process reconfiguration, but it should not be treated as a universal endpoint. Automation produces the strongest gains in processes that are modular, repetitive, data-rich and governed by relatively stable rules, such as fraud screening, replenishment triggers, invoice matching, predictive maintenance alerts and routine quality inspection. In tacit, exception-heavy, ambiguous or politically sensitive processes, full automation may be inappropriate. The value of big data may lie instead in decision support, simulation, early warning or scenario analysis. Firms should therefore distinguish automation from augmentation. The more relevant question is not whether data can automate a process, but which parts of the process should be automated, which should be augmented and which require human judgement because of uncertainty, ethics, tacit knowledge or stakeholder consequences.

Process reconfiguration also involves exception detectability. Traditional operations rely on periodic reviews, manual inspections and lagging indicators. Big data allows firms to detect anomalies earlier and at lower levels of aggregation. A machine-learning system may identify a quality anomaly before it becomes a batch failure. A logistics platform may detect a delivery risk before a customer complains. A demand-sensing system may identify market shifts before a monthly planning cycle. Exception detectability matters because operational performance is often damaged by the delay between deviation and recognition. The faster and more variable the environment, the greater the value of early detection.

Data-driven process design also creates risks. Automation can amplify errors where data quality is poor or models are miscalibrated. Measurable signals can crowd out tacit knowledge. Employees may perceive analytics as surveillance rather than support. Algorithmic recommendations can reduce learning when users follow outputs without understanding their assumptions. Process transparency can become process control, raising questions of autonomy and trust. These risks show why social and governance alignment are inseparable from efficiency. Data-driven process design changes the relationship among workers, managers, technologies and performance systems.

A useful research direction is process modularity. Modular processes have repeatable units, clear inputs, outputs and performance criteria. They are more amenable to instrumentation and automation. Highly interdependent processes with tacit coordination, ambiguous objectives or unstable task boundaries require different analytics designs. Future studies should

compare modular and non-modular processes, routine and exception-heavy workflows and standardised and customised services. This would move the field away from general claims that analytics improves operations toward more precise explanations of where and why it does so.

7. Supply-Chain and Production Operations

Supply-chain and production operations offer clear evidence that big data changes operational coordination. Supply chains generate data from suppliers, logistics providers, production systems, retailers, customers, market signals and external risk sources. They also require coordination across organisational boundaries, making them sensitive to data quality, timeliness and trust. Lee and Mangalaraj (2022) show that big data analytics supports supply-chain management through forecasting, logistics, inventory, risk management, sustainability and performance improvement. Gopal et al. (2024) evaluate big data practices such as machine learning, cloud computing, RFID, blockchain, IoT, ERP and business intelligence against criteria including supplier integration, customer integration, cost, utilisation, flexibility, demand management, time and value. These studies point to a shift from optimising individual functions to synchronising data across actors.

The main mechanism is process visibility. In supply chains, visibility means observing demand, inventory, production status, transport flows, supplier reliability and customer requirements across time and organisational boundaries. Big data expands visibility by combining internal transaction records with external signals, connected devices and platform data. Yet visibility is not equivalent to transparency. Firms may see more data without knowing whether it is reliable, comparable or actionable. Cross-boundary visibility requires shared definitions, platform interfaces, data-sharing agreements and trust. This connects supply-chain analytics to governance research: inter-organisational data sharing cannot be sustained if partners fear opportunistic use, quality manipulation, loss of bargaining power or privacy exposure.

Forecasting and demand management illustrate the operational logic. Traditional forecasting depends on historical sales, periodic updates and hierarchical planning cycles. Big data enables demand sensing through real-time sales, search behaviour, weather signals, mobility data, events and platform interactions. The value is not only higher forecast accuracy. It is the ability to adjust production, inventory, replenishment and logistics before demand deviations become costly. This is decision-latency compression at the supply-chain level. Yet forecasting improvements will not translate into performance if procurement lead times, production flexibility, supplier responsiveness or logistics capacity remain rigid. Analytics value requires complementary operational flexibility.

Production operations reveal the same knowledge-action gap. Predictive maintenance can reduce downtime only if maintenance teams can act on alerts, spare parts are available and schedules allow intervention. Quality analytics can identify defects only if process engineers translate patterns into corrective action. Energy analytics can optimise consumption only where equipment and production requirements permit adjustment. Big data creates value when analytical insight is matched by operational capacity to respond. This matching condition is under-theorised in survey research that measures analytics capability and performance outcomes without measuring responsiveness.

Resilience has become a central theme. Lin et al. (2025) show that organisational resilience mediates the relationship between BDA capability and firm performance and that environmental dynamism shapes this relationship. This shifts attention from static efficiency to adaptive capacity. In stable settings, analytics may improve cost, quality and utilisation. Under disruption, value may arise through early warning, scenario analysis, rapid reconfiguration, alternative sourcing, inventory repositioning and customer communication. The meaning of performance is therefore contingent: firms may obtain value not by optimising a steady-state process but by adapting under turbulence.

The supply-chain literature also reveals the importance of platform architectures. Many firms do not collect all relevant supply-chain data internally. They depend on logistics platforms, e-commerce platforms, supplier portals, IoT ecosystems and third-party data providers. Platform-mediated coordination can improve visibility and lower transaction costs, but it can also create dependency, data asymmetry and governance challenges. Large firms may develop proprietary data-integration architectures; SMEs may rely on platform analytics. This difference links supply-chain transformation to firm-size heterogeneity and ecosystem governance.

8. Performance Measurement and Value Realisation

Performance is the most common outcome in big data research, yet it remains conceptually uneven. Some studies examine financial performance, others examine operational performance, innovation, decision quality, supply-chain performance, resilience or customer outcomes. Ansari and Ghasemaghaei (2023) meta-analysed the relationship between BDAC and firm performance, including moderators such as performance type, country and respondent type. Oesterreich et al. (2022) conducted a broader meta-analysis and conceptualised BDA as a socio-technical system in which technical and social factors jointly create business value. These findings are important, but they also show that performance should not be treated as a single endpoint.

Big data creates value through at least five pathways. Efficiency value arises through cost reduction, faster cycles, improved resource utilisation, reduced waste and fewer manual interventions. Quality value arises through defect detection, process standardisation, root-cause analysis and service consistency. Responsiveness value arises through faster sensing, shorter decision latency and dynamic adjustment. Innovation value arises through new services, individualised offerings, data-based business models and experimentation. Resilience value arises through disruption detection, alternative planning, adaptive coordination and recovery speed. Treating all these pathways as general firm performance obscures the mechanisms through which big data works.

Industry characteristics shape these pathways. Mueller et al. (2018) show that the effect of BDA on performance varies by industry characteristics. Industries differ in data intensity, process digitisation, regulation, capital intensity, customer interaction frequency and supply-chain complexity. In digital retail, behavioural data may directly influence pricing, assortment, promotion and service. In manufacturing, sensor data may operate through maintenance, quality and scheduling. In financial services, governance and privacy may condition data use. In logistics, route and location data reshape coordination. A universal performance model will miss these sectoral mechanisms.

Grover et al. (2018) distinguish value creation from value realisation, a distinction that remains essential. Value creation refers to the generation of insight, capability or operational potential. Value realisation refers to the capture of measurable benefits through organisational action. Firms often create insight without realising value because they do not redesign incentives, authority, workflows or measurement systems. A churn model creates value only when service teams can intervene, offers are viable, contact rules permit action and outcomes are measured. A predictive maintenance model creates value only when scheduling and spare-parts logistics can respond.

A stronger measurement agenda would separate leading, intermediate and lagging indicators. Leading indicators include data quality, availability, governance automation, analytical literacy, model deployment and process instrumentation. Intermediate indicators include process visibility, decision latency, exception-detection rate, forecast-adjustment speed, automation depth and coordination frequency. Lagging indicators include cost, quality, customer retention, delivery reliability, innovation output, resilience and financial performance. This layered approach would help researchers avoid attributing performance changes to analytics capability when the operational mechanism is unmeasured. It would also help managers diagnose where value leakage occurs.

9. Data Platforms, Ecosystems and SME Heterogeneity

Technology architecture is often underrepresented in management reviews of big data, yet it matters because it determines what data can be combined, how quickly insights are produced, how expensive analytics becomes and how easily operational users can access data products. Recent work on lakehouses, enterprise data marketplaces and data products suggests that firms are moving from isolated warehouses and data lakes toward integrated and platformised data environments. Armbrust et al. (2021) argue that lakehouse architectures combine features of data warehouses and data lakes while addressing problems such as staleness, reliability, cost, lock-in and limited use-case support. Schneider et al. (2024) review lakehouse concepts and technologies and note that such architectures seek to simplify enterprise analytics environments that suffer from replication, slow processes and high operating cost.

The operational implications are substantial. If data is scattered across disconnected systems, managers rely on stale reports or

manually reconciled spreadsheets. If data is centralised but not curated for operational use, business users remain dependent on specialists for routine questions. If architecture supports reusable data products, governed access, lineage and integration with workflows, operational teams can use data more directly. Architecture therefore shapes decision latency, visibility and scalability. A successful pilot analytics project may fail to transform operations if the underlying platform cannot support repeated deployment across functions and units.

Enterprise data marketplaces add an organisational layer to architecture. Eichler et al. (2023) conceptualise the enterprise data marketplace as a platform for democratising company data. Such marketplaces make data assets discoverable, requestable and reusable. For operations, this can reduce friction between data producers and users. Production teams may discover supplier-quality data maintained by procurement; logistics teams may use customer-delivery preference data maintained by sales; service teams may reuse product telemetry maintained by engineering. These flows are central to operating-model reconfiguration, but they require incentives and governance. Data producers may not document or improve assets if they receive no benefit from reuse. Data consumers may misuse assets if metadata and restrictions are unclear.

Big data operations are increasingly ecosystem-based. Firms depend on external platforms for sales, logistics, payments, advertising, cloud computing and analytics. Platform dependence expands data access but can reduce autonomy. This is particularly relevant for SMEs. Bar-Gill et al. (2024) show in a field experiment on eBay that access to a data-rich seller dashboard increased revenues for e-retailers, partly through active performance monitoring. This evidence matters because it shows that small firms can capture analytics value without building full internal infrastructure. The platform supplies data and interface routines; the SME supplies managerial attention and operational action.

SME research challenges the assumption that smaller firms are simply less mature versions of large firms. Babalghaith and Aljarallah (2024) identify technological, organisational and environmental determinants of BDA adoption in SMEs, including compatibility, complexity, uncertainty, top management support, organisational readiness and data-driven culture. Bhatti et al. (2025) show that BDA capabilities improve MSME innovation and performance through digital platform and network capabilities. This double mediation logic is theoretically significant. It suggests that SMEs may gain value not by internalising all components of a large-firm analytics stack but by using analytics to participate more effectively in platforms and networks. This distinction calls for an ecosystem-based capability view. In traditional resource-based logic, capability tends to reside inside the firm. In platformised data environments, capability may be distributed across cloud providers, marketplaces, software vendors, logistics platforms, e-commerce platforms and partners. A firm's operating model is shaped by how well it orchestrates these distributed analytical resources. The key questions become: who controls the data, who defines the metrics, who owns the models, whether data is portable and whether platform analytics serves the firm's goals, the platform owner's goals or both. These questions should be central to future studies of big data and firm operations.

10. Integrated Discussion: The Data-to-Operations Reconfiguration Framework

The reviewed literature supports an integrated framework explaining how big data transforms firm operations. The Data-to-Operations Reconfiguration Framework consists of three upstream components, four mechanisms, five outcome domains and four boundary conditions. The upstream components are data institutionalisation, decision-system redesign and operational embedding. The mechanisms are process visibility, decision-latency compression, exception detectability and cross-boundary coordination. The outcome domains are efficiency, resilience, innovation, service personalisation and financial value. The boundary conditions are industry data intensity, regulatory burden, SME constraints and environmental dynamism.

Data institutionalisation converts dispersed digital traces into governed, trusted, accessible and legally usable assets. It includes governance, privacy, quality, metadata, lineage, architecture, access rights and platform arrangements. This component is supported by governance research (Abraham et al., 2023; Eichler et al., 2023; Nadal et al., 2022), privacy research (Gotsch & Schoegel, 2023) and policy developments around data sharing and industrial access (European Commission, 2024, 2025; OECD, 2024). Data institutionalisation is foundational because operations depend on reliable and timely signals. Without it, data abundance may increase noise, risk and inconsistency.

Decision-system redesign refers to changes in evidence standards, decision rights, analytical culture, centralisation, escalation logic and accountability. It explains why some firms obtain value from analytics while others produce unused dashboards.

Data must enter managerial routines. It must change what managers attend to, how quickly they act, how they justify decisions and how learning occurs after outcomes are observed. Operational embedding then integrates data and analytics into workflows, automation, process controls, supply-chain interfaces, service systems and performance measures. This is the point at which analytics becomes work. A model that remains in a data-science environment does not transform operations; a model embedded in replenishment, maintenance, routing, customer triage or supplier risk management can do so.

The first mechanism is process visibility. Big data increases visibility when it allows firms to observe process states, dependencies, deviations and performance drivers at greater granularity and speed. Visibility changes control because firms can compare units, identify bottlenecks and coordinate interventions. It may improve accountability, but it can also create surveillance concerns and defensive behaviour. The second mechanism is decision-latency compression. Many operational benefits depend on the speed with which firms move from signal to action. Big data can reduce latency through automated data collection, alerts, predictive models and real-time dashboards, but only when decision rights and operational capacity permit action.

The third mechanism is exception detectability. Operations fail not only because firms lack information but because they recognise exceptions too late. Big data enables earlier detection of anomalies, supplier disruptions, demand shifts, fraud, dissatisfaction and equipment risk. Exception detectability also raises design problems: false positives, alert fatigue, threshold selection and trust in models. The fourth mechanism is cross-boundary coordination. Big data allows functions, business units, suppliers, platforms and customers to coordinate around shared signals. This mechanism is central to supply chains and ecosystems, but it depends on interoperability, common metrics, trust and governance.

The framework generates six propositions for future empirical research. Proposition 1: data governance quality improves operational value primarily through process visibility and decision-latency compression rather than through a direct effect on performance. Proposition 2: centralisation in data use moderates the relationship between analytical culture and operational responsiveness in a curvilinear way; moderate centralisation improves coherence, while excessive centralisation slows response. Proposition 3: platformised data architectures increase coordination capacity, which improves supply-chain resilience and process synchronisation. Proposition 4: automation produces stronger gains in modular, repetitive, data-rich processes than in tacit or exception-heavy processes. Proposition 5: SMEs capture greater value when BDA is complemented by digital platform capability and network capability than when they pursue closed internal analytics stacks. Proposition 6: under high environmental dynamism, big data affects performance mainly through resilience and reconfiguration speed rather than through static efficiency alone.

These propositions shift the unit of analysis. Instead of asking whether BDA capability improves firm performance, researchers can ask which mechanism is activated, in which process domain, under which governance arrangement and with which contextual constraints. This move is essential for theoretical advancement. The field does not need another generic claim that big data improves performance; it needs precise explanations of how data reorganises operating models.

11. Research Gaps and Future Agenda

Several gaps follow from the synthesis. The first concerns governance-mediated operational value. Data governance and operations research remain insufficiently connected. Governance studies focus on quality, metadata, ownership, privacy and access, while operations studies focus on forecasting, automation, supply-chain performance and process improvement. Future research should test whether governance automation improves visibility, whether metadata quality reduces decision latency, whether privacy-by-design increases sustainable data reuse and whether enterprise data marketplaces improve cross-functional coordination.

The second gap concerns decision-right redistribution. Existing research recognises analytical culture and centralisation, but more work is needed on how big data reallocates authority. Does analytics empower frontline workers or strengthen central monitoring? Under what conditions does centralised data governance coexist with decentralised operational action? How should firms design escalation rules when model-generated alerts conflict with managerial judgement? These questions require qualitative studies, field experiments and longitudinal designs that capture decision practices rather than only survey perceptions.

The third gap concerns micro-foundations of process embedding. Many studies measure analytics capability at firm level, but process transformation occurs at the level of workflows, tasks, roles and routines. Future research should examine how specific processes are instrumented, how employees interpret analytics, how models are integrated into workflow systems and how feedback loops are maintained. Process mining, digital trace data, ethnographic observation and mixed-method research could provide stronger evidence than cross-sectional survey models alone.

The fourth gap concerns platform dependence and ecosystem governance. As firms rely on external platforms for analytics, data access and operational coordination, capability becomes distributed. Research should examine whether platform analytics reduces capability gaps for SMEs, whether it creates dependency, how data portability affects autonomy and how platform owners influence the metrics firms use to manage operations. This line of inquiry is practically important because many SMEs will not build large internal data-science teams but will operate through platform analytics.

The fifth gap concerns resilience and dynamic environments. Literature has begun to link BDA with organisational resilience, but more work is needed to distinguish efficiency-oriented and resilience-oriented value. In stable environments, analytics may optimise existing operations. In turbulent environments, it may support sensing, adaptation, reconfiguration and recovery. Future studies should measure disruption frequency, adaptation speed, recovery quality and environmental dynamism. They should also examine whether firms over-invest in efficiency analytics while under-investing in resilience analytics.

The sixth gap concerns negative consequences and trade-offs. Big data operations can produce surveillance, privacy risk, algorithmic rigidity, model dependence, alert fatigue, data-quality vulnerabilities and power asymmetries. The same visibility that improves coordination may increase employee stress. The same platform dashboard that improves SME revenue may deepen platform dependence. The same automation that reduces cost may reduce local learning. Future research should examine the conditions under which big data transformation creates operational fragility rather than resilience.

12. Practical Implications and Conclusion

For executives, big data transformation should be governed as operating-model redesign rather than as an IT programme. Investment in cloud infrastructure, analytical tools or data-science talent will not transform operations unless decision rights, workflows, incentives and performance measures are redesigned. Senior leaders should ask four diagnostic questions: which operational decisions are too slow, which processes are insufficiently visible, which exceptions are detected too late and which organisational boundaries prevent coordination. These questions are more useful than asking only whether the firm has adopted advanced analytics.

For data and technology leaders, architecture should be evaluated by operational usability. A sophisticated platform has limited value if operational teams cannot discover, trust and act on data assets. Governance should be automated where possible, metadata should be readable by business users and data products should be linked to decision use cases. Enterprise data marketplaces should not become catalogues of poorly documented assets. They should become governed interfaces between data producers and operational users.

For operations managers, analytics should be embedded selectively. Not every process should be automated. Managers should distinguish processes suitable for full automation, processes suitable for decision augmentation and processes requiring human judgement. They should monitor decision latency, exception detection, alert quality and workflow adoption, not only final performance outcomes. For SMEs, value does not necessarily require a complete internal analytics stack. SMEs can capture value through platform dashboards, external analytics services, network relationships and disciplined routines, while also monitoring data portability and dependency risks.

The transformation of firm operations under big data is not a linear story of technology adoption followed by performance improvement. It is a reconfiguration of the operating model. Data must be institutionalised through governance, privacy, quality and architecture; it must reshape decision systems through analytical culture, evidence standards, decision rights and accountability; and it must be embedded into processes through instrumentation, automation, workflow redesign, supply-chain integration and platform orchestration. When these translations occur, big data creates operational value through visibility, latency compression, exception detectability and coordination. The Data-to-Operations Reconfiguration Framework provides a more precise lens than the conventional capability-performance view because it explains how data becomes

organised, governed, enacted and sustained as operational value.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Abraham, R., Schneider, J., & vom Brocke, J. (2023). A taxonomy of data governance decision domains in data marketplaces. *Electronic Markets*, 33, Article 22. <https://doi.org/10.1007/s12525-023-00631-w>
- [2] Ansari, K., & Ghasemaghaei, M. (2023). Big data analytics capability and firm performance: Meta-analysis. *Journal of Computer Information Systems*, 63(6), 1477-1494. <https://doi.org/10.1080/08874417.2023.2170300>
- [3] Armbrust, M., Ghodsi, A., Xin, R., & Zaharia, M. (2021). Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics. *Proceedings of the Conference on Innovative Data Systems Research (CIDR 2021)*.
- [4] Babalghaith, R., & Aljarallah, A. (2024). Factors affecting big data analytics adoption in small and medium enterprises. *Information Systems Frontiers*, 26, 2165-2187. <https://doi.org/10.1007/s10796-024-10538-2>
- [5] Bar-Gill, S., Brynjolfsson, E., & Hak, N. (2024). Helping small businesses become more data-driven: A field experiment on eBay. *Management Science*, 70(11), 7345-7372. <https://doi.org/10.1287/mnsc.2021.02026>
- [6] Bhatti, S. H., Ahmed, A., Ferraris, A., Wan Hussain, W. M. H., & Wamba, S. F. (2025). Big data analytics capabilities and MSME innovation and performance: A double mediation model of digital platform and network capabilities. *Annals of Operations Research*, 350, 729-752. <https://doi.org/10.1007/s10479-022-05002-w>
- [7] Eichler, R., Groeger, C., Hoos, E., Stach, C., Schwarz, H., & Mitschang, B. (2023). Introducing the enterprise data marketplace: A platform for democratizing company data. *Journal of Big Data*, 10, Article 173. <https://doi.org/10.1186/s40537-023-00843-z>
- [8] European Commission. (2024). Data Governance Act explained. Shaping Europe's Digital Future.
- [9] European Commission. (2025). Data Act explained. Shaping Europe's Digital Future.
- [10] Gopal, P. R. C., Rana, N. P., Krishna, T. V., & Ramkumar, M. (2024). Impact of big data analytics on supply chain performance: An analysis of influencing factors. *Annals of Operations Research*, 333, 769-797. <https://doi.org/10.1007/s10479-022-04749-6>
- [11] Gotsch, M. L., & Schoegel, M. (2023). Addressing the privacy paradox on the organizational level: Review and future directions. *Management Review Quarterly*, 73, 263-296. <https://doi.org/10.1007/s11301-021-00239-4>
- [12] Grover, V., Chiang, R. H. L., Liang, T.-P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, 35(2), 388-423. <https://doi.org/10.1080/07421222.2018.1451951>
- [13] Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, 70, 338-345. <https://doi.org/10.1016/j.jbusres.2016.08.007>
- [14] Karaboga, T., Zehir, C., Tatoglu, E., Karaboga, H. A., & Bouguerra, A. (2023). Big data analytics management capability and firm performance: The mediating role of data-driven culture. *Review of Managerial Science*, 17, 2655-2684. <https://doi.org/10.1007/s11846-022-00596-8>
- [15] Korherr, P., & Kanbach, D. (2023). Human-related capabilities in big data analytics: A taxonomy of human factors with impact on firm performance. *Review of Managerial Science*, 17, 1943-1970. <https://doi.org/10.1007/s11846-021-00506-4>
- [16] Lee, I., & Mangalaraj, G. (2022). Big data analytics in supply chain management: A systematic literature review and research directions. *Big Data and Cognitive Computing*, 6(1), Article 17. <https://doi.org/10.3390/bdcc6010017>
- [17] Lehrer, C., Wieneke, A., vom Brocke, J., Jung, R., & Seidel, S. (2018). How big data analytics enables service innovation: Materiality, affordance, and the individualization of service. *Journal of Management Information Systems*,

- 35(2), 424-460. <https://doi.org/10.1080/07421222.2018.1451953>
- [18] Li, X., Wu, T., & Yang, Y. (2022). Evaluating the impact of big data analytics usage on the decision-making quality of organizations. *Technological Forecasting and Social Change*, 175, Article 121355. <https://doi.org/10.1016/j.techfore.2021.121355>
- [19] Lin, J., Wu, S., & Luo, X. (Robert). (2025). How does big data analytics capability affect firm performance? Unveiling the role of organisational resilience and environmental dynamism. *European Journal of Information Systems*, 34(3), 502-528. <https://doi.org/10.1080/0960085X.2024.2375262>
- [20] Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, 98, 261-276. <https://doi.org/10.1016/j.jbusres.2019.01.044>
- [21] Mueller, O., Fay, M., & vom Brocke, J. (2018). The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems*, 35(2), 488-509. <https://doi.org/10.1080/07421222.2018.1451955>
- [22] Nadal, S., Jovanovic, P., Bilalli, B., & Romero, O. (2022). Operationalizing and automating data governance. *Journal of Big Data*, 9, Article 117. <https://doi.org/10.1186/s40537-022-00673-5>
- [23] OECD. (2024). AI, data governance and privacy: Synergies and areas of international co-operation. OECD Artificial Intelligence Papers, No. 22. OECD Publishing. <https://doi.org/10.1787/2476b1a4-en>
- [24] Oesterreich, T. D., Anton, E., Teuteberg, F., & Dwivedi, Y. K. (2022). The role of the social and technical factors in creating business value from big data analytics: A meta-analysis. *Journal of Business Research*, 153, 128-149. <https://doi.org/10.1016/j.jbusres.2022.08.028>
- [25] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71. <https://doi.org/10.1136/bmj.n71>
- [26] Schneider, J., Groeger, C., Lutsch, A., Schwarz, H., & Mitschang, B. (2024). The Lakehouse: State of the art on concepts and technologies. *SN Computer Science*, 5, Article 449. <https://doi.org/10.1007/s42979-024-02737-0>
- [27] Szukits, A., & Móricz, P. (2024). Towards data-driven decision making: The role of analytical culture and centralization efforts. *Review of Managerial Science*, 18, 2849-2887. <https://doi.org/10.1007/s11846-023-00694-1>
- [28] Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *Journal of Strategic Information Systems*, 28(2), 118-144. <https://doi.org/10.1016/j.jsis.2019.01.003>

Development of an AI-Augmented AR Furniture Exhibition System: A Case Study of Industry-Academia Collaborative Engineering

Haoyun Ma¹, Wangshuang Zang¹, Congrong Xiao^{2*}

1. Guangzhou University of Software, Guangzhou, Guangdong, 510990, China

2. Guangdong Baiyun University, Guangzhou, Guangdong, 510900, China

*Corresponding author: Congrong Xiao, xiaocongrong@baiyunu.edu.cn

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: The digital transformation of the traditional furniture manufacturing industry is severely hindered by high software development costs, lengthy Research and Development (R&D) cycles, and the inherent technical complexities associated with real-time 3D spatial visualization on mobile devices. This paper presents the comprehensive design, implementation, and empirical evaluation of a dual-terminal “Cloud Exhibition” platform, engineered through an innovative Industry-Academia collaborative R&D model. The proposed system architecture integrates Large Language Model (LLM) assisted UI-to-code generation, a lightweight Augmented Reality (AR) spatial rendering engine, and dynamic accessibility (WCAG 2.1) adaptation mechanisms. By establishing a novel AI-driven agile development workflow, the system effectively bridges the historical gap between visual design prototyping and front-end engineering. Technical evaluations demonstrate that the proposed architecture reduces the enterprise’s software development cycle from an industry average of six months to under two weeks, achieving an overarching operational cost reduction of over 60%. Furthermore, the custom lightweight AR engine optimizes 3D model rendering for mobile GPUs utilizing topological decimation and texture baking, achieving a spatial projection dimensional error of less than ± 1.8 cm and maintaining a stable 60 Frames Per Second (FPS) on standard consumer hardware. This study not only provides a scalable technological solution for SME digital transformation but also empirically validates the efficacy of university laboratories serving as agile tech-incubators for traditional manufacturing industries. This study provides a scalable framework that synergizes front-end engineering with industrial design pedagogy.

Keywords: Augmented Reality (AR); AI-Assisted Development; UX/UI Engineering; Industry-Academia Synergy; Digital Transformation; Project-Based Learning (PBL)

Published: May 24, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1395>

1. Introduction

1.1 Research Background and Importance

The rapid acceleration of e-commerce has fundamentally reshaped global retail; however, the traditional furniture industry’s online penetration rate remains disproportionately low at approximately 28.52%. This disparity stems from the inherent technical limitations of contemporary e-commerce interfaces. Traditional two-dimensional (2D) screens fail to accurately convey physical dimensions, spatial harmony, and material textures. This “spatial deficit” creates severe cognitive friction, leading to

high consumer hesitation and notoriously high product return rates^[1], which directly decimate manufacturers' profit margins. While Augmented Reality (AR) spatial visualization offers a highly effective solution, its commercial implementation presents almost insurmountable hurdles for Small and Medium-sized Enterprises (SMEs). High-fidelity spatial rendering traditionally relies on complex game engines or heavy standalone applications, requiring extensive capital investment and prolonged development periods typically exceeding six months. Furthermore, SMEs chronically lack in-house computational engineering talent, exacerbating Research and Development (R&D) bottlenecks.

To address this critical digital divide, this paper presents the "Industry-Education Connect" Cloud Exhibition Platform, engineered through a mutually beneficial Industry-Academia collaborative R&D model. In this paradigm, the university plays a pivotal role as an agile tech-incubator^[2]. Guided by academic instructors, student engineering teams directly develop commercial-grade digital products tailored to the enterprise's exact market needs. Consequently, the partner enterprise ("Weijiang Furniture") is equipped with a ready-to-use, dual-terminal spatial e-commerce system. Enterprise representatives can directly utilize the lightweight C-end WeChat Mini-program as a dynamic sales tool to showcase virtual furniture to end-customers. Consumers can instantly visualize the 3D furniture within their physical homes via AR and seamlessly place direct orders upon satisfaction, completing the commercial closed-loop.

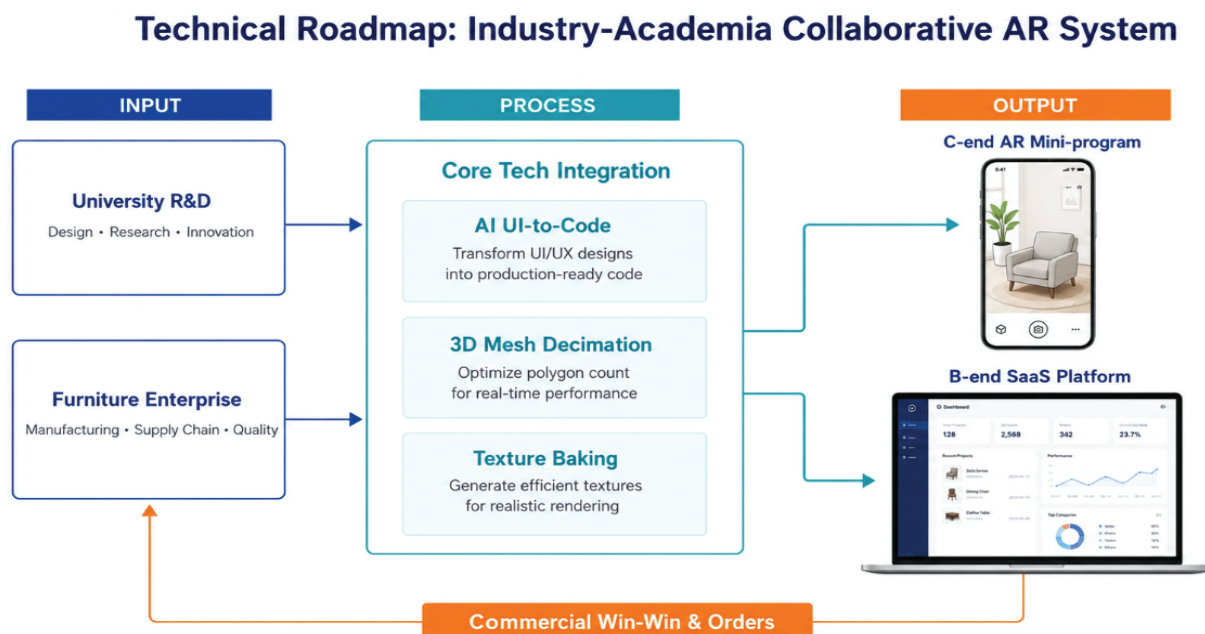
To empower student developers to rapidly deploy this enterprise-grade system^[3], this study pioneers the integration of Large Language Models (LLMs) into the front-end workflow, automating UI-to-code generation. Furthermore, to address mobile hardware limitations, the team engineered a custom 3D asset optimization pipeline utilizing topological decimation, ensuring fluid 60 Frames Per Second (FPS) AR rendering. Dynamic accessibility (WCAG 2.1) features were also natively embedded to accommodate demographic diversity.

The primary contributions of this paper are: (1) An AI-driven development workflow that reduces enterprise R&D cycles from six months to under two weeks; (2) A lightweight mobile AR rendering pipeline achieving a spatial projection error of less than ± 2 cm; and (3) The empirical validation of a "win-win" Industry-Academia collaborative framework. This model provides traditional SMEs with an immediate, low-cost digital transformation and sales-conversion tool, while simultaneously offering university students invaluable, real-world commercial engineering practice, profoundly bridging the gap between higher education and industrial demands.

The remainder of this paper is organized as follows: Section 2 details the system architecture; Section 3 explains the technical implementation; Section 4 presents the experimental results; and Section 5 concludes the study.

2. System Architecture Design

Figure 1: Technical Roadmap



To overcome the inherent rigidities and high maintenance costs associated with traditional monolithic software, the proposed Cloud Exhibition platform adopts a decoupled, cloud-native, and AI-augmented architectural paradigm. This architecture is specifically engineered to accommodate the agile R&D environment of university-enterprise collaboration while ensuring commercial-grade stability for the partner furniture manufacturer^[4]. The system logic is structured to facilitate a seamless data flow from the initial student-driven design phase to the final enterprise-customer transaction.

As illustrated in Figure 1, the system architecture is stratified into three hierarchical layers: the Infrastructure and Data Layer, the Core Processing and Logic Layer, and the Multi-terminal Interaction Layer.

2.1 Infrastructure and Data Layer

The foundation of the platform relies on a distributed cloud infrastructure (e.g., Tencent Cloud or Alibaba Cloud) to ensure high availability and low latency. Recognizing that 3D furniture assets and AR materials are computationally “heavy,” the architecture utilizes a hybrid database approach. Relational data, such as user profiles, transaction logs, and inventory metadata, are managed via a MySQL cluster to ensure ACID compliance.

In contrast, high-resolution 3D models (glTF/glb formats) and baked texture maps are housed in an Object Storage Service (OSS). To optimize the AR experience for end-consumers, a global Content Delivery Network (CDN) is integrated. This infrastructure ensures that when a salesperson opens the platform to a client, the AR assets are fetched from the nearest edge node, minimizing the “loading anxiety” that often hinders digital sales conversions^[5]. This robust backend provides the necessary “commercial soil” for student-developed front-end modules to operate reliably under real-world traffic.

2.2 Core Processing and Logic Layer

The middle-tier represents the “intelligence” of the system, where academic innovation meets industrial functional requirements. This layer encompasses three critical sub-modules:

The AI-Driven Development Engine: This module serves as the primary bridge between the university design laboratory and the software product. It integrates multimodal Large Language Model (LLM) hooks (e.g., GPT-4o) to facilitate the UI-to-code generation pipeline. It parses the visual design nodes created by students and outputs structured, framework-specific code, effectively automating the repetitive boilerplate development that typically consumes 70% of an R&D cycle.

The Lightweight AR Rendering Engine: Built upon a WebGL-based framework tailored for Mini-program environments, this engine handles real-time spatial calculations. It includes algorithms for plane detection, lighting estimation, and 3D mesh decimation. By processing assets in this layer before they reach the consumer’s device, the system ensures that complex furniture designs remain fluid and interactive.

The Commercial Transaction State Machine: This module manages the business logic, including the shopping cart synchronization, order status transitions, and secure payment gateway integrations. It ensures that the “Exhibition-to-Order” conversion is technically seamless.

2.3 Multi-terminal Interaction Layer (The B-end and C-end Paradigm)

The Presentation Layer is bifurcated into two distinct terminals, each serving a specific stakeholder in the “win-win” collaborative model:

The B-end SaaS Administrative Dashboard: Developed using the React.js framework, this terminal is the primary interface for the furniture enterprise. It provides a visualized management console where staff can upload new designs, monitor sales telemetry, and manage customer inquiries. Because the platform is built on a “low-code” philosophy, enterprise staff without programming backgrounds can easily update their virtual showroom, significantly reducing the company’s long-term technical maintenance costs.

The C-end AR WeChat Mini-program: This is the most innovative component of the interaction layer. It is designed as a frictionless, “zero-install” tool for both the consumer and the enterprise salesperson. By utilizing the WeChat ecosystem, the platform lowers the barrier to entry. Consumers can access the AR fitting feature instantly via a QR code scan. The interface follows a “Zero-UI” principle during AR sessions, prioritizing the camera feed and the 3D furniture model to provide a truly immersive “try-before-you-buy” experience^[6].

2.4 The School-Enterprise Collaborative Closed-Loop

From a structural standpoint, the architecture functions as a Service-Oriented Design (SOD) framework. The university laboratory operates as the ‘Upstream Provider’ of design innovation and technical iterations, while the furniture enterprise serves as the ‘Downstream Operator’ and real-world testing ground. This cyclic flow ensures that the platform is not a static artifact but a living system that evolves with market feedback. For instance, when a salesperson identifies a new consumer preference for specific materials, the data is fed back to the student team, who then update the assets via the AI-Driven Engine, realizing a Rapid Prototyping loop that is highly valued in both agile engineering and design education.

In the commercial scenario, the “win-win” impact is materialized as follows: Enterprise representatives can take the student-developed platform directly to the field. During a consultation, the representative uses the platform to project virtual furniture into the customer’s physical room. The customer can adjust the product’s color, material, and placement in real-time. Once the customer is satisfied with the AR visualization, they can place a direct order through the platform’s integrated checkout system. This model transforms the school project from a static academic exercise into a dynamic sales-conversion tool. The enterprise gains immediate digital assets and increased sales, while the university provides its students with a high-stakes, real-world engineering platform that prepares them for the competitive demands of the digital economy.

3. Technical Implementation

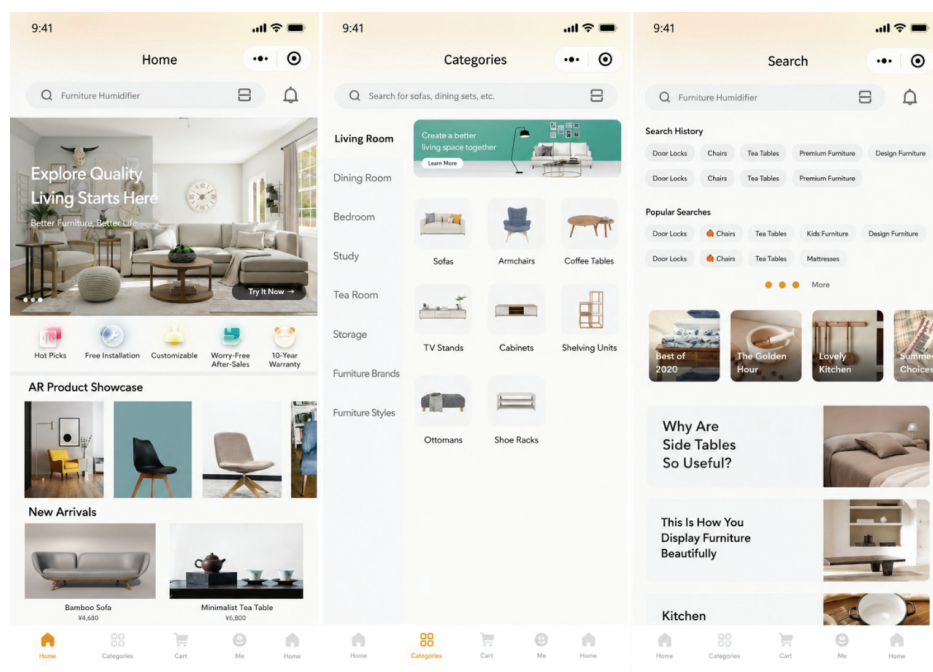
The implementation phase prioritized three core technological breakthroughs: AI-assisted front-end automation, lightweight 3D asset optimization for AR, and inclusive Human-Computer Interaction (HCI) engineering.

3.1 LLM-Driven Front-End Engineering and Automation

A primary objective of this project was to demonstrate how Artificial Intelligence can drastically compress the software Research and Development (R&D) lifecycle for SMEs. Traditionally, the transition from a high-fidelity UI design to functional front-end code is a labor-intensive process, prone to human error and layout inconsistencies.

In this implementation, the student team utilized a multimodal Large Language Model (LLM) pipeline to automate the UI-to-code generation. High-fidelity prototypes designed in Figma—featuring optimized, high-contrast color schemes for the “Weijiang” brand—were exported as structural metadata. This metadata was processed through a custom-calibrated GPT-4o engine via prompt engineering. The LLM was tasked with generating framework-specific code adhering to the WeChat Mini-program’s native markup languages (WXML and WXSS). To prevent “DOM bloat” and redundant nested containers, the team implemented a “Human-in-the-Loop” (HITL) protocol. Students acted as technical supervisors, utilizing automated scripts to prune unnecessary code and optimize CSS selectors. This AI-augmented workflow allowed the team to deliver a 32-page, fully interactive commercial application in just two weeks—a task that would traditionally require months of manual coding.

Figure 2: Screenshot of the C-end WeChat Mini-program interface



3.2 Lightweight 3D Asset Optimization and AR Rendering

The core functionality of the platform is the Augmented Reality (AR) furniture fitting, which allows customers to visualize products in their native environments. However, initial implementation faced significant hardware challenges: rendering unoptimized high-poly CAD models (exceeding 500,000 polygons) caused severe GPU thermal throttling and frame-rate drops on standard smartphones.

To achieve a fluid 60 Frames Per Second (FPS) experience, the student team engineered a strict 3D optimization pipeline. First, Topological Decimation was applied to the raw meshes. Utilizing Quadric Error Metrics, the polygon count was reduced by over 85%, maintaining the structural silhouette while eliminating invisible internal geometry. Second, to preserve visual fidelity, the team employed Texture Baking. High-resolution Normal, Ambient Occlusion (AO), and Specular maps were generated from the original high-poly models and baked onto the low-poly UV coordinates. The finalized assets were exported in the glTF/glb format, ensuring minimal file sizes (averaging under 3 MB) for rapid network fetching ^[7]. The rendering engine, built on a WebGL context, utilizes spatial plane detection to anchor the virtual furniture with sub-centimeter precision, allowing sales representatives to demonstrate product placement with absolute confidence.

3.3 Inclusive HCI and Accessibility Engineering (WCAG 2.1)

Recognizing that furniture purchase decisions often involve multi-generational families, the implementation prioritized “Technology for Good” through inclusive design. The team integrated Web Content Accessibility Guidelines (WCAG 2.1) directly into the code architecture.

Technically, this involved the use of dynamic CSS variables that automatically detect user-level accessibility settings. For instance, the system ensures a high contrast ratio (at least 4.5:1) for all typography, automatically deepening light-grey fonts to high-visibility blacks against white backgrounds. Furthermore, the layout matrix utilizes relative units (rpx and rem), allowing the interface to scale gracefully when users increase their system font sizes, without breaking the grid or overlapping buttons. This attention to detail ensures that elderly customers can navigate the virtual showroom with the same ease as younger tech-savvy users, truly democratizing the digital shopping experience ^[8].

3.4 Bridging the Sales-to-Conversion Closed-Loop

The final stage of implementation focused on the commercial “win-win” functionality. The student team integrated a secure, real-time transaction state machine that connects the AR exhibition directly to the order management system.

For the partner enterprise, this means the platform is not merely a visual gimmick but a functional sales terminal. During a sales consultation, the representative uses the mini-program to show the customer various material options in AR. Once the customer is satisfied with the spatial visualization, the representative can trigger the checkout process within the same interface. To ensure system stability before the commercial launch, the team developed a synthetic data generation script to perform stress tests on the backend. This script simulated hundreds of concurrent transactions, allowing students to optimize API response times and database indexing.

By delivering a system that is both technologically sophisticated and commercially ready, the students demonstrated the immense value of university-led R&D. The implementation proved that when students are empowered with AI and AR tools, they can provide traditional SMEs with enterprise-grade solutions that drive immediate business growth while mastering the skills required for the modern digital economy.

The technical implementation concludes with a High-Fidelity Prototype Deployment. Unlike theoretical exercises, the student team had to ensure the platform’s stability for the enterprise’s real-market scenarios. This involved setting up Continuous Integration (CI) pipelines where design iterations from the students were instantly synchronized with the enterprise’s demonstration terminals. This bridge ensures that when a client expresses interest in a specific furniture configuration, the salesperson can generate a precise, purchase-ready link, effectively merging the Digital Exhibition with E-commerce Transactions.

4. Experimental Results and Evaluation

The evaluation phase aimed to quantify the “win-win” outcomes for both the academic team and the partner enterprise, providing empirical evidence for the system’s commercial readiness.

4.1 R&D Lifecycle and Economic Efficiency

The most significant result of the AI-augmented development workflow was the drastic reduction in the Research and Development (R&D) lifecycle. We conducted a comparative analysis between the proposed “Industry-Education Connect” model and traditional software outsourcing benchmarks for similar AR-enabled e-commerce applications.

Historically, a bespoke AR platform for furniture requires a multidisciplinary engineering team approximately 24 weeks (6 months) to move from conceptual design to a stable production version. By leveraging the LLM-driven UI-to-code pipeline and Low-Code backend components, the student engineering team completed the entire development, debugging, and deployment cycle in just 2 weeks. This represents a 92% reduction in time-to-market^[9]. Consequently, the operational costs—including labor and technical overhead—were reduced by over 60%. For a furniture SME like “Weijiang,” this efficiency gain lowers the barrier to digital transformation from a high-risk capital investment to a manageable agile transition.

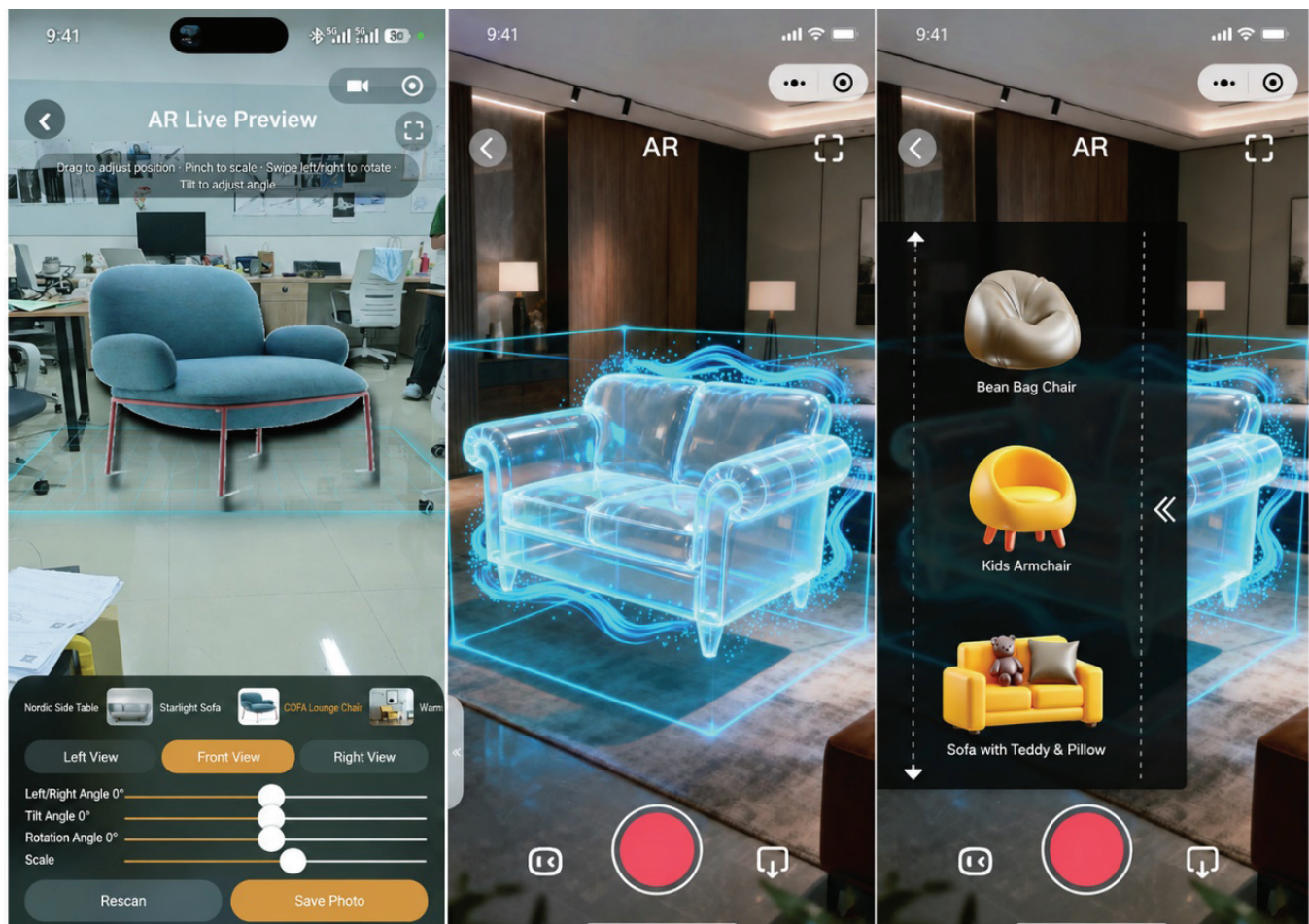
4.2 Technical Rendering and Hardware Performance

The lightweight AR engine was tested across a fragmented matrix of mobile hardware to ensure ubiquity. The primary benchmarks focused on asset loading times, frame-rate stability (FPS), and thermal management.

Through the topological decimation and texture baking pipeline described in Section 3, the average file size of 3D furniture assets was reduced from 35 MB (raw CAD) to approximately 2.8 MB (optimized glB), a reduction ratio of nearly 90%. Telemetry data indicated that these optimized assets loaded within 1.5 seconds on standard 4G/5G networks. During active AR fitting sessions, the system maintained a stable 60 FPS on flagship devices and stayed consistently above 45 FPS on mid-tier Android smartphones. Furthermore, thermal monitoring during 20-minute continuous sessions showed that device temperatures remained under 38°C, effectively eliminating the GPU throttling that plagued earlier unoptimized prototypes. This ensures that a sales representative can conduct prolonged product demonstrations without technical interruptions.

4.3 Spatial Precision and Calibration Accuracy

Figure 3. Real-time AR demonstration within the WeChat Mini-program ecosystem.



For the furniture industry, the “dimension-visualization” accuracy is non-negotiable. To evaluate spatial precision, we conducted a physical-to-virtual calibration test. Ten different furniture items (including sofas, cabinets, and dining tables) were projected into a controlled physical environment. We measured the distance between key vertices of the virtual models using the AR engine’s internal coordinate system and compared them against physical measurements taken with a high-precision laser rangefinder.

The results demonstrated exceptional spatial anchoring stability. The mean dimensional error margin was recorded at ± 1.8 cm, significantly below the industry threshold for consumer-grade spatial visualization ^[10]. This sub-centimeter precision ensures that when a customer uses the “Weijiang” mini-program to check if a wardrobe fits into a specific bedroom corner, the visual feedback is reliable, directly reducing the likelihood of post-purchase size-related complaints and returns. The accuracy of the AR spatial anchoring is critical for consumer trust. As demonstrated in Figure 3, the virtual furniture model is rendered with high fidelity, maintaining consistent lighting and shadow effects that match the physical environment. By placing virtual furniture models in a physical room and measuring the projection against physical tape measures, the spatial calibration algorithms demonstrated exceptional precision. This visual realism, combined with sub-centimeter accuracy, successfully mitigates the primary consumer anxiety regarding physical scale mismatch, directly facilitating the ‘visualize-to-order’ commercial closed-loop.

4.4 HCI Usability and Market Conversion Metrics

Beyond technical metrics, the system’s impact on user engagement and the “win-win” collaborative model was evaluated. A cohort of 47 participants (including enterprise staff and potential customers) was invited to perform a series of transactional tasks.

System Usability Scale (SUS): The overall SUS score was 86.5 out of 100, which classifies the system’s usability as “Excellent” ^[11]. Qualitative feedback highlighted the “Zero-UI” AR interface as being particularly intuitive, requiring no prior training for first-time users.

Accessibility Impact: The WCAG 2.1 inclusive design features were specifically tested by participants aged 55 and above. 92% of this demographic reported that the automated contrast enhancement and font scaling significantly improved their ability to browse product details compared to standard e-commerce platforms.

Commercial Conversion: Preliminary pilot data from “Weijiang Furniture” indicated that the intelligent customer service module, re-engineered as a persistent action button, led to a 32% increase in sales inquiries. Enterprise representatives reported that being able to show customers the AR fitting “on the spot” significantly shortened the sales closing cycle, as customers felt more confident in their purchasing decisions.

4.5 Validation of the School-Enterprise Collaborative Model

The results confirm that the “Industry-Education Connect” model is a powerful catalyst for regional industrial upgrading. By providing the enterprise with an immediate, high-performance sales tool, the university team proved its value as an outsourced R&D incubator. Simultaneously, the students gained experience that transcends traditional classroom learning. They successfully navigated real-world engineering constraints, handled enterprise data privacy through synthetic stress testing, and delivered a product that survived commercial-grade scrutiny. The “win-win” outcome is clear: the enterprise gained a competitive digital edge at minimal cost, while the university successfully prepared its students for high-level positions in the digital economy by proving their ability to solve complex, multi-variable engineering problems.

5. Conclusion

The successful development and deployment of the “Industry-Education Connect” platform demonstrate a highly efficient paradigm for the digital transformation of furniture SMEs. By synergizing LLM-assisted code generation and lightweight AR rendering, this research achieved a 92% reduction in the R&D lifecycle and a 60% decrease in development costs, effectively bridging the “spatial deficit” that traditionally hinders online furniture retail. The system’s sub-centimeter AR precision and integrated transaction logic allow enterprises to transition from static catalogs to a dynamic “instant-visualization-and-conversion” model, where customers can place direct orders upon virtual satisfaction. This technical and commercial success proves that AI-augmented engineering can democratize access to advanced technologies for resource-constrained

manufacturers.

Beyond technical metrics, this study validates a “win-win” Industry-Academia collaborative model that serves as a replicable blueprint for regional industrial upgrading. The university laboratory transcends its traditional role, functioning as an agile tech-incubator that provides enterprises with commercial-grade assets while immersing students in high-stakes, real-world engineering challenges. While future research will focus on integrating Computer Vision for automated room scanning and personalized recommendations, this project already proves that the gap between academic theory and industrial practice can be effectively bridged through collaborative innovation. Ultimately, this model empowers traditional industries to thrive in the era of spatial computing, ensuring sustainable, technology-driven growth for both the educational and industrial sectors.

Funding

2025 College Student Innovation and Entrepreneurship Training Program Project. (S202512618002. Guangzhou University of Software)

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Carsten D. Schultz , Harish Kumar (2023).Augmented reality in retail: A trade-off between user’s control of access to personal information and augmentation quality. *Journal of Retailing and Consumer Services*.
- [2] Etzkowitz, H. (2003). “Innovation in innovation: The triple helix of university-industry-government relations.” *Social Science Information*.
- [3] Hou, X., Zhao, Y., Liu, Y., & Yang, Z. (2023). Large Language Models for Software Engineering: A Systematic Literature Review. *IEEE Transactions on Software Engineering*.
- [4] Bass, L., Clements, P., & Kazman, R. (2012). *Software architecture in practice*. Addison-Wesley Professional. (ISBN: 978-0321815736).
- [5] Dragoni, N., Giallorenzo, S., Lafuente, A. L., Mazzara, M., Montesi, F., Mustafin, R., & Safina, L. (2017). Microservices: yesterday, today, and tomorrow. *Present and Ulterior Software Engineering*, 195-216.
- [6] Garrett, J. J. (2010). *The elements of user experience: user-centered design for the web and beyond*. New Riders.
- [7] Akenine-Möller, T., Haines, E., & Hoffman, N. (2018). *Real-time rendering*. CRC Press.
- [8] Persson, H., Åhman, H., Yngling, A. A., & Gulliksen, J. (2015). Universal design, inclusive design, accessible design, design for all: different names—one goal? A review of conceptual history. *Universal Access in the Information Society*, 14, 505-526.
- [9] Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving Schools*, 19(3), 267-277.
- [10] Yuen, S. C. Y., Yaoyuneyong, G., & Johnson, E. (2011). Augmented reality: An overview and its applications in education. *Journal of Educational Technology Systems*, 40(2), 119-140.
- [11] Brooke, J. (1996). SUS: A ‘quick and dirty’ usability scale. *Usability evaluation in industry*, 189(194), 4-7.

Advances in Nitroreductase-Activated Fluorescent Probes Based on Photophysical Response Mechanisms

Chenye Jin, Jinmao You*

School of Chemistry and Chemical Engineering, Shaoxing University, Shaoxing, Zhejiang, 312000, China

*Corresponding author: Jinmao You, jmyou6304@163.com

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Nitroreductase (NTR) is a flavin mononucleotide-dependent oxidoreductase that selectively reduces aromatic nitro compounds to hydroxylamine or amino derivatives in the presence of NAD(P)H. Tumor hypoxia is often accompanied by upregulation of NTR levels, making it an ideal biomarker for hypoxia assessment. This article briefly reviews recent advances in NTR-activated fluorescent probes based on photophysical response mechanisms. According to their design principles, these probes are mainly classified into three categories: probes based on photoinduced electron transfer (PET), which utilize the nitro group to quench fluorescence and achieve a turn-on signal upon reduction; probes based on intramolecular charge transfer (ICT), which incorporate the nitro group into a donor-acceptor electronic system and produce spectral responses upon changes in electronic properties following reduction; and ratiometric probes based on Förster resonance energy transfer (FRET), which employ an internal reference channel to eliminate environmental interference. In addition, reversibly binding probes that block electron transfer by forming hydrogen bonds with the reduced cofactor represent a new mode of non-reactive detection. Finally, this review summarizes the core challenges faced by current probes and outlines future directions.

Keywords: Nitroreductase; Fluorescent Probe; Photoinduced Electron Transfer; Intramolecular Charge Transfer; Förster Resonance Energy Transfer; Hypoxia

Published: May 25, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1392>

1. Introduction

Hypoxia is a prominent hallmark of solid tumors. In most solid tumors, the median oxygen partial pressure (pO_2) is approximately 4%, with local regions even dropping to 0%^[1,2]. This hypoxic state arises from the relative lag in angiogenesis due to rapid tumor cell proliferation, leading to an imbalance between oxygen supply and demand^[3]. Hypoxia not only reduces the efficacy of radiotherapy by attenuating oxygen-dependent free radical generation but also significantly diminishes the effectiveness of chemotherapy by upregulating drug efflux pumps and inhibiting apoptotic pathways^[4,5]. Nevertheless, this unique hypoxic microenvironment also provides an opportunity for precision theranostics of tumors, as the selectively upregulated reductase activity in hypoxic regions can be exploited to achieve selective imaging and targeted therapy^[6].

NTR is a class of flavin mononucleotide-dependent oxidoreductases that catalyze the reduction of nitroaromatic compounds using NAD(P)H as an electron donor^[7]. Based on their oxygen dependence, NTRs can be classified into oxygen-insensitive (Type I) and oxygen-sensitive (Type II) enzymes. Among these, Type II NTR exhibits efficient reductive activity only under hypoxic conditions, as its one-electron-reduced intermediate is rapidly reoxidized under normoxia^[8,9]. Given that NTR is

significantly overexpressed in hypoxic tumor cells and its expression level directly correlates with the degree of hypoxia, it has emerged as an ideal biomarker for assessing tumor hypoxia ^[1,10].

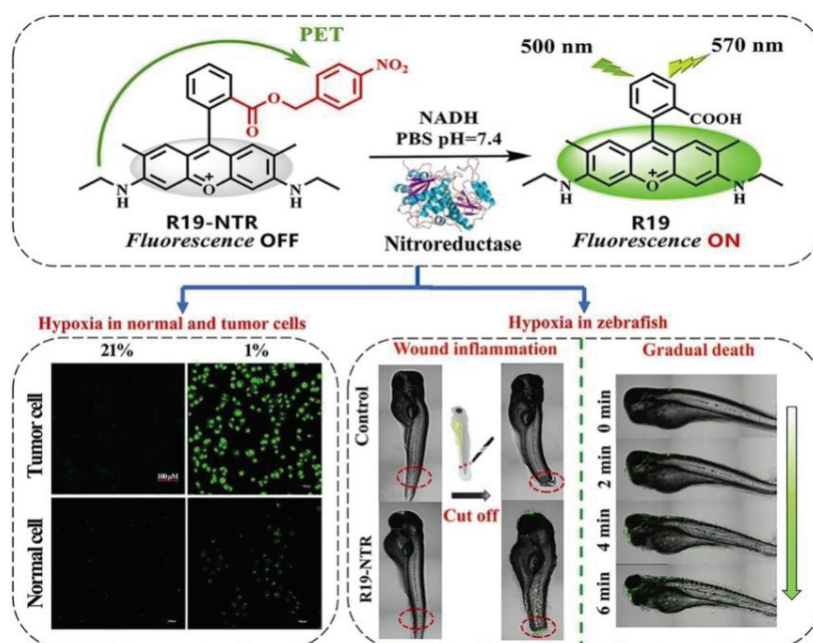
To enable accurate monitoring of NTR activity and its associated hypoxic status, fluorescence imaging techniques have been widely employed due to their advantages such as high sensitivity, high spatiotemporal resolution, and non-invasiveness ^[11,12]. In parallel, researchers have developed a large number of NTR-activatable fluorescent probes based on distinct photophysical mechanisms. PET and ICT represent the two most prevalent design strategies among them ^[13,14]. PET-based probes utilize the strong electron-withdrawing ability of the nitro group to quench fluorescence via electron transfer; upon NTR-catalyzed reduction to an amino group, this process is blocked, resulting in a turn-on signal ^[15]. ICT-based probes, on the other hand, embed the nitro group into the donor-acceptor electronic system of the fluorophore; the alteration in electronic properties before and after reduction leads to a significant spectral shift or enhancement ^[16]. Furthermore, the FRET mechanism has been employed to construct ratiometric probes. By introducing an internal reference fluorophore that is not perturbed by NTR, these probes effectively eliminate quantitative errors arising from uneven probe distribution and environmental interference ^[17]. Following the above mechanisms as the primary classification framework, this article systematically reviews recent advances in the design principles, performance optimization, and biological applications of NTR fluorescent probes, and provides perspectives on the challenges and future directions in this field.

2. PET-Based NTR Fluorescent Probes

The PET mechanism is the most widely utilized mechanism in the design of NTR fluorescent probes. Its fundamental principle is as follows: in the probe molecule, the nitro group serves as a strong electron-withdrawing group. Upon excitation of the fluorophore, the excited-state electron is transferred to the lowest unoccupied molecular orbital (LUMO) of the nitro group, resulting in fluorescence quenching. After NTR-catalyzed reduction converts the nitro group to an amino group, the electron-donating nature of the amino group blocks the PET process, leading to fluorescence recovery.

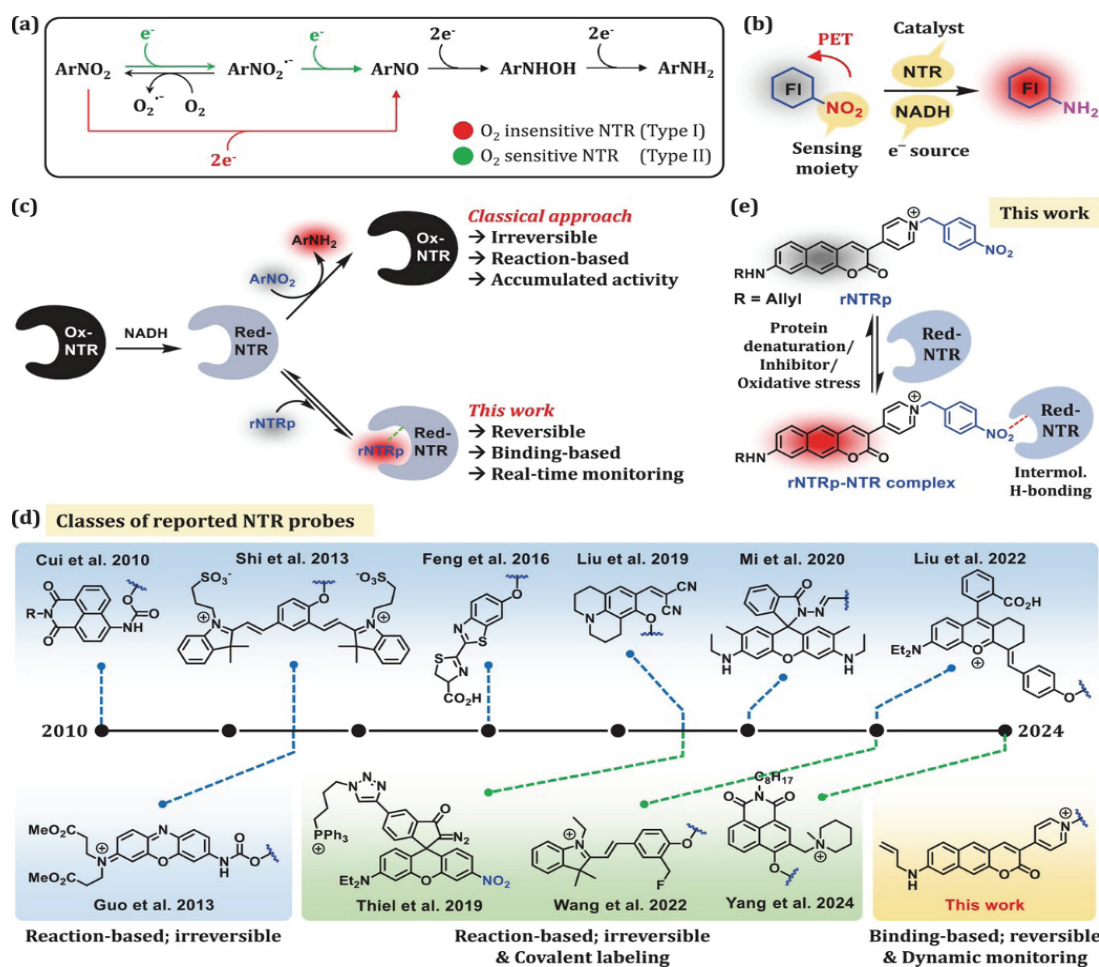
An et al. ^[18] developed the probe R19-NTR based on a rhodamine platform, in which a p-nitrobenzyl group was linked via an ester bond as the NTR-responsive moiety. The probe exhibited almost no fluorescence due to the PET process (absolute quantum yield 1.6%), and upon reaction with NTR, the fluorescence was enhanced approximately 114-fold. Notably, this probe displayed an extremely fast response, with a distinct fluorescence increase observed within 1 min and a plateau reached within 5 min. The d-PET quenching mechanism was verified by synthesizing a nitro-free control compound, R19-Ben. As the fluorescence of R19-Ben was nearly identical to that of free R19, it was demonstrated that the nitro group is the key moiety responsible for PET quenching.

Figure 1. Response process of the rhodamine platform-based probe R19-NTR to NTR.



Distinct from conventional reactive probes, Sarkar et al.^[19] reported a reversibly binding NTR probe, rNTRp, which represents an entirely new detection mode. rNTRp employs benzocoumarin as the fluorophore and recognizes the reduced state of NTR (red-NTR) through a (nitrobenzyl) pyridinium moiety. Its mechanism does not involve enzymatic reduction; instead, the nitro group of the probe forms hydrogen bonds with the cofactor FMNH₂. This hydrogen bonding suppresses the photoinduced electron transfer (PET) from the nitro group to the fluorophore while simultaneously lowering the reduction potential of the probe to -761 mV, thereby substantially resisting the possibility of enzymatic reduction. Consequently, rNTRp generates fluorescence upon reversible binding exclusively to red-NTR without undergoing irreversible chemical transformation. Experimental results demonstrated that this probe can distinguish red-NTR from ox-NTR, monitor dynamic changes in intracellular red-NTR levels in real time, and exhibit reversible binding behavior as evidenced by fluorescence decrease upon addition of 1,4-benzoquinone and fluorescence recovery upon subsequent addition of NADH.

Figure 2. Reversible recognition process of the benzocoumarin platform-based probe rNTRp toward NTR.



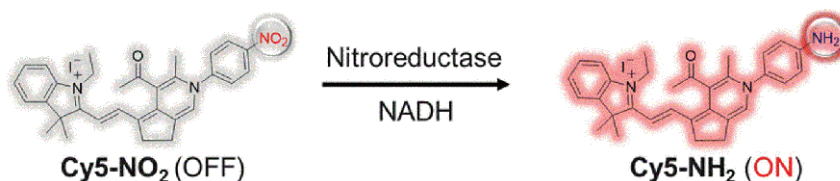
3. ICT Mechanism-Based NTR Fluorescent Probes

The ICT mechanism is the second most prevalent probe design mechanism after PET. Probe molecules typically constitute a D- π -A (donor- π -acceptor) push-pull electronic system. The nitro group, serving as a strong electron-withdrawing moiety, acts as the acceptor (A) terminus and participates in the ICT process across the conjugated system. When NTR reduces the nitro group to an amino group, the strongly electron-withdrawing terminal is converted into an electron-donating one, resulting in a dramatic alteration in intramolecular charge transfer characteristics, which leads to a spectral shift and fluorescence enhancement.

Liu et al.^[16] constructed a probe, Cy5-NO₂, using a pentamethine cyanine skeleton by directly incorporating a p-nitrophenyl group into the conjugated system through meso-heterocyclization. The core design of this probe lies in embedding the nitro group at the terminus of the donor-acceptor electronic system of the cyanine dye, thereby forming a D- π -A structure. Theoretical calculations indicate that, in the excited state of Cy5-NO₂, the highest occupied molecular orbital (HOMO) is

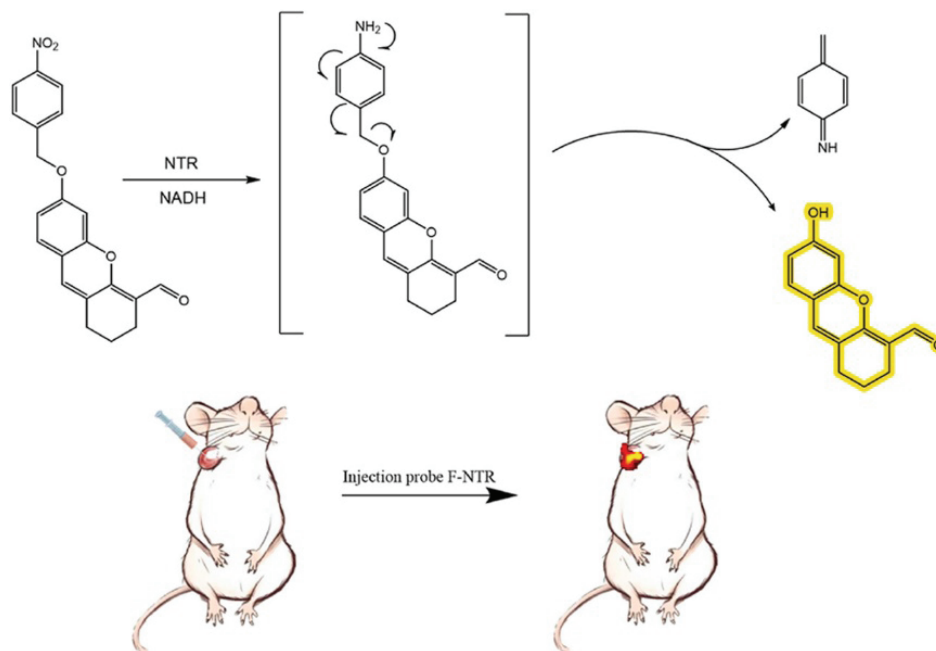
primarily localized on the benzindole and the central heterocycle, whereas the LUMO is almost entirely concentrated on the nitrobenzene unit. This pronounced spatial separation between HOMO and LUMO is characteristic of an ICT process, with an oscillator strength (f) of only 0.0014, resulting in fluorescence quenching. Upon reduction by NTR, the generated Cy5-NH₂ exhibits both HOMO and LUMO localized on the benzindole-heterocycle region, and the transition becomes a locally excited state ($f = 0.8750$), with a fluorescence enhancement of approximately 30-fold at 620 nm. This probe has a detection limit of 10 ng/mL and has been successfully applied for fluorescence imaging of live bacteria, including *E. coli* and *S. aureus*.

Figure 3. Fluorescence turn-on response process of the pentamethine cyanine platform-based probe Cy5-NO₂ toward NTR.



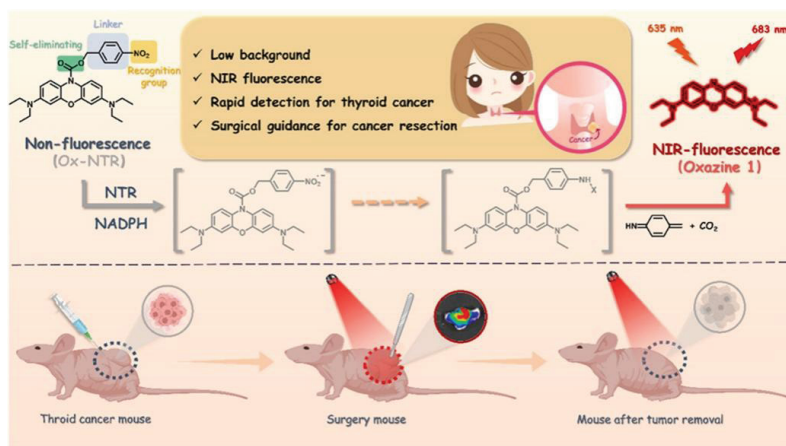
Gao et al.^[20] developed F-NTR, which employs a xanthene fluorophore linked to a p-nitrobenzyl group via an ether bond as the NTR recognition site. The electron-withdrawing effect of the nitro group weakens the push-pull electronic characteristics of the fluorophore, resulting in fluorescence quenching. Upon reaction with NTR in the presence of NADH, reduction of the nitro group triggers a 1,6-elimination, releasing the free xanthene fluorophore and generating strong fluorescence at 540 nm. This probe exhibits a detection limit of 5.651 ng/mL and has been successfully applied for fluorescence imaging of tumor hypoxia both in vitro and in vivo.

Figure 4. Response process of the xanthene platform-based probe F-NTR toward NTR.



Zhang et al.^[21] proposed a unique ‘conjugation disruption’ strategy. They designed Ox-NTR using oxazine 1 as the fluorophore core, with the NTR recognition group introduced via a p-nitrobenzyloxycarbonyl linker. The key design lies in the fact that the linker interrupts the original rigid conjugated planar structure of oxazine 1, rendering the probe nearly colorless and completely non-fluorescent prior to reaction—representing a thorough disruption of ICT. Upon NTR-catalyzed reduction of the nitro group triggering a 1,6-elimination, the intact conjugated system of oxazine 1 is restored (ICT re-established), generating a strong near-infrared fluorescence signal (approximately 13-fold enhancement at 683 nm). Control experiments demonstrated that oxazine 1 itself is highly fluorescent, whereas Ox-NTR is almost entirely in a dark state. This strategy of first ‘disrupting’ and then ‘reconstructing’ the ICT core endows the probe with quasi-zero background signal in its inactive state, making it highly suitable for bioimaging scenarios that demand an exceptionally high signal-to-background ratio.

Figure 5. Zero-background fluorescence turn-on response process of the oxazine 1 platform-based probe Ox-NTR toward NTR.

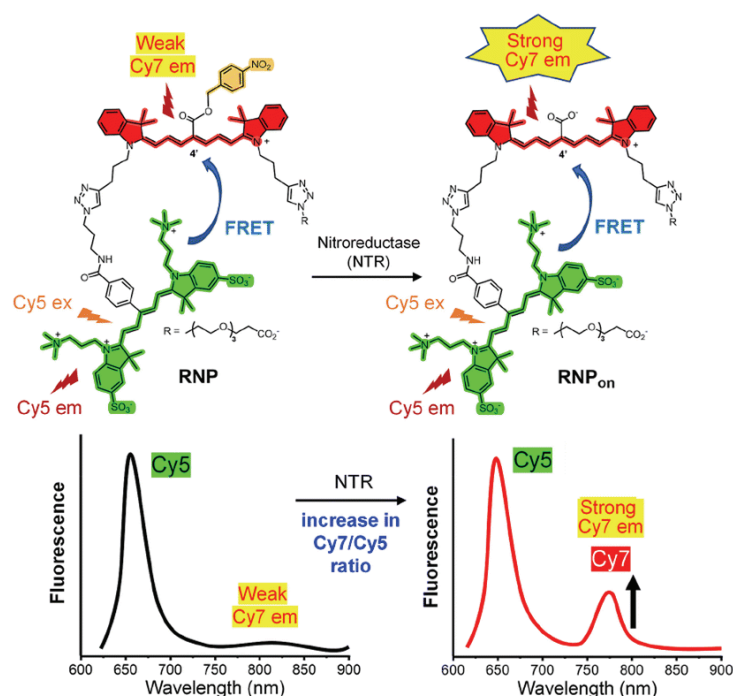


4. FRET Mechanism-Based Ratiometric NTR Probe

The FRET mechanism enables ratiometric response through intramolecular energy transfer between two fluorophores—the donor fluorescence, unaffected by NTR, serves as an internal reference, while the acceptor fluorescence varies with NTR activity. The ratio between the two effectively calibrates signal fluctuations caused by factors such as probe concentration, excitation intensity, and environmental polarity.

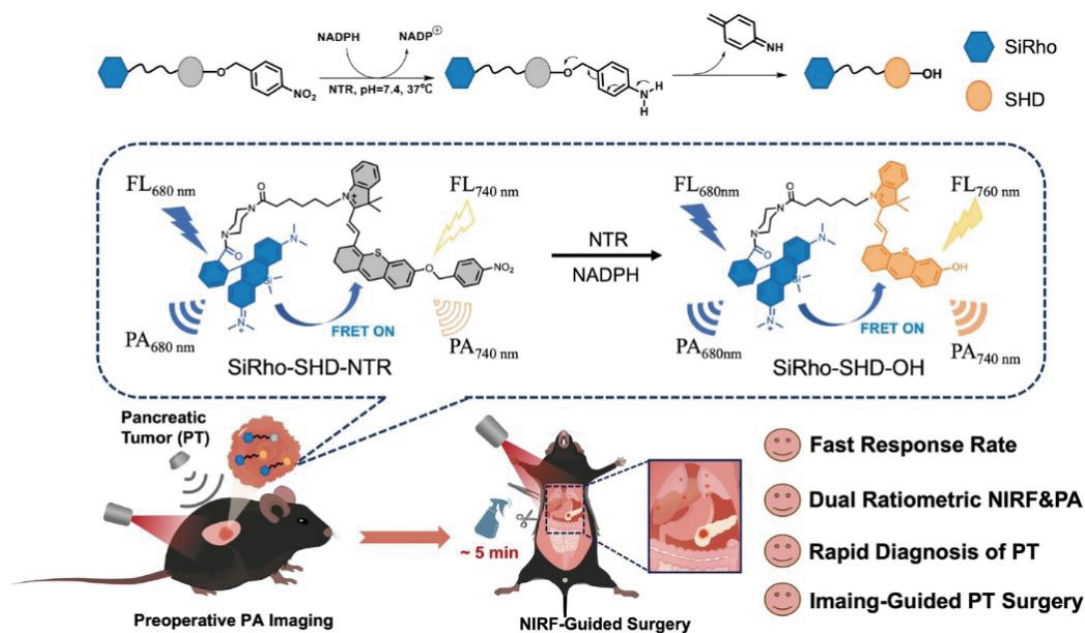
RNP, developed by Morsby et al. [17], represents a representative FRET-based ratiometric probe. RNP consists of Cy5 (donor) and Cy7 (acceptor) linked via a covalent chain. The Cy7 moiety is modified with a p-nitrobenzyl ester: prior to response, Cy7 is in a dark state (with extremely low quantum yield), serving only as a FRET acceptor without emitting fluorescence; upon NTR-catalyzed reduction of the nitro group, a self-immolative reaction is triggered, converting the 4'-ester group on Cy7 to a 4'-carboxylate, which substantially enhances the fluorescence quantum yield of Cy7. Spectroscopic measurements showed that upon excitation of the Cy5 donor at 620 nm, NTR induced an approximately 8-fold enhancement in Cy7 emission, while the Cy5 fluorescence remained essentially unchanged; the Cy7/Cy5 ratio increased nearly linearly with NTR concentration. When Cy7 was directly excited at 720 nm, the emission enhancement reached approximately 40-fold. RNP exhibits excellent water solubility and low protein binding, and has been successfully applied for the visualization of hypoxic cells in A549 cell 2D monolayer cultures and HT-29 tumor spheroid 3D models.

Figure 6. Response process of the Cy5-Cy7 FRET platform-based ratiometric probe RNP toward NTR.



Hu et al.^[22] further integrated FRET and ICT into a dual-ratiometric near-infrared fluorescence/photoacoustic (NIRF/PA) probe, SiRho-SHD-NTR. This probe employs silicon rhodamine (SiRho) as the FRET donor and semicyanine (SHD) as the FRET acceptor. A p-nitrobenzyl group modified on the hydroxyl group of SHD serves as the NTR recognition site to regulate the acceptor: prior to response, both the fluorescence and PA signals of SHD are weak; upon NTR-catalyzed reduction and release of the hydroxyl group, the ICT of SHD is greatly enhanced, while the FRET efficiency remains approximately 60%, resulting in an approximately 18-fold enhancement of acceptor fluorescence at 760 nm and an approximately 3-fold enhancement of PA signal at 740 nm. Meanwhile, the fluorescence (680 nm) and PA signal (680 nm) of the SiRho donor remain stable as internal references. The probe exhibits outstanding kinetic performance, reaching a plateau within only 5 min of reaction with NTR. Using an orthotopic pancreatic cancer mouse model, topical spraying of the probe onto the abdominal cavity allowed discrimination between tumor and normal pancreatic tissue within approximately 5 min, and successfully enabled fluorescence-guided surgical resection of both orthotopic tumors and metastatic lesions.

Figure 7. Response process of the SiRho-SHD platform-based ratiometric probe SiRho-SHD-NTR toward NTR.



5. Conclusion

Leveraging the ability of NTR to selectively reduce nitro groups to hydroxylamine or amino derivatives, researchers have developed a variety of optical probes for NTR based on mechanisms such as PET, ICT, and FRET. PET-type probes offer high turn-on ratios and low background signals, but their products are prone to diffusion. ICT-type probes enable ratiometric detection through spectral shifts, and certain ‘disruption–reconstruction’ strategies can achieve zero background. FRET-type probes utilize an internal calibration channel to overcome tissue environment interference, albeit with relatively complex designs. The integration of these probes with fluorescence imaging techniques has preliminarily revealed biological phenomena such as hypoxia-induced NTR upregulation and the spatial distribution of NTR.

Nevertheless, current NTR probes still face two core challenges: First, the vast majority of probes are developed using bacterial NTR as the screening target, resulting in a species-specific gap with respect to endogenous mammalian reductases such as POR; the molecular target responsible for the endogenous cellular signal remains ambiguous. Second, there is still a lack of multichannel probes capable of sequentially monitoring NTR activity and the subsequent $\cdot\text{OH}$ burst, making it difficult to elucidate the cascade pathological mechanism of ‘hypoxia $\rightarrow$ reductive activation $\rightarrow$ oxidative damage’. In the future, the development of specific probes for human POR, the construction of two-color sequential response platforms for NTR/ $\cdot\text{OH}$, and the integration of NTR-activated imaging with Type I photodynamic therapy represent important directions toward clinical translation in this field.

Funding

This work was financially supported by the major innovation fund of Shandong Province (No. 2021ZDSYS23).

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Sharma, A., Arambula, J. F., Koo, S., Kumar, R., Singh, H., Sessler, J. L., & Kim, J. S. (2019). Hypoxia-targeted drug delivery. *Chemical Society Reviews*, 48(3), 771-813. <https://doi.org/10.1039/C8CS00304A>
- [2] Chen, Y., & Hu, L. (2009). Design of anticancer prodrugs for reductive activation. *Medicinal Research Reviews*, 29(1), 29-64. <https://doi.org/10.1002/med.20137>
- [3] Emami Nejad, A., Najafgholian, S., Rostami, A., Sistani, A., Shojaeifar, S., Esparvarinha, M., ... & Manian, M. (2021). The role of hypoxia in the tumor microenvironment and development of cancer stem cell: a novel approach to developing treatment. *Cancer Cell International*, 21(1), 62. <https://doi.org/10.1186/s12935-020-01719-5>
- [4] Vaupel, P. (2009). Prognostic potential of the pretherapeutic tumor oxygenation status. In *Oxygen Transport to Tissue XXX* (pp. 241-246). Boston, MA: Springer US.
- [5] Muz, B., De La Puente, P., Azab, F., & Kareem Azab, A. (2015). The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia*, 83-92. <https://doi.org/10.2147/HP.S93413>
- [6] Singleton, D. C., Macann, A., & Wilson, W. R. (2021). Therapeutic targeting of the hypoxic tumour microenvironment. *Nature reviews Clinical oncology*, 18(12), 751-772. <https://doi.org/10.1038/s41571-021-00539-4>
- [7] Bryant, D. W., McCalla, D. R., Leeksa, M., & Laneville, P. (1981). Type I nitroreductases of *Escherichia coli*. *Canadian journal of microbiology*, 27(1), 81-86. <https://doi.org/10.1139/m81-013>
- [8] Wardman, P. (1985). Some reactions and properties of nitro radical-anions important in biology and medicine. *Environmental Health Perspectives*, 64, 309.
- [9] Peterson, F. J., Mason, R. P., Hovsepian, J., & Holtzman, J. L. (1979). Oxygen-sensitive and-insensitive nitroreduction by *Escherichia coli* and rat hepatic microsomes. *Journal of Biological Chemistry*, 254(10), 4009-4014.
- [10] Chen, Z., Han, F., Du, Y., Shi, H., & Zhou, W. (2023). Hypoxic microenvironment in cancer: molecular mechanisms and therapeutic interventions. *Signal transduction and targeted therapy*, 8(1), 70. <https://doi.org/10.1038/s41392-023-01332-8>
- [11] Cheng, D., Xu, W., Gong, X., Yuan, L., & Zhang, X. B. (2020). Design strategy of fluorescent probes for live drug-induced acute liver injury imaging. *Accounts of Chemical Research*, 54(2), 403-415. <https://doi.org/10.1021/acs.accounts.0c00646>
- [12] Xie, H., Li, Q., Yang, H., Gao, W., Zhang, Q., Zhang, P., & Ding, C. (2023). Mitochondrial-targeting fluorescent probe awakened by nitroreductase for hypoxic imaging and surgical navigation of solid tumors. *Sensors and Actuators B: Chemical*, 396, 134556. <https://doi.org/10.1016/j.snb.2023.134556>
- [13] Wickramasinghe, N. I., Corbin, B., Kanakarathna, D. Y., Pang, Y., Abeywickrama, C. S., & Wijesinghe, K. J. (2023). Bright NIR-emitting styryl pyridinium dyes with large stokes' shift for sensing applications. *Biosensors*, 13(8), 799. <https://doi.org/10.3390/bios13080799>
- [14] Wu, L., Huang, J., Pu, K., & James, T. D. (2021). Dual-locked spectroscopic probes for sensing and therapy. *Nature Reviews Chemistry*, 5(6), 406-421. <https://doi.org/10.1038/s41570-021-00277-2>
- [15] Zhang, B., Chen, H., Shi, L., Guo, R., Wang, Y., Zheng, Y., ... & Zhang, X. (2024). Nitroreductase-based "turn-on" fluorescent probe for bacterial identification with visible features. *Acs Sensors*, 9(9), 4560-4567. <https://doi.org/10.1021/acssensors.3c02785>
- [16] Liu, J., Abdulkadir, A. Z., Wu, S., Gao, Y., Butuyuyu, B. J., Wong, K. M. C., ... & Zhang, P. (2025). Cyanine-scaffold fluorogenic probes for visual detection of nitroreductase in living bacteria. *Journal of Materials Chemistry B*, 13(38), 12224-12233. <https://doi.org/10.1039/D5TB00758E>
- [17] Morsby, J. J., Zhang, Z., Burchett, A., Datta, M., & Smith, B. D. (2024). Ratiometric near-infrared fluorescent probe for

- nitroreductase activity enables 3D imaging of hypoxic cells within intact tumor spheroids. *Chemical Science*, 15(10), 3633-3639. <https://doi.org/10.1039/D3SC06058F>
- [18] An, Z., Sun, Y., Wang, D., Ge, J., Zhao, C., Lu, H., ... & Wang, Q. (2024). Ultra-sensitive and fast nitroreductase-responsive fluorescence probe for real-time bioimaging of hypoxia in tumor cells and zebrafish wound inflammation. *Sensors and Actuators B: Chemical*, 420, 136458. <https://doi.org/10.1016/j.snb.2024.136458>
- [19] Sarkar, S., Shil, A., Jun, Y. W., Yang, Y. J., Cheng, R., Wu, W., ... & Ahn, K. H. (2025). Real-Time Detection of Reduced Nitroreductase with a Reversible Fluorescent Probe. *Advanced Science*, 12(42), e08689. <https://doi.org/10.1002/adv.202508689>
- [20] Gao, W., Shi, A., Hou, Y., Zhang, P., Zhang, Q., & Ding, C. (2025). A turn on fluorescent probe for nitroreductase activity and its application in real-time imaging of tumor hypoxia. *Talanta*, 290, 127804. <https://doi.org/10.1016/j.talanta.2025.127804>
- [21] Zhang, S., Ma, M., Zhao, C., Li, J., Xu, L., Zhang, Z., ... & Song, D. (2024). A novel low-background nitroreductase fluorescent probe for real-time fluorescence imaging and surgical guidance of thyroid cancer resection. *Biosensors and Bioelectronics*, 261, 116514. <https://doi.org/10.1016/j.bios.2024.116514>
- [22] Hu, C., Liu, H., Zhang, Z., Li, L., Mao, G. J., Cheng, W., & Zhou, L. (2025). A Self-Calibrating Fluorescent-Photoacoustic Integrated Probe Enables Fast Visualizing Pancreatic Cancer and Imaging-Guided Tumor Surgery. *Small*, 21(20), 2408527. <https://doi.org/10.1002/sml.202408527>

Practical Pathways for Enhancing Artificial Intelligence Literacy among University Teachers in Ningbo under the Background of Digital-Intelligent Transformation

Kebiao Yuan*, Yun Ren

School of Economics and Management, Ningbo University of Technology, Ningbo, Zhejiang, 315211, China

*Corresponding author: Kebiao Yuan, ykbjob@163.com

Copyright: 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Against the backdrop of digital-intelligent transformation, artificial intelligence has evolved from an auxiliary tool into a technological force that is increasingly embedded in university teaching design, research support, administrative services, and social engagement, and is reshaping the modes of knowledge production, talent cultivation, and organizational operation in higher education institutions. For university teachers, artificial intelligence literacy does not merely refer to the ability to use a particular tool; rather, it represents a professional capacity to understand technological logic, judge application boundaries, and transform artificial intelligence into an enabling force for curriculum reform, research innovation, student development, and professional growth. Ningbo has a solid manufacturing foundation, a well-developed port and shipping logistics system, and abundant application scenarios in digital trade, intelligent manufacturing, and industry-education integration, which provide a favorable practical foundation for enhancing artificial intelligence literacy among university teachers. Based on a review of studies on educational digital transformation, teacher digital literacy, and artificial intelligence literacy, and in light of the actual development of higher education institutions in Ningbo, this paper analyzes the major problems in the development of university teachers' artificial intelligence literacy, including uneven conceptual awareness, incomplete competency structures, fragmented training systems, insufficient disciplinary integration, weak ethical governance, and imperfect evaluation and incentive mechanisms. To address these problems, this paper proposes practical pathways such as standards-based guidance, tiered cultivation, scenario-driven application, curriculum reconstruction, integration of teaching and research, university-government-enterprise collaboration, ethics-first governance, and closed-loop evaluation. The study argues that the enhancement of artificial intelligence literacy among university teachers in Ningbo should not remain at the level of general tool training, but should be oriented toward real teaching, research, and local industrial contexts, enabling teachers to gradually move from simply using artificial intelligence to using it effectively, cautiously, and creatively.

Keywords: Digital-Intelligent Transformation; Ningbo Higher Education Institutions; University Teachers; Artificial Intelligence Literacy; Educational Digital Transformation; Practical Pathways

Published: Jun 16, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1504>

1. Research Background and Problem Statement

1.1 Digital-Intelligent Transformation Is Reshaping the Higher Education Ecosystem

Digital-intelligent transformation can be understood as a process in which digitalization and intelligentization become integrated across technology, organization, and practice. It is based on data resources, supported by artificial intelligence algorithms, and carried by digital platforms. In operation, it is characterized by intelligent decision-making, human-machine collaboration, and dynamic optimization. Unlike earlier stages of educational informatization, which mainly focused on hardware configuration, network access, and online resource development, digital-intelligent transformation places greater emphasis on the deep impact of technology on educational structures, teaching relationships, operational mechanisms, and value logic. Zhu and Hu argue that educational digital transformation is not a simple upgrading of informatization. Rather, it involves systematic changes in new paradigms, new thinking, new capabilities, new environments, new resources, new applications, and new cultures ^[1]. Yu further points out that educational digital transformation includes changes in technology, business processes, and human-centered development, and it involves the entire process from localized informatization applications to school organizational transformation and then to the formation of a new ecosystem of smart education services ^[2]. Therefore, under the background of digital-intelligent transformation, the key issue in higher education reform is no longer whether technology is available, but whether technology can truly enter the processes of teaching, research, and governance, thereby promoting substantive changes in the operation of higher education institutions.

In university contexts, the influence of digital-intelligent transformation does not emerge in isolation. It penetrates various aspects of education, teaching, and organizational operation. In terms of teaching resources, digital textbooks, online courses, virtual simulation experiments, knowledge graphs, and intelligent resource libraries provide new conditions for teachers to reorganize teaching content and update course cases. In the teaching process, learning platforms, classroom interaction systems, intelligent teaching assistants, and learning analytics systems can track students' learning conditions and provide teachers with richer data support for process diagnosis and personalized guidance. From the perspective of teaching evaluation, artificial intelligence can support feedback on student assignments, learning profiles, formative assessment, and comprehensive competency analysis, thereby shifting evaluation from single-result judgment to process-oriented and comprehensive evaluation. In research activities, artificial intelligence has also begun to participate in literature retrieval, data processing, model construction, manuscript revision, and academic trend analysis, influencing the research process of university teachers. In university governance, the application of data middle platforms, smart campus platforms, and intelligent decision-making systems is also transforming university management from extensive administration toward refined and collaborative development.

University teachers play a connecting role between technological change and educational practice. Without the necessary artificial intelligence literacy, educational digitalization may remain limited to platform construction, tool use, and resource accumulation, and it cannot be truly transformed into a process that improves the quality of talent cultivation. Conversely, if teachers can understand the basic logic of artificial intelligence, master methods of human-machine collaborative teaching, develop data analysis awareness, and form ethical judgment capabilities, then artificial intelligence can become an important force for improving curriculum teaching, promoting student development, enhancing research capacity, and serving local industries. Based on the above analysis, enhancing university teachers' artificial intelligence literacy is not only an inevitable requirement for teachers to adapt to technological change, but also a key factor determining whether the digital-intelligent transformation of higher education institutions can be successfully realized.

1.2 Policy Requirements Have Made Teachers' Artificial Intelligence Literacy an Issue of the Times

In recent years, educational digitalization and the empowerment of education through artificial intelligence have received sustained attention at the national level. According to relevant documents issued by the Ministry of Education, by 2024, China should establish a national strategy for educational digitalization, promote smart campus construction, explore effective pathways for digitally empowered large-scale personalized teaching and innovative teaching, use artificial intelligence to promote educational transformation, improve standards for teachers' and students' digital literacy, deepen the application of artificial intelligence in teacher development, and strengthen data security, algorithm security, and ethical security ^[3]. This indicates that teacher digital literacy and artificial intelligence literacy are no longer merely local technical capacity issues. They have become part of the broader agenda of building China into an education power and modernizing the teaching

workforce.

The “Artificial Intelligence + Education” Action Plan further proposes that artificial intelligence should empower student learning, teacher teaching, educational governance, and scientific research, while promoting full-staff training, competency assessment, and evaluation reform^[4]. The Teacher Digital Literacy standard issued by the Ministry of Education defines teacher digital literacy as the awareness, ability, and responsibility of teachers to discover, analyze, and solve educational and teaching problems based on the acquisition, processing, use, management, and evaluation of digital information and resources, thereby optimizing, innovating, and transforming educational and teaching activities^[5]. The Framework for Artificial Intelligence Literacy of University Teachers in Zhejiang Province (Trial) further proposes that university teachers’ artificial intelligence literacy consists of three components: basic artificial intelligence literacy, ethical norms, and application literacy^[6]. From national policy to local frameworks, these documents provide institutional support for higher education institutions in Ningbo to establish teacher artificial intelligence literacy standards, enrich training systems, and improve evaluation mechanisms.

1.3 The Practical Urgency of Enhancing Artificial Intelligence Literacy among University Teachers in Ningbo

Ningbo is a major manufacturing city, a port city, and an open economy city, with a solid industrial foundation and abundant artificial intelligence application scenarios. The public consultation draft of the Ningbo Artificial Intelligence Innovation and Development Action Plan (2025–2030), issued by the Ningbo Municipal Bureau of Economy and Information Technology, represents the local government’s policy orientation toward accelerating artificial intelligence development and promoting industrial upgrading empowered by artificial intelligence^[7]. Ningbo’s “Hundred Scenarios and Thousand Products AI Empowerment” plan takes AI-enabled new industrialization as its starting point and promotes artificial intelligence empowerment for new industrialization from four aspects: AI-enabled scenarios, AI-enabled manufacturing, AI empowerment in digital economy parks, and artificial intelligence product supply^[8]. These local arrangements show that the development of artificial intelligence in Ningbo is not merely a matter of technological renewal in the industrial sector. It will also influence the direction of talent cultivation, teacher development, and industry-education integration in higher education institutions.

Higher education institutions in Ningbo are diverse in type, including comprehensive universities, application-oriented undergraduate institutions, vocational colleges, and Chinese-foreign cooperative universities. Although these institutions differ in their educational positioning, they all need to answer the same question: whether their teachers possess the professional capacity to support talent cultivation in the intelligent era. Comprehensive universities rely on artificial intelligence to promote interdisciplinary development and research innovation. Application-oriented undergraduate institutions need to integrate artificial intelligence into professional courses and engineering practice. Vocational colleges should focus on job-based scenarios such as intelligent manufacturing, smart logistics, and digital trade to strengthen digital teaching and intelligent practical training. Zhang’s study on the implementation path of digital teaching in vocational colleges shows that teacher digital literacy and digital teaching in vocational colleges are coupled and mutually reinforcing, and that digital teaching in vocational colleges needs to be promoted simultaneously in teaching content, teaching design, teaching modes, and teaching evaluation^[9]. At the same time, Ningbo has been selected into the relevant construction list of the second batch of national city-level industry-education consortiums, which has laid a platform foundation for universities, government, and enterprises to collaboratively enhance teachers’ artificial intelligence literacy^[10].

2. Theoretical Foundations and Analytical Framework

2.1 Research on Teacher Digital Literacy Provides a Foundational Support for Artificial Intelligence Literacy

Teacher artificial intelligence literacy is not a new concept detached from the existing system of teacher professional competence. Rather, it is a new type of professional literacy further developed on the basis of teacher digital literacy. Wu, Gui, Zhou, and others argue that teacher digital literacy is an important component of teacher development in the digital era, and its evaluation should be promoted through tiered classification, digital-intelligent drive, and continuous monitoring^[11]. Zhang and Yang analyze the internal mechanism through which teacher digital literacy promotes digital education

development. They argue that teacher digital literacy can advance the high-quality development of digital education through teaching method innovation, improved resource utilization, enhanced evaluation methods, and teacher professional growth^[12]. Mei points out that in the process of enhancing university teachers' digital literacy, there are still problems such as deviation from educational goals, insufficient support systems, lack of frontier technologies, and neglect of subject characteristics. Corresponding support systems should therefore be established from the dimensions of concepts, capabilities, institutions, and culture^[13]. Zhuang and others argue that cultivating university teachers' digital literacy should focus on stimulating internal motivation, building communication platforms, improving training mechanisms, establishing assessment and evaluation systems, and creating a favorable ecosystem^[14].

The above studies indicate that teachers' digital literacy does not simply refer to technical operational ability. It includes digital awareness, digital technology knowledge and skills, digital application, digital social responsibility, and professional development. Artificial intelligence literacy further deepens these dimensions by placing greater emphasis on teachers' understanding and application of data, algorithms, models, human-machine collaboration, intelligent evaluation, and ethical governance.

2.2 Research on Teacher Artificial Intelligence Literacy Emphasizes the Unity of Competency Development and Ethical Norms

Existing research on teacher artificial intelligence literacy has produced a number of valuable findings, most of which focus on competency structures and value norms. Li, Sun, Jiang, and others argue that university teachers' artificial intelligence literacy should respond to the new requirements imposed on university teachers in the intelligent era in areas such as teaching and educating students, research innovation, social service, and cultural inheritance and innovation. Teachers should not only master basic artificial intelligence knowledge and application skills, but also use intelligent technologies to promote teaching evaluation reform, research innovation, and industry-university-research collaboration^[15]. Wang argues that teachers' professional literacy in the intelligent era should reflect the unity of intelligence and humanity, and that a proper balance should be found among technology application, intelligent talent cultivation, and value guidance^[16]. Guo and Hao argue that teachers in the intelligent era should possess abilities related to technological cognition, innovative teaching, human-machine collaboration, resource integration, data application, and ethical security^[17].

Liu and Yin suggest that improving teachers' digital literacy through artificial intelligence should be demand-oriented, standards-led, scenario-based, and feedback-driven^[18]. Wei, Yang, and Xu argue that scenario-based teacher digital literacy training should not remain at the level of abstract competency training. Instead, it should shift toward task-driven and practice-generated training in real educational contexts^[19]. Therefore, teacher artificial intelligence literacy should not be simply regarded as the ability to use AI tools. It should be understood as a comprehensive professional capacity involving value cognition, technical understanding, teaching application, research innovation, and ethical governance.

2.3 Policy Standards and AI-TPACK Theory Provide Important References

From the perspective of policy and theory, the cultivation of teacher artificial intelligence literacy should be based on national and local standards, while also relying on theories of disciplinary teaching integration to construct practical frameworks. The Teacher Digital Literacy standard defines teacher digital literacy from the dimensions of digital awareness, digital technology knowledge and skills, digital application, digital social responsibility, and professional development. The Framework for Artificial Intelligence Literacy of University Teachers in Zhejiang Province (Trial) summarizes university teachers' artificial intelligence literacy into three basic elements: basic artificial intelligence literacy, ethical norms, and application literacy^{[5][6]}. Cen argues that the improvement of teacher competence should involve the linkage of policy standards, training systems, application scenarios, and evaluation mechanisms^[20].

AI-TPACK theory also provides an analytical tool for enhancing teachers' artificial intelligence literacy. Yan, Fu, Zhu, and others argue that teachers need to understand the relationship among artificial intelligence technology, disciplinary content, and teaching methods, and integrate the three organically in specific teaching contexts^[21]. This has strong implications for higher education institutions in Ningbo. The application of artificial intelligence differs across disciplines. Engineering, business, humanities and social sciences, and art and design programs have quite different needs for artificial intelligence

application. Therefore, enhancing artificial intelligence literacy among university teachers in Ningbo should not adopt a one-size-fits-all approach. Instead, it should be promoted according to disciplinary characteristics, professional positioning, and teaching scenarios.

3. The Practical Foundation for Enhancing Artificial Intelligence Literacy among University Teachers in Ningbo

3.1 Policy Foundation

Enhancing artificial intelligence literacy among university teachers in Ningbo is supported not only by policy foundations, but also by a relatively solid higher education foundation, digital economy foundation, intelligent manufacturing foundation, port and shipping logistics foundation, and technological innovation foundation. To strengthen the objectivity of the analysis, this paper selects relevant Ningbo statistical data for 2025 to quantitatively present the practical foundation, as shown in Table 1.

Table 1. Practical Foundation Data for Enhancing Artificial Intelligence Literacy among University Teachers in Ningbo

Dimension	Representative Indicator	2025 Data	Significance for Enhancing University Teachers' Artificial Intelligence Literacy
Higher education foundation	Number of higher education institutions in Ningbo	16	Indicates that Ningbo has the institutional foundation for enhancing artificial intelligence literacy among university teachers.
Higher education foundation	Number of students enrolled in higher education institutions in Ningbo	257,000	Indicates that the enhancement of teachers' artificial intelligence literacy will affect the quality of talent cultivation for a large student population.
Digital economy foundation	Added value of core digital economy industries	RMB 159.9 billion	Provides practical support for teachers to reform courses related to the digital economy, artificial intelligence, and data analysis.
Digital economy foundation	Growth rate of the added value of core digital economy industries	12.1%	Indicates rapid development of Ningbo's digital economy and higher requirements for teachers' digital-intelligent teaching capacity.
Intelligent manufacturing foundation	Growth rate of added value of high-tech manufacturing	11.9%	Provides industrial scenarios for course construction in intelligent manufacturing, industrial internet, machine vision, and digital twins.
Intelligent manufacturing foundation	Growth rate of industrial robot output	46.4%	Indicates that Ningbo's manufacturing industry is undergoing significant intelligent transformation, which can be converted into practical teaching and industry-education integration projects.
Port and shipping logistics foundation	Cargo throughput of Ningbo-Zhoushan Port	1.43 billion tons	Provides local cases for professional courses on smart ports, smart logistics, and intelligent supply chain decision-making.
Port and shipping logistics foundation	Container throughput of Ningbo-Zhoushan Port	43.872 million TEUs	Indicates abundant port and shipping logistics scenarios, which are conducive to scenario-based AI teaching and research.
Technological innovation foundation	Number of valid high-tech enterprises	10,050	Provides an enterprise foundation for university teachers to conduct university-enterprise cooperation, horizontal research projects, and industrial case development.
Technological innovation foundation	Number of registered technology contracts	6,393	Indicates active industry-university-research cooperation in Ningbo and supports teachers in integrating AI applications into research and social service.

Source: Ningbo Municipal Bureau of Statistics, 2025 Statistical Communiqué on Ningbo's National Economic and Social Development.

Table 1 shows that enhancing artificial intelligence literacy among university teachers in Ningbo has strong regional practical

support. On the one hand, Ningbo has 16 higher education institutions and 257,000 enrolled university students, indicating that the enhancement of teachers' artificial intelligence literacy has a broad educational coverage. On the other hand, the added value of Ningbo's core digital economy industries reached RMB 159.9 billion, while indicators related to high-tech manufacturing, industrial robots, port and shipping logistics, and high-tech enterprises also performed strongly. This shows that the enhancement of artificial intelligence literacy among university teachers in Ningbo should not remain at the level of general tool use. Instead, it should be integrated with local scenarios such as intelligent manufacturing, smart ports, digital trade, industrial innovation, and industry-education integration.

The enhancement of artificial intelligence literacy among university teachers in Ningbo has relatively clear policy support. At the national level, the Outline of the Plan for Building China into an Education Power (2024–2035) and the Artificial Intelligence + Education Action Plan provide direction for the development of university teachers' artificial intelligence literacy, while the Teacher Digital Literacy standard provides a basic reference for the development of teachers' digital competence ^{[3][4][5]}. At the provincial level, the Framework for Artificial Intelligence Literacy of University Teachers in Zhejiang Province (Trial) further clarifies the basic components of university teachers' artificial intelligence literacy ^[6]. At the local level, Ningbo's artificial intelligence innovation and development actions, the “Hundred Scenarios and Thousand Products AI Empowerment” plan, and the construction of city-level industry-education consortiums provide industrial scenarios and collaborative mechanisms for enhancing teachers' literacy ^{[7][8][10]}. Overall, national policy provides direction, the provincial framework provides competency structure, and local arrangements provide scenario support and regional coordination.

3.2 Industrial Foundation

Ningbo has a relatively complete industrial system. Intelligent manufacturing, green petrochemicals, new energy vehicles, port and shipping logistics, foreign trade services, cross-border e-commerce, smart healthcare, cultural tourism, and urban governance all generate rich demand for artificial intelligence applications. Ningbo's “Hundred Scenarios and Thousand Products AI Empowerment” plan emphasizes the leading role of application scenarios in the large-scale application of new artificial intelligence technologies and products, providing practical resources for university teachers to conduct curriculum development, case construction, practical teaching, and research cooperation ^[8].

To further illustrate the practical foundation of Ningbo's industrial digital-intelligent development, this paper selects the growth rates of relevant indicators such as the digital economy, high-tech manufacturing, and equipment manufacturing in Ningbo in 2025 for comparison, as shown in Figure 1.

Source: Ningbo Municipal Bureau of Statistics, 2025 Statistical Communiqué on Ningbo's National Economic and Social Development.

Figure 1. Comparison of Growth Rates of Digital-Intelligent Industries in Ningbo in 2025

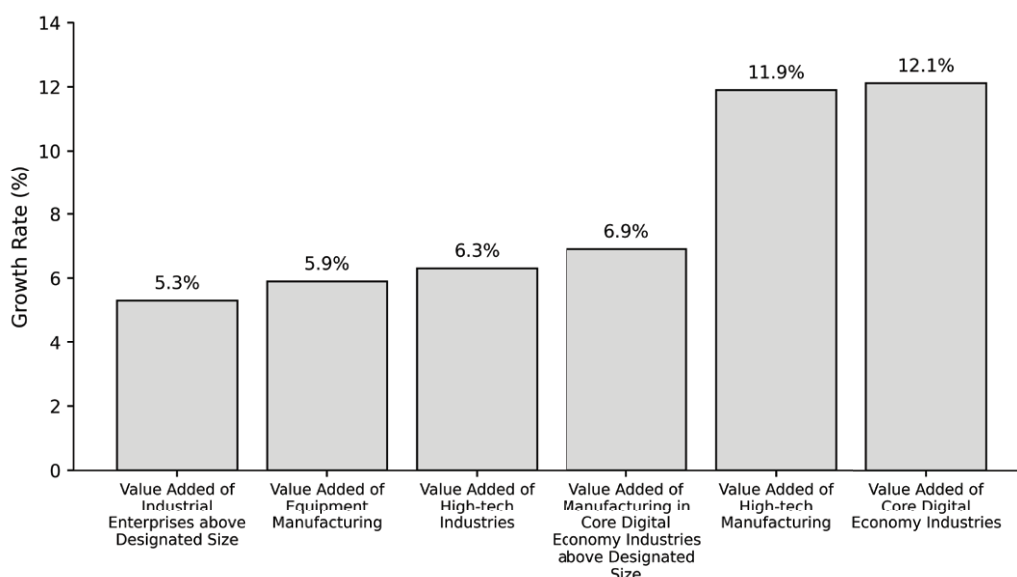


Figure 1 shows that in 2025, the added value of Ningbo's core digital economy industries increased by 12.1%, and the added value of high-tech manufacturing increased by 11.9%, both significantly higher than the average growth rate of 5.3% for industrial enterprises above designated size. This indicates that Ningbo's industrial structure is accelerating its transformation toward the digital economy, high-tech manufacturing, and intelligent manufacturing. For university teachers, this industrial transformation means that the enhancement of artificial intelligence literacy should not be limited to the use of general tools. It should be further oriented toward the digital economy, intelligent manufacturing, equipment manufacturing, industrial data analysis, and scenario-based industrial teaching, so as to better serve Ningbo's local industrial development and the cultivation of application-oriented and interdisciplinary talent.

The above analysis shows that the integration of artificial intelligence and manufacturing can support teaching cases such as industrial visual inspection, intelligent scheduling, predictive equipment maintenance, and digital twin factories. The integration of artificial intelligence and port and shipping logistics can be transformed into course projects on port scheduling, intelligent warehousing, and supply chain forecasting. The integration of artificial intelligence and digital trade can serve practical teaching in intelligent customer service, cross-border e-commerce data analysis, user profiling, and intelligent marketing. Therefore, industrial scenarios transform abstract artificial intelligence learning into practical problem-solving processes, which is consistent with the concepts of "scenario implementation" and "scenario-based training" ^{[18][19]}.

3.3 Institutional and Organizational Foundation

The diversity of higher education institutions in Ningbo provides practical space for classified promotion of teacher artificial intelligence literacy. Comprehensive universities can engage in basic artificial intelligence research, interdisciplinary development, and intelligent research methods. Application-oriented undergraduate institutions can focus on intelligent transformation of professional courses, industry-education integration projects, and engineering practice capacity development. Vocational colleges can focus on job competencies, practical training scenarios, and the application of intelligent equipment. Chinese-foreign cooperative universities can learn from international digital education resources and cross-cultural intelligent education experience.

At the same time, higher education institutions in Ningbo generally have teacher development centers, teaching promotion centers, academic affairs training systems, teaching competition systems, and curriculum construction mechanisms. These platforms provide organizational foundations for enhancing teachers' artificial intelligence literacy. Universities can integrate artificial intelligence literacy training into new teacher induction training, young teacher teaching capacity improvement, backbone teacher training, online excellent course construction, and teaching innovation competition cultivation, instead of creating an entirely separate system. Zhu, Zhang, and others argue that teacher digital literacy assessment scenarios should be constructed according to the concept of "technology empowerment, scenario integration, and process accompaniment" ^[22]. The same idea can also be applied to enhancing artificial intelligence literacy among university teachers in Ningbo.

4. Major Problems in Enhancing Artificial Intelligence Literacy among University Teachers in Ningbo

4.1 Uneven Conceptual Awareness

At present, university teachers in Ningbo differ in their understanding and acceptance of artificial intelligence. Some teachers have already applied generative artificial intelligence to lesson preparation, information collection, case creation, manuscript revision, data analysis, teaching feedback, and other areas. Some teachers remain cautious and regard artificial intelligence as a temporary technological trend. Others worry that artificial intelligence will increase their workload, weaken their professional authority, and lead to academic misconduct, and therefore oppose its use.

The deeper reason is that some teachers regard artificial intelligence merely as a tool for improving efficiency, such as assisting in writing materials, making slides, and generating test questions. They do not fully recognize that artificial intelligence may change course content, learning methods, evaluation methods, and teachers' roles. In the intelligent era, teachers need to achieve role transformation in human-machine collaboration, resource integration, data application, and ethical security ^[17]. Without corresponding conceptual renewal, it is difficult for teachers to develop a high level of artificial intelligence literacy.

4.2 Incomplete Competency Structure

Although some teachers have begun to use artificial intelligence tools, their competency structures remain incomplete. This is mainly reflected in several aspects: they can use tools but do not understand the basic principles; they can generate content but lack the ability to judge its quality; they can use AI to improve efficiency but cannot design human-machine collaborative learning activities; they can use intelligent scoring but lack awareness of academic norms and authorship boundaries; and they can use AI to assist research but do not fully understand academic standards and authorship boundaries. University teachers' artificial intelligence literacy includes tool use ability, but it also includes educational concepts, artificial intelligence knowledge and skills, teaching innovation capacity, research innovation capacity, and ethical awareness ^[15]. Therefore, higher education institutions in Ningbo should not focus only on tool lists and operational techniques. They should pay more attention to the systematic and complete structure of teachers' artificial intelligence literacy.

4.3 Fragmented Training Systems

In recent years, artificial intelligence-related training in universities has increased. However, most training still takes the form of lectures, tool demonstrations, and short-term workshops. The content mainly focuses on introductions to generative artificial intelligence tools and prompt techniques, without forming systematic course design, disciplinary case development, teaching practice tracking, and outcome evaluation. Teachers may feel inspired during training sessions, but when they return to specific courses, they still encounter problems such as not knowing how to design AI-supported assignments, how to adjust course evaluation, and how to handle students' use of AI.

Teacher digital literacy training should move toward scenario-based development and should not be separated from teachers' real tasks and educational contexts ^[19]. Therefore, if higher education institutions in Ningbo rely only on scattered lectures and short-term training to improve teachers' artificial intelligence application capacity, it will be difficult to form stable and transferable competence.

4.4 Insufficient Disciplinary Integration

The enhancement of artificial intelligence literacy must ultimately be implemented in professional curriculum reform. In reality, however, some teachers regard artificial intelligence as a public basic course or as content belonging to computer science majors, believing that it has little connection with their own disciplines. Even when some teachers are willing to try, they often use AI only as an auxiliary tool and do not seriously consider the internal relationship among professional knowledge systems, job competency requirements, and artificial intelligence applications. To generate educational value, artificial intelligence technology must be integrated with disciplinary content and teaching methods ^[21]. Without disciplinary integration, teachers' artificial intelligence literacy can easily remain at a shallow application level.

4.5 Insufficient Awareness of Ethical Governance

After artificial intelligence enters university teaching and research, it also produces new risks and challenges. The first is data security risk. Teachers and students may upload materials containing personal information, course data, experimental data, or enterprise cooperation data to external platforms. The second is academic integrity risk. Students may use AI to write assignments, papers, and laboratory reports, while teachers may also improperly use AI-generated content in research. The third is intellectual property risk. The copyright ownership, citation methods, and scope of use of AI-generated texts, images, codes, and cases need to be further standardized. The fourth is evaluation fairness risk. Intelligent scoring and learning profiling may be affected by data quality, algorithmic bias, and model limitations. The Teacher Digital Literacy standard emphasizes teachers' digital social responsibility ^[5], and the Framework for Artificial Intelligence Literacy of University Teachers in Zhejiang Province regards ethical norms as an important component of university teachers' artificial intelligence literacy ^[6]. Therefore, ethical governance awareness should become an important part of enhancing artificial intelligence literacy among university teachers in Ningbo.

4.6 Imperfect Evaluation and Incentive Mechanisms

Whether teachers are willing to continuously strengthen their artificial intelligence literacy is closely related to university evaluation and incentive mechanisms. If teachers invest considerable time in intelligent curriculum transformation, AI teaching case construction, and intelligent teaching assistant maintenance, but these efforts are not reflected in teaching

workload, performance assessment, professional title evaluation, awards, and honors, then their enthusiasm will be affected. The evaluation of teacher digital literacy should reflect tiered classification, digital-intelligent drive, and continuous monitoring, and should also use scenario-based assessment and process evaluation to reflect teachers' real abilities ^{[11][22]}. Therefore, higher education institutions in Ningbo should establish an evaluation system that reflects teachers' practical contributions, rather than using participation in training and submission of materials as the main evaluation criteria.

5. A Target Framework for Enhancing Artificial Intelligence Literacy among University Teachers in Ningbo

The enhancement of artificial intelligence literacy among university teachers in Ningbo should not remain at the level of mastering a few AI tools. Instead, it should point toward a more complete form of digital-intelligent educational competence. Digital-intelligent educational competence refers to teachers' comprehensive ability to correctly understand, select, apply, and evaluate artificial intelligence technologies in digital-intelligent educational environments according to educational laws and disciplinary characteristics, and to apply them in student learning, curriculum reform, research innovation, and educational governance.

Based on the practical needs of higher education institutions in Ningbo, teachers' artificial intelligence literacy should consist of three basic levels. The first is the ability to "understand AI." Teachers should have a comprehensive understanding of the basic concepts, technical principles, application fields, and limitations of artificial intelligence, rather than regarding artificial intelligence merely as a tool for automatically generating texts or courseware. The second is the ability to "use AI." Teachers should be able to apply artificial intelligence to teaching design, course resource development, classroom interaction, personalized tutoring, learning assessment, and research, so that technology can genuinely serve teaching and research. The third is the ability to "govern AI." Teachers should be able to identify risks related to data security, algorithmic bias, academic integrity, intellectual property, and educational ethics. Only by integrating understanding, application, and governance can teachers' artificial intelligence literacy avoid remaining at the level of simple operation.

In terms of competency structure, artificial intelligence literacy among university teachers in Ningbo can be understood from five dimensions. Value cognition literacy requires teachers to understand the relationship between artificial intelligence and the development of higher education, and to recognize that technology application cannot be separated from talent cultivation, educational equity, and high-quality development. Basic knowledge literacy requires teachers to understand basic concepts such as machine learning, deep learning, generative artificial intelligence, large models, knowledge graphs, natural language processing, and learning analytics. Teaching application literacy refers to whether teachers can use artificial intelligence in intelligent lesson preparation, resource generation, learning analysis, human-machine collaborative classrooms, intelligent Q&A, and formative assessment. Research innovation literacy refers to teachers' ability to use artificial intelligence properly for literature retrieval, data processing, research review, model construction, and manuscript revision. Ethical governance literacy refers to teachers' adherence to data security, privacy protection, intellectual property, academic integrity, algorithmic fairness, and educational ethics when using artificial intelligence.

From the perspective of developmental levels, different teachers do not need to advance according to the same standard at the same pace. Most teachers should reach the basic application level, meaning that they can master basic concepts and common tools of artificial intelligence and complete basic work such as intelligent lesson preparation, information organization, classroom support, and assignment feedback. Teachers who already have a foundation in curriculum reform can further develop into integration-innovation teachers, meaning that they can carry out intelligent curriculum transformation according to disciplinary characteristics, design human-machine collaborative learning activities, and use data for learning diagnosis and formative assessment. Teaching backbone teachers and program leaders can develop toward a leading-demonstration level by promoting AI-empowered course groups, professional case libraries, and teaching reform projects through team building, while also participating in university-enterprise cooperation, regional promotion, and demonstration applications.

6. Practical Pathways for Enhancing Artificial Intelligence Literacy among University Teachers in Ningbo

6.1 Standards-Based Guidance: Establishing School-Based Artificial Intelligence Literacy Standards

Higher education institutions in Ningbo can refer to the national Teacher Digital Literacy standard^[5] and the Framework for Artificial Intelligence Literacy of University Teachers in Zhejiang Province. On this basis, they can combine the implications of AI-TPACK theory regarding the integration of artificial intelligence technology, disciplinary content, and teaching methods to formulate teacher artificial intelligence literacy standards that better fit their own institutional realities^{[4][6][21]}. School-based standards should not simply copy higher-level documents. They should be transformed into specific requirements that teachers can understand, implement, and evaluate.

Universities can first identify basic requirements that all teachers should meet, including basic understanding of artificial intelligence, use of common tools, basic teaching applications, data security awareness, and academic integrity norms. On this basis, classified requirements can be proposed according to disciplinary differences. Engineering teachers can pay more attention to engineering data analysis, intelligent manufacturing cases, and simulation practice. Business teachers can focus more on business intelligence, intelligent decision-making, and digital trade cases. Humanities and social science teachers can emphasize text analysis, digital humanities, and intelligent communication ethics. Art and design teachers can focus on generative creative tools, visual expression, and copyright norms. Vocational college teachers can develop around job scenarios, intelligent equipment, and practical training teaching. Through such treatment, artificial intelligence literacy standards will no longer be abstract statements. Instead, they will enter the professional development process of teachers in different disciplines.

6.2 Tiered Cultivation: Building a Training System with Full Coverage and Classified Promotion

Higher education institutions in Ningbo can establish a tiered training system consisting of “general popularization—disciplinary integration—project practice—demonstration leadership.” General popularization training should be oriented toward all teachers and mainly solve the problems of “not knowing, not daring to use, and not knowing how to use AI properly.” The content should include artificial intelligence development trends, basic principles of generative AI, use of common tools, prompt design, educational application scenarios, data security, intellectual property, and academic integrity. Such training should not be overly technical. It should enable teachers from different disciplinary backgrounds to form basic understandings of when AI can be used and when it should be used cautiously.

Disciplinary integration training can be organized by colleges, programs, or course groups and should be directly linked to teachers’ own courses. Intelligent manufacturing programs can conduct training on industrial AI case development. Digital trade programs can conduct training on AI-enabled cross-border e-commerce. Journalism and communication programs can conduct training on AIGC and communication ethics. Project-based practical training should not remain at the listening stage. It should require teachers to produce specific outcomes, such as designing an AI-empowered teaching unit, transforming a course assignment, creating an intelligent course knowledge base, developing a local industrial case, or writing a teaching reform paper. For backbone teachers, workshops, enterprise practice, inter-university exchanges, and action research can be used to improve their capacity in course group construction, intelligent evaluation design, teaching achievement cultivation, and team leadership.

6.3 Scenario-Driven Application: Promoting the Integration of Artificial Intelligence into the Whole Teaching Process

Artificial intelligence literacy can become a stable teaching capacity only when it enters real teaching contexts. For higher education institutions in Ningbo, instead of treating AI training as an additional task, it is more effective to integrate artificial intelligence into pre-class, in-class, after-class, and evaluation processes, so that teachers can experience the value and boundaries of technology in specific tasks.

Before class, teachers can use artificial intelligence to decompose teaching objectives, organize knowledge points, collect cases, design teaching activities, and generate learning resources. They can also create course knowledge graphs to help students develop an overall understanding of the knowledge structure. During class, teachers can design learning activities in which students dialogue with, compare, question, and revise AI-generated content, enabling students to analyze the rationality, limitations, and possible improvements of AI-generated outputs. After class, teachers can use AI to analyze

students' assignments, tests, and discussion data, identify common errors and individual learning difficulties, and provide personalized learning suggestions. Artificial intelligence can support formative assessment by recording learning trajectories, providing personalized feedback, and identifying learning risks, but it cannot completely replace teachers' judgment. Especially in the evaluation of innovation capacity, value judgment, and practical performance, teachers should still play a leading role.

6.4 Curriculum Reconstruction: Building AI-Empowered Professional Curriculum Systems

The enhancement of teachers' artificial intelligence literacy should ultimately return to curriculum reform. The Artificial Intelligence Action Plan proposes promoting artificial intelligence as a public basic course in higher education so that all students can master artificial intelligence knowledge^[4]. Higher education institutions in Ningbo can offer artificial intelligence general education courses for all students, covering artificial intelligence basics, data literacy, algorithmic thinking, generative AI applications, artificial intelligence ethics, and industry cases. Such courses can improve students' cognitive level and also lay a common foundation for the intelligent transformation of professional courses.

At the same time, each program should promote intelligent curriculum transformation according to its talent cultivation objectives. Programs related to intelligent manufacturing can add content on industrial AI, machine vision, industrial internet, and digital twins. International trade and e-commerce programs can add content on intelligent marketing, intelligent customer service, cross-border data analysis, and supply chain forecasting. Journalism and communication programs can add content on intelligent communication, algorithmic recommendation, AIGC content governance, and media ethics. Art and design programs can add content on AIGC design, interactive experience, and copyright norms. Higher education institutions in Ningbo can jointly build local industrial AI case libraries with enterprises, developing real cases in intelligent factories, smart ports, cross-border e-commerce, urban governance, smart healthcare, cultural tourism, and other areas. This can prevent artificial intelligence courses from being separated from professional teaching and can also promote the transformation of single-course reform toward the overall development of professional course groups.

6.5 Integration of Teaching and Research: Promoting the Coordination of Teaching, Research, and Social Service

University teachers undertake teaching, research, and social service. The enhancement of artificial intelligence literacy should serve all three dimensions. Teachers can use artificial intelligence for literature retrieval, research hotspot analysis, text mining, data processing, model construction, and manuscript revision. Li and others argue that artificial intelligence can improve university teachers' ability to collect, organize, analyze, and present research data, thereby improving research efficiency and research quality^[15]. However, artificial intelligence should only serve as an auxiliary tool in research. Teachers should not treat AI-generated content as original research outcomes, nor should they fabricate literature or data.

Teachers can also treat AI-supported teaching practice as a teaching research issue. They can explore the influence of AI-assisted feedback on student engagement, the effect of human-machine collaborative learning on critical thinking, the limits of generative AI in professional writing, and the ways in which intelligent evaluation supports learning outcomes. The result of such teaching research is that teachers shift from being users of experience to research-oriented practitioners. Teachers can also connect with Ningbo's industrial scenarios to carry out horizontal research projects, enterprise training, technical services, and case development. The construction of city-level industry-education consortiums provides an institutional platform for university teachers to participate in industrial practice, and real problems generated when teachers serve enterprises can also feed back into curriculum teaching and research innovation^[10].

From the perspective of the regional innovation environment, university teachers in Ningbo have a relatively strong practical foundation for AI-empowered teaching research, research transformation, and social service. To further illustrate this point, this paper compares the growth rates of indicators such as high-value invention patents, valid invention patents, high-tech enterprises, and registered technology contracts in Ningbo in 2025, as shown in Figure 2.

Source: Ningbo Municipal Bureau of Statistics, 2025 Statistical Communiqué on Ningbo's National Economic and Social Development.

Figure 2. Comparison of Growth Rates of Science and Technology Innovation Indicators in Ningbo in 2025

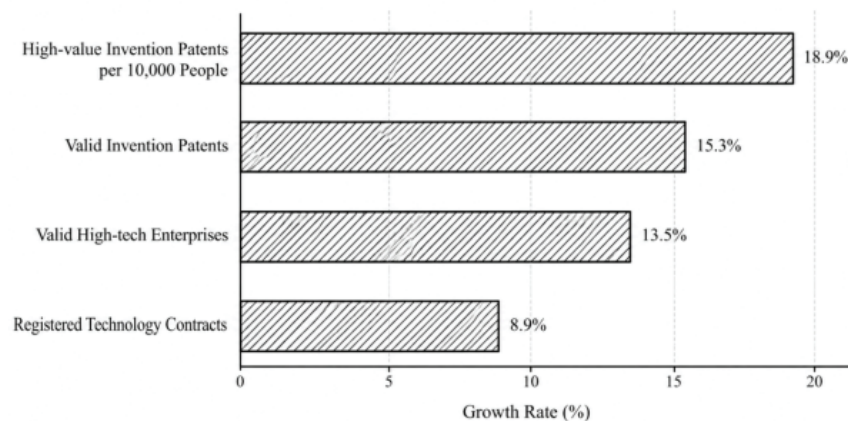


Figure 2 shows that in 2025, the number of high-value invention patents per 10,000 people in Ningbo increased by 18.9%, valid invention patents increased by 15.3%, valid high-tech enterprises increased by 13.5%, and registered technology contracts increased by 8.9%. These data indicate that Ningbo has active technological innovation actors and a strong foundation for technology achievement transformation. For university teachers, this regional innovation environment provides practical support for AI-empowered teaching research, horizontal project cooperation, enterprise case development, and social service. The enhancement of teachers' artificial intelligence literacy should not only serve classroom teaching, but also extend to research innovation, industry-education integration, and local industrial services.

6.6 University-Government-Enterprise Collaboration: Building a Regional Ecosystem for Enhancing Teachers' Artificial Intelligence Literacy

The enhancement of artificial intelligence literacy among university teachers in Ningbo cannot rely solely on internal university efforts. It also requires the participation of government, universities, enterprises, industry associations, and other actors. The government can incorporate the enhancement of university teachers' artificial intelligence literacy into policy systems related to artificial intelligence industrial development, educational digitalization, talent development, and industry-education integration, and support teacher development through special projects, training platforms, practice bases, and case promotion. Universities should play the leading role by establishing school-level coordination bodies led by teacher development centers and jointly participated in by academic affairs offices, human resources offices, research offices, informatization departments, and secondary colleges.

Enterprises should not only provide internship positions, but also introduce real business scenarios, data samples, tool platforms, and practical instructors into teacher training and curriculum construction. Higher education institutions in Ningbo can also establish an alliance for enhancing teacher artificial intelligence literacy, sharing training courses, expert resources, case libraries, tool platforms, and evaluation standards. In this way, the enhancement of teachers' artificial intelligence literacy will no longer be merely an internal matter of individual universities, but can gradually form a regional development community with Ningbo characteristics.

6.7 Ethics-First Governance: Improving Governance Mechanisms for Educational Applications of Artificial Intelligence

The enhancement of artificial intelligence literacy should begin with ethics. Universities should formulate guidance manuals for teachers' use of artificial intelligence, specifying permitted, cautious, and prohibited application scenarios. Teachers may use AI to assist lesson preparation, case generation, and public information organization, but they must not casually upload student privacy data, examination materials, enterprise confidential data, or unpublished research data. Course syllabi should also specify which artificial intelligence technologies students may use, to what extent they may use them, whether they need to disclose the usage process, and whether they need to submit dialogue records or reflective explanations.

Universities should establish classified and graded management mechanisms for educational data. When using students' personal information, classroom records, examination results, research data, and enterprise cooperation data, use should be

restricted to internal platforms. Universities can build trusted on-campus AI platforms and private course knowledge bases to create secure and controllable intelligent tool environments for teachers. At the same time, artificial intelligence ethics should be integrated into teacher training and student courses, helping teachers and students understand issues such as algorithmic bias, privacy protection, intellectual property, deepfakes, academic integrity, and responsibility attribution. Only when ethical requirements are placed first can the educational application of artificial intelligence develop steadily.

6.8 Closed-Loop Evaluation: Establishing a Support Mechanism for Teachers' Artificial Intelligence Literacy Development

Evaluation mechanisms should serve teacher growth rather than add new formalistic burdens. Universities can form teachers' artificial intelligence literacy profiles through self-assessment, training records, curriculum practice, student feedback, peer evaluation, and outcome recognition. Such profiles can identify teachers' strengths and weaknesses and provide a basis for personalized training. Teachers' participation in AI training, completion of curriculum transformation, creation of intelligent resources, publication of teaching research outcomes, participation in enterprise projects, and acquisition of relevant certifications can be included in teacher development records and used as references for professional title evaluation, post appointment, performance rewards, and awards.

At the same time, universities should improve incentive policies and incorporate AI-empowered teaching reform achievements into teaching workload recognition, teaching project approval, curriculum construction funding, teaching achievement awards, and professional title evaluation. Good courses, outstanding cases, and demonstration teams should receive funding, honors, and platform support. Universities should also regularly evaluate the effects of teachers' artificial intelligence literacy enhancement, focusing on teacher participation, the number of transformed courses, changes in teaching quality, students' learning experience, ethical risk incidents, and outcome production. Evaluation results should be used to improve subsequent work. Only by forming a closed loop of training, practice, evaluation, and improvement can the enhancement of teachers' artificial intelligence literacy become sustainable.

7. Implementation Steps and Support Mechanisms

The enhancement of artificial intelligence literacy among university teachers in Ningbo should be promoted in stages. In the short term, the priority is to solve the problems of teachers "not knowing, not daring to use, and not knowing how to use AI properly." Universities can use full-staff general training to help teachers understand basic concepts, typical application scenarios, and major ethical risks of artificial intelligence. They can also quickly formulate guidelines for educational applications of artificial intelligence, stipulate basic behavioral norms for teachers and students using artificial intelligence, and select some courses with a good foundation for early pilot projects to generate replicable and promotable cases. It is not appropriate to require all teachers to carry out comprehensive reforms from the very beginning.

In the medium term, the focus should shift from popularization training to deeper application-oriented exploration, solving problems such as being able to use AI but not using it deeply, using it in scattered ways, and having difficulty transforming its use into concrete outcomes. Universities should establish a tiered certification system for teachers' artificial intelligence literacy, encouraging teachers at different levels to complete corresponding practical tasks. Secondary colleges should carry out intelligent curriculum transformation according to professional characteristics and create AI-driven course groups and Ningbo local industrial case libraries. At the same time, universities should support teachers in conducting teaching research based on AI-empowered teaching, incorporating curriculum reform, teaching cases, and teaching reform papers into the process of teacher professional development.

In the long term, higher education institutions in Ningbo should gradually form a regional ecosystem for enhancing teachers' artificial intelligence literacy. The government, universities, enterprises, and industry organizations should jointly build open and shared training platforms, case resource libraries, practice bases, and research communities, realizing resource sharing, experience sharing, and outcome sharing. Through such regional collaboration, Ningbo can form a local model for enhancing university teachers' artificial intelligence literacy and provide useful experience for other cities in the Yangtze River Delta.

Universities should establish school-level leadership groups for AI-empowered education, clarify leading and participating units, and treat the enhancement of teachers' artificial intelligence literacy as an important part of institutional development

planning. Universities should also increase financial investment to support teacher training, platform construction, course resource development, case library construction, and teaching reform projects. Institutions with sufficient conditions can create secure and controllable artificial intelligence education platforms to provide teachers with support for intelligent lesson preparation, resource management, intelligent teaching assistants, learning analytics, intelligent evaluation, and teaching research. At a deeper level, universities should form a digital-intelligent educational culture that encourages exploration, allows trial and error, values sharing, and upholds ethics, so that teachers are willing to learn actively, practice continuously, and innovate collaboratively.

8. Conclusion and Future Prospects

Under the background of digital-intelligent transformation, enhancing university teachers' artificial intelligence literacy has become a key component of educational digital transformation. It is also a practical support for improving the quality of talent cultivation, promoting research innovation, and serving local development. Higher education institutions in Ningbo are located at the intersection of the national educational digitalization strategy, Zhejiang Province's artificial intelligence education deployment, and Ningbo's industrial digital-intelligent transformation. They have policy foundations, industrial foundations, institutional foundations, and organizational foundations for promoting the enhancement of teachers' artificial intelligence literacy.

At present, the enhancement of artificial intelligence literacy among university teachers in Ningbo still faces problems such as uneven conceptual awareness, incomplete competency structures, fragmented training systems, insufficient disciplinary integration, weak ethical governance awareness, and imperfect evaluation and incentive mechanisms. Solving these problems cannot be achieved through several tool training sessions or short-term lectures. It requires joint promotion from multiple aspects, including systematic capacity building, institutional support, professional curriculum integration, and continuous training. Especially in a city like Ningbo, which has a strong foundation in manufacturing and port and shipping logistics, the enhancement of university teachers' artificial intelligence literacy should be integrated with local industrial scenarios, professional curriculum reform, and industry-education integration practice.

In the future, higher education institutions in Ningbo can establish practical pathways for enhancing teachers' artificial intelligence literacy through standards-based guidance, tiered cultivation, scenario-driven application, curriculum reconstruction, integration of teaching and research, university-government-enterprise collaboration, ethics-first governance, and closed-loop evaluation. The purpose of enhancing artificial intelligence literacy is not to allow technology to replace teachers, but to enable teachers to better understand, master, reflect on, and creatively use technology. In the digital-intelligent era, university teachers should understand educational laws and master the basic principles of intelligent technologies. They should not merely rely on artificial intelligence to improve the efficiency of teaching and research, but should also uphold the fundamental goal of education and comply with academic norms. They should not only use artificial intelligence to improve courses and research, but also guide students to develop correct views of technology, learning, and values. Only by continuously promoting the cultivation of teachers' artificial intelligence literacy can higher education institutions in Ningbo form new advantages for high-quality development in the wave of educational digitalization and provide talent and intellectual support for Ningbo to build itself into a modern coastal metropolis and a highland for artificial intelligence innovation and application.

Funding

This research was supported by China Higher Education Society 2025 Annual Higher Education Scientific Research Planning Project (grant no. 25ZJ0306): Research on the Innovation Evaluation Mechanism and Path of Vocational Education System Empowered by Artificial Intelligence, and Ningbo Educational Science Planning Project (grant no. 2025YGHGJ05): Research on the Practical Pathways for Enhancing Artificial Intelligence Literacy among University Teachers in Ningbo under the Background of Digital-Intelligent Transformation.

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Zhu, Z.-T., & Hu, J. (2022). An exploration of the essence of educational digital transformation and research prospects. *China Educational Technology*, (4), 1–8, 25.
- [2] Yu, S.-Q. (2023). Key pathways of educational digital transformation. *Journal of East China Normal University (Educational Sciences)*, 41(3), 62–71.
- [3] Central Committee of the Communist Party of China & State Council. (2025). Outline of the Plan for Building China into an Education Power (2024–2035) [Z/OL]. (2025-01-19). https://www.gov.cn/zhengce/202501/content_6999913.htm.
- [4] Ministry of Education, National Development and Reform Commission, Ministry of Industry and Information Technology, et al. (2026). Notice of the Ministry of Education and Other Four Departments on Issuing the “Artificial Intelligence + Education” Action Plan: Jiaokexin [2026]No. 1 [Z/OL]. (2026-04-02). https://www.gov.cn/zhengce/zhengceku/202604/content_7065138.htm.
- [5] Ministry of Education. (2022). Teacher Digital Literacy: JY/T 0646—2022 [S/OL]. (2022-11-30). http://www.moe.gov.cn/srcsite/A16/s3342/202302/t20230214_1044634.html.
- [6] Office of the Education Department of Zhejiang Province. (2025). Framework for Artificial Intelligence Literacy of University Teachers in Zhejiang Province (Trial) [Z/OL]. (2025-08-28). <https://tc.zwu.edu.cn/41/94/c2273a213396/page.htm>.
- [7] Ningbo Municipal Bureau of Economy and Information Technology. (2025). Announcement on Publicly Soliciting Opinions on the Ningbo Artificial Intelligence Innovation and Development Action Plan (2025–2030) (Draft for Comments) [EB/OL]. (2025-07-25). <https://app.ningbo.gov.cn/jact/front/mailpubdetail.do?sysid=63&transactId=51183>.
- [8] Ningbo Enterprise Comprehensive Service Platform. (2026). Developing the intelligent economy: Ningbo advances toward the “new” [EB/OL]. (2026-04-09). https://qf.ningbo.gov.cn/qfpt/fwbk/zcfw/rmzx/art/2026/art_bcbeb860fad2421a8cf9643636bd90a7.html.
- [9] Zhang, S.-B. (2023). Implementation pathways of digital teaching in vocational colleges from the perspective of teacher digital literacy. *Journal of Ningbo Institute of Education*, 25(6), 71–75.
- [10] Ministry of Education. (2024). The Ministry of Education deploys work to strengthen the construction of city-level industry-education consortiums [EB/OL]. (2024-11-14). <https://qspfw.moe.gov.cn/html/hotnews/20241114/22433.html>.
- [11] Wu, D., Gui, X.-J., Zhou, C., et al. (2023). Teacher digital literacy: Connotation, standards, and evaluation. *E-Education Research*, 44(8), 108–114, 128.
- [12] Zhang, G.-Y., & Yang, J.-S. (2024). Research on the mechanism of teacher digital literacy driving the development of digital education. *China Educational Technology*, (11), 37–44.
- [13] Mei, B. (2024). Enhancing university teachers’ digital literacy: Practical problems and system construction. *China Higher Education*, (12), 50–54.
- [14] Zhuang, R.-L., Yu, X.-J., & Mi, K.-N. (2025). Cultivating university teachers’ digital literacy under the background of digital transformation: Value connotation, practical challenges, and breakthrough pathways. *Higher Education Forum*, (1), 39–42.
- [15] Li, Y., Sun, L.-Y., Jiang, Q.-Y., et al. (2025). Artificial intelligence literacy of university teachers and improvement strategies. *Open Education Research*, 31(1), 23–33.
- [16] Wang, D. (2022). Research on teachers’ intelligent education literacy from the perspective of artificial intelligence: Connotation, challenges, and cultivation strategies. *Journal of the Chinese Society of Education*, (3), 91–96.
- [17] Guo, J., & Hao, J.-J. (2021). Role positioning and literacy framework of teachers in the intelligent era. *China Educational Technology*, (6), 121–127.
- [18] Liu, B.-Q., & Yin, H.-H. (2024). Artificial intelligence empowering the enhancement of teachers’ digital literacy: Strategies, scenarios, and evaluation feedback mechanisms. *Modern Educational Technology*, 34(7), 23–31.
- [19] Wei, F., Yang, K.-X., & Xu, R.-Y. (2026). Scenario-based teacher digital literacy training: Connotation framework, practical model, and generative AI empowerment mechanism. *China Educational Technology*, (1), 127–135.

- [20] Cen, Y. (2024). Practical dilemmas and global actions in cultivating teacher digital literacy. *Journal of Teacher Education*, 11(4), 38–46.
- [21] Yan, Z.-M., Fu, J.-L., Zhu, Y.-L., et al. (2020). Artificial intelligence technological pedagogical content knowledge (AI-TPACK): Connotation, teaching practice, and future issues. *Journal of Distance Education*, 38(5), 23–34.
- [22] Zhu, L., Zhang, J., Wu, X.-X., et al. (2024). Assessment of teacher digital literacy from the perspective of digital transformation: Development trends, scenario construction, and practical suggestions. *E-Education Research*, 45(2), 113–120.

Heavy Metal Passivation and Microbial Community Succession during Anoxic-Aerobic Digestion of Municipal Sludge

Ruihua Mu*, Kaige Zhang, Wenzhuo Chen*, Xiaoqian Shi, Fengqiu An, Ye Liu

Xi'an Key Laboratory of Textile Chemical Engineering Auxiliaries, School of Environmental and Chemical Engineering, Xi'an Polytechnic University, Xi'an 710048, Shaanxi, China

*Corresponding author: Ruihua Mu, 20131102@xpu.edu.cn; Wenzhuo Chen, wenzhuochen@xpu.edu.cn

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Although anoxic-aerobic digestion has garnered widespread attention as an emerging sludge treatment technology, the mechanisms governing microbial community succession and heavy metal passivation remain incompletely elucidated. In this study, a sequencing batch reactor (SBR) was operated under alternating anoxic-aerobic conditions (dissolved oxygen = 4 mg/L; sludge retention time = 30 days) to systematically investigate these underlying mechanisms. The alternating environment effectively promoted the migration of heavy metals from the liquid to the solid phase, thereby achieving successful passivation. The liquid-phase zinc (Zn) concentration peaked at 5.76 mg/L before subsequently decreasing, whereas the solid-phase Zn concentration remained stable, reaching a maximum of 57.43 mg/g. Highly toxic cadmium (Cd) was not detected at any point during the operation. Furthermore, the system demonstrated significant nutrient recovery potential, as liquid-phase phosphate accumulated to 2902.81 µg/L. High-throughput sequencing revealed that the alternating redox environment fundamentally reshaped the microbial community structure. Although overall microbial diversity decreased, the functional phyla Chloroflexi (34.08%) and Proteobacteria (19.50%) maintained their dominance. Additionally, the abundance of pathogenic genera, including *Acinetobacter*, was markedly reduced. These specific microbial successions were intrinsically driven by the anoxic-aerobic digestion process and exhibited a strong correlation with heavy metal passivation. Ultimately, these findings offer a sustainable and effective approach for the safe disposal and resource utilization of municipal sludge.

Keywords: Municipal Sludge; Anoxic-Aerobic Digestion; Heavy Metal Passivation; Nitrogen and Phosphorus; Microflora

Published: Jun 25, 2026

DOI: <https://doi.org/10.62177/jaet.v3i2.1383>

1. Introduction

As economic development accelerates, rapid urbanization has driven the continuous expansion of sewage treatment facilities and processing capacities. Consequently, the production of residual sludge from wastewater treatment processes has risen consistently. Current estimates indicate that, assuming an 80% moisture content, the annual production of residual sludge ranges between 35 and 45 million tons. This figure is projected to escalate to 90 million tons by 2025, underscoring the severe challenges associated with sludge treatment and disposal. ^[1,2]

Presently, a substantial disparity exists between sludge generation and available treatment capabilities. Statistics reveal

that less than 20% of sludge undergoes safe, standardized treatment, leaving vast amounts of energy-rich residual sludge underutilized. With evolving environmental philosophies, the governance paradigm is shifting from a water-centric approach to one that places equal emphasis on both sludge and water management. Developing efficient treatment technologies that achieve volume reduction, resource recovery, detoxification, and stabilization has therefore become an urgent priority.^[3,4]

Aerobic composting is currently a traditional technology for the resource utilization of municipal sludge, converting organic components into humus through microbial action while retaining nutrients such as nitrogen and phosphorus^[5,6]. However, this technology has inherent limitations, including a large footprint, the generation of odorous gases, high operating costs, and complex management requirements^[7]. In contrast, the anoxic-aerobic digestion process, as an emerging treatment strategy, demonstrates significant advantages: its intermittent oxygen supply mode effectively reduces aeration energy consumption; the alternating redox environment inhibits ammonia volatilization and odor generation while favoring phosphorus retention; and the digested products can be utilized as soil conditioners for land application, offering broad agricultural application prospects^[8].

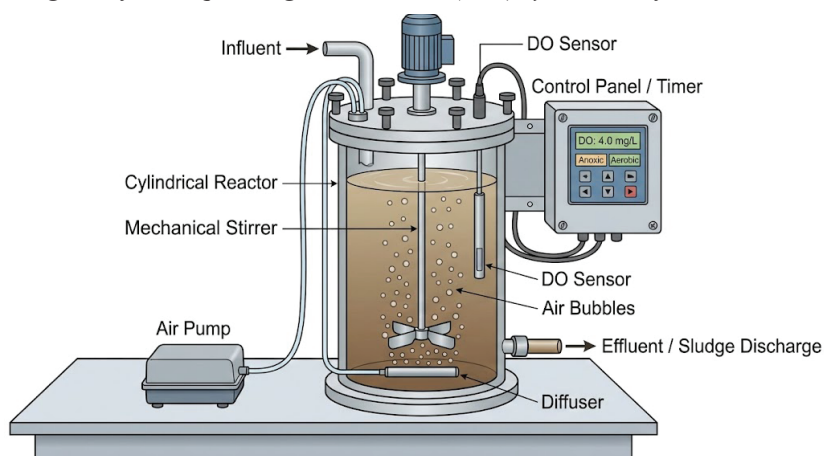
Although the effectiveness of anoxic-aerobic digestion in sludge reduction and stabilization has been preliminarily validated, existing studies have largely focused on it as a volume-reduction method. There is a distinct lack of systematic quantitative assessment regarding the correlations between microbial community shifts and the migration behaviors of heavy metals during this process^[9-12]. In particular, optimizing sludge stabilization through microenvironmental regulation and facilitating heavy metal passivation, rather than treating it merely as a fermentation byproduct, remains a gap in current research. To address this, this study established a sequencing batch reactor (SBR) simulating alternating anoxic-aerobic conditions ($\text{DO} = 4 \text{ mg}\cdot\text{L}^{-1}$, $\text{SRT} = 30$ days) to investigate the roles of functional microbial groups in heavy metal passivation and nutrient transformation. By integrating inductively coupled plasma mass spectrometry (ICP-MS), UV spectrophotometry, and 16S rRNA high-throughput sequencing technologies, this study systematically analyzed the migration and distribution of heavy metals, nitrogen and phosphorus metabolic pathways, and the functional succession of microbial communities in both liquid and solid phases^[13,14]. Unlike traditional qualitative descriptions, this research aims to establish a quantitative link between “process parameters, microbial communities, and heavy metal distribution” through multi-dimensional data analysis.

2. Materials and Methods

2.1 Materials and Experimental Setup

The sludge samples used in this study were obtained from the secondary sedimentation tank of a wastewater treatment plant (A/A/O process) at Xi'an Polytechnic University (Xi'an, China). The sludge had a sludge retention time (SRT) of 30 days. Key chemical reagents, including hydrochloric acid, nitric acid, and hydrofluoric acid, were of analytical grade and purchased from Xilong Scientific Co., Ltd. (Chengdu, China). The experiment was conducted in a sequencing batch reactor (SBR) with an effective working volume of 1 L (Figure 1). The reactor operated under alternating anoxic-aerobic conditions with a hydraulic retention time (HRT) ratio of 1:1. The operational cycle was set to 6 h, with the temperature maintained at 25°C. During the aerobic phase, dissolved oxygen (DO) was maintained at $4 \text{ mg}\cdot\text{L}^{-1}$ using an aeration pump. The total SRT was controlled at 30 days.

Figure 1. Schematic diagram of the sequencing batch reactor (SBR) system used for anoxic-aerobic sludge digestion



2.2 Chemical Analysis and Quantification Methods

Sludge samples were collected every three days and separated into liquid and solid phases via centrifugation (4000 rpm, 10 min) for subsequent analysis. For heavy metal quantification, heavy metals (Cr, Cu, Zn, Cd, Pb) in the liquid phase were digested using an adjustable electric heating plate (Model ML-2.2-4, Kewei Yongxing, Beijing, China) following the standard method GB 7475-87. Solid phase samples were digested using the aqua regia-perchloric acid method according to standard HJ 491-2019. The concentrations of heavy metals in the digests were quantified using Inductively Coupled Plasma Mass Spectrometry (ICP-MS, 7900X, Agilent Technologies, Santa Clara, CA, USA).

Additionally, nitrogen and phosphorus species were analyzed according to standard protocols. Specifically, ammonia nitrogen ($\text{NH}_4^+\text{-N}$) was quantified using Nessler's reagent spectrophotometry. Nitrite nitrogen ($\text{NO}_2^-\text{-N}$) was measured using the N-(1-naphthyl)-ethylenediamine dihydrochloride spectrophotometric method. Nitrate nitrogen ($\text{NO}_3^-\text{-N}$) was analyzed via ultraviolet spectrophotometry, and phosphate ($\text{PO}_4^{3-}\text{-P}$) was determined using the molybdenum-antimony anti-spectrophotometric method. For solid samples, Total Nitrogen (TN) was measured using alkaline potassium persulfate digestion followed by UV spectrophotometry. Total Phosphorus (TP) was quantified using the potassium persulfate digestion-molybdenum antimony anti-spectrophotometric method. All absorbance measurements were performed using a UV-Vis spectrophotometer (T6-1650E, Puxi General Instrument Co., Ltd., Beijing, China).

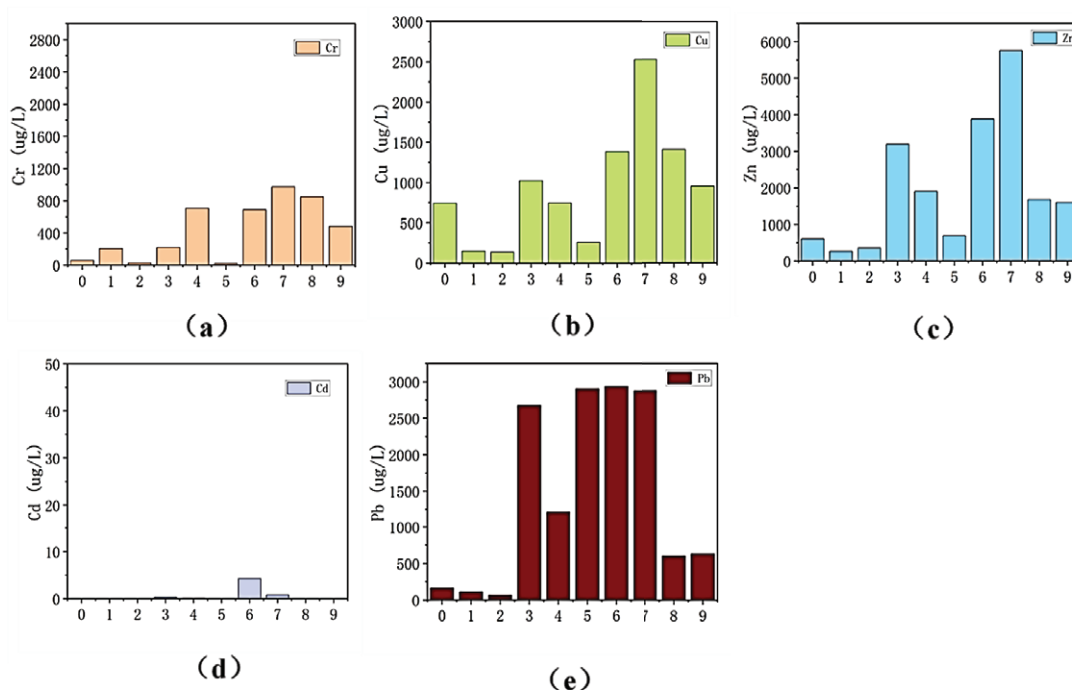
2.3 Microbial Community Analysis

Microbial species and their abundance were identified using 16S rRNA high-throughput sequencing. Sludge samples were collected before the reaction cycle (A0) and after the cycle (A9). DNA was extracted, and the V3–V4 region of the 16S rRNA gene was amplified. For bioinformatics analysis, raw sequencing data were processed to obtain valid sequences. Operational Taxonomic Units (OTUs) were clustered with a 97% similarity cutoff using UPARSE software (version 7.1), and chimeric sequences were identified and removed using UCHIME. The taxonomy of each OTU representative sequence was analyzed by RDP Classifier against the Silva 16S rRNA database (v138) using a confidence threshold of 0.7. Alpha diversity indices (Chao, ACE, Shannon, Simpson, and Coverage) were calculated using MOTHUR to evaluate species richness and diversity.

3. Results

3.1 Heavy Metal Migration and Distribution

Figure 2. Heavy metal content in liquid sludge (a) Cr, (b) Cu, (c) Zn, (d) Cd, (e) Pb



Significant dynamic fluctuations were observed in liquid-phase heavy metal concentrations (Table 1, Figure 2)^[15,16]. Initially, concentrations remained low, but a sharp release was observed mid-cycle. Specifically, Zn concentrations surged to a

peak of $5760.44 \mu\text{g}\cdot\text{L}^{-1}$ on the 6th sampling day, followed by a substantial decline in the final stages. This surge in liquid-phase Zn coincided with a temporary decrease in solid-phase Zn (Figure 3c), confirming the solubilization of metal-organic complexes. In contrast, Cd levels remained negligible throughout the process. In the solid phase (Table 2, Figure 3), heavy metal concentrations remained relatively stable compared to the liquid phase. Zn was the predominant element, peaking at $57.43 \text{ mg}\cdot\text{g}^{-1}$. Notably, the concentrations of Cr, Cu, and Zn in the solid phase showed a slight downward trend after the 6th sampling but stabilized towards the end of the cycle^[17,18].

Figure 3. Heavy metal content in solid sludge (a) Cr, (b) Cu, (c) Zn, (d) Cd, (e) Pb

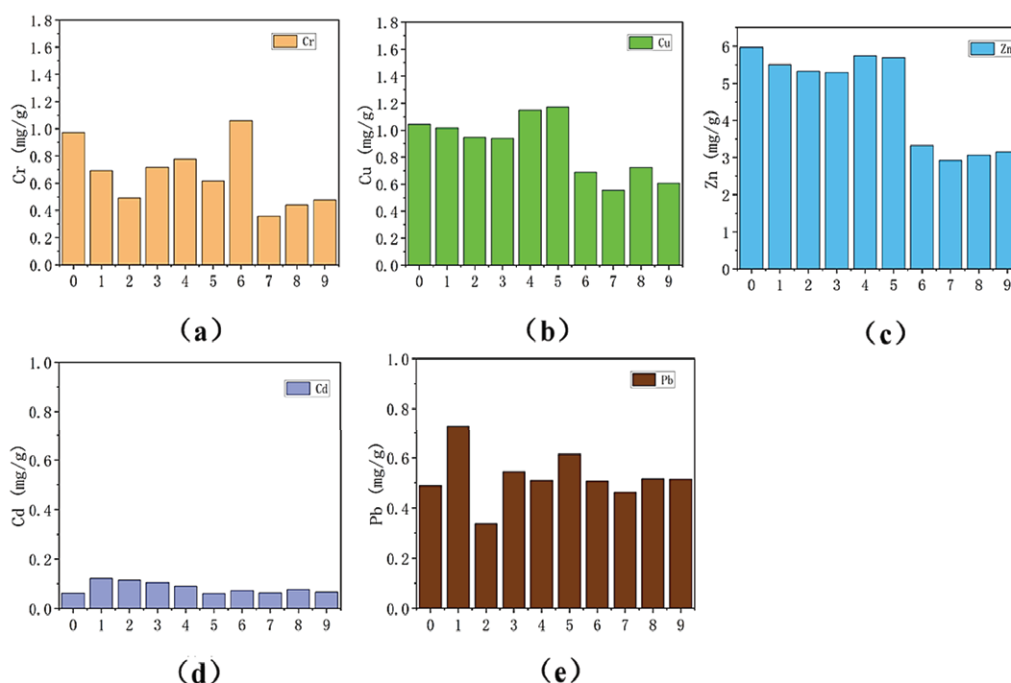


Table 1. Heavy metal content ($\text{mg}\cdot\text{L}^{-1}$) in liquid sludge

Concentration	Cr	Cu	Zn	Cd	Pb
0	57.75	742.4	606.09	0	164.72
1	206.59	147.11	265.94	0	111.27
2	27.64	134.24	356.21	0	66.79
3	219.07	1022.93	3195.56	0.34	2675.67
4	706.11	746.06	1905.42	0.09	1211.17
5	20.87	256.26	691.45	0	2902.81
6	687.7	1382.8	3886.32	4.28	2936.11
7	973.59	2526.4	5760.44	0.81	2875.87
8	847.72	1411.27	1680.97	0	605.83
9	479.92	954.69	1600.88	0	638.32

Table 2. Heavy metal content ($\text{mg}\cdot\text{L}^{-1}$) in solid sludge

Concentration	Cr	Cu	Zn	Cd	Pb
0	3.891	4.173	23.888	0.247	0.165
1	4.152	6.098	33.039	0.727	0.111
2	4.919	9.456	53.249	1.141	0.067
3	7.169	9.393	52.896	1.048	2.676

Concentration	Cr	Cu	Zn	Cd	Pb
4	7.764	11.481	57.426	0.904	1.211
5	6.170	11.716	56.908	0.602	2.903
6	10.589	6.878	33.298	0.726	2.936
7	3.579	5.538	29.215	0.632	2.876
8	4.390	7.237	30.638	0.770	0.606
9	4.766	6.078	31.505	0.664	0.638

3.2 Nitrogen and Phosphorus Transformation

The variations of nitrogen and phosphorus in the liquid phase are illustrated in Figure 4. Phosphate concentrations were remarkably higher than those of nitrogen species, peaking at $2902.81 \mu\text{g}\cdot\text{L}^{-1}$ on the 18th day (6th sampling) before declining [19,20]. Ammonia nitrogen and nitrate nitrogen also showed accumulation trends, whereas nitrite nitrogen remained at low levels. In the solid phase, Total Nitrogen (TN) (Figure 5a) initially decreased and then stabilized, reaching its lowest point on the 9th day. Total Phosphorus (TP) (Figure 5b) exhibited a trend similar to the liquid phase, increasing initially to a peak in the early stages (2nd sampling) and then gradually decreasing [21,22].

Figure 4. Nitrogen and phosphorus content in liquid sludge (a) nitrite nitrogen, (b) ammonia nitrogen, (c) phosphate, (d) nitrate nitrogen

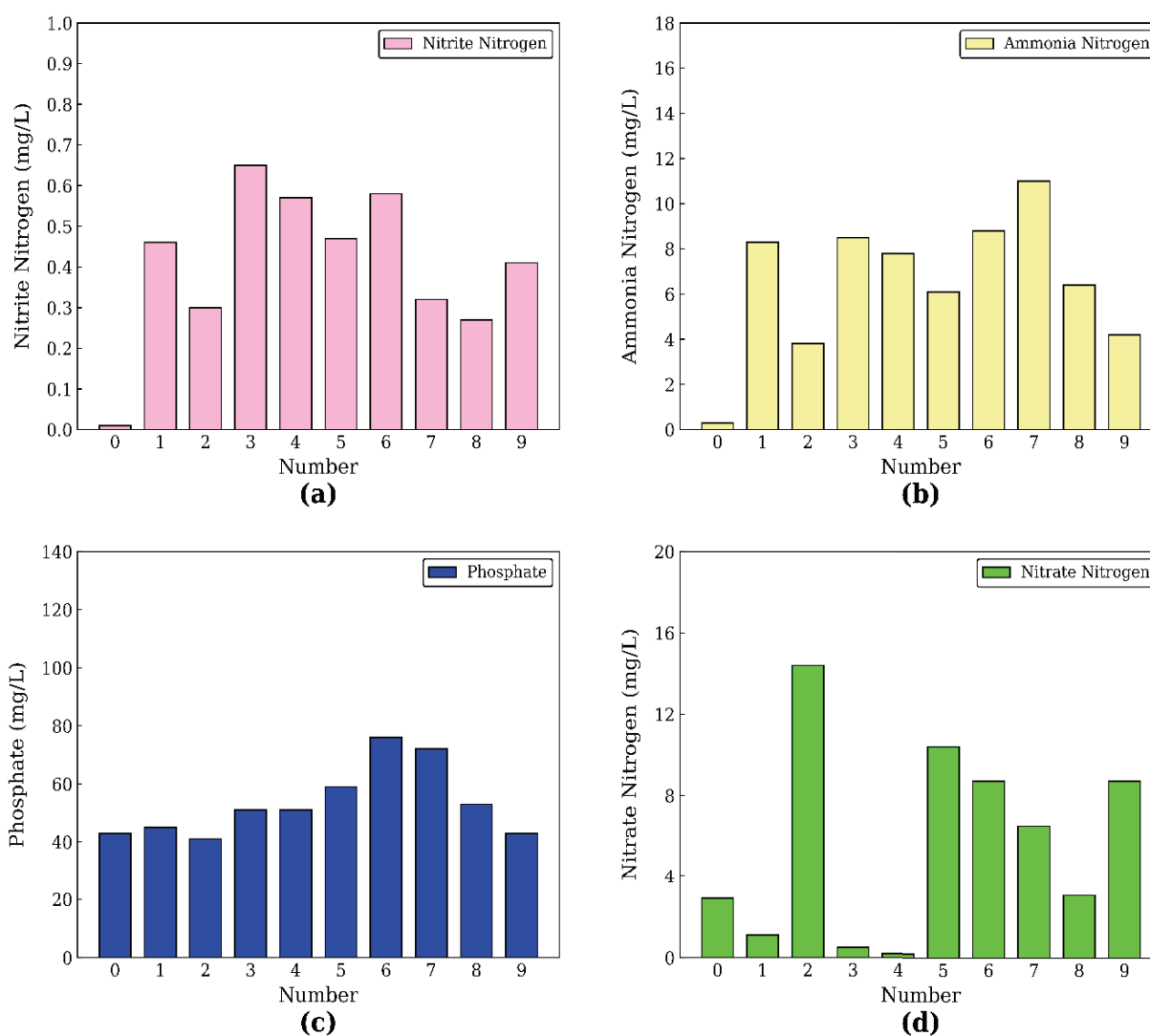
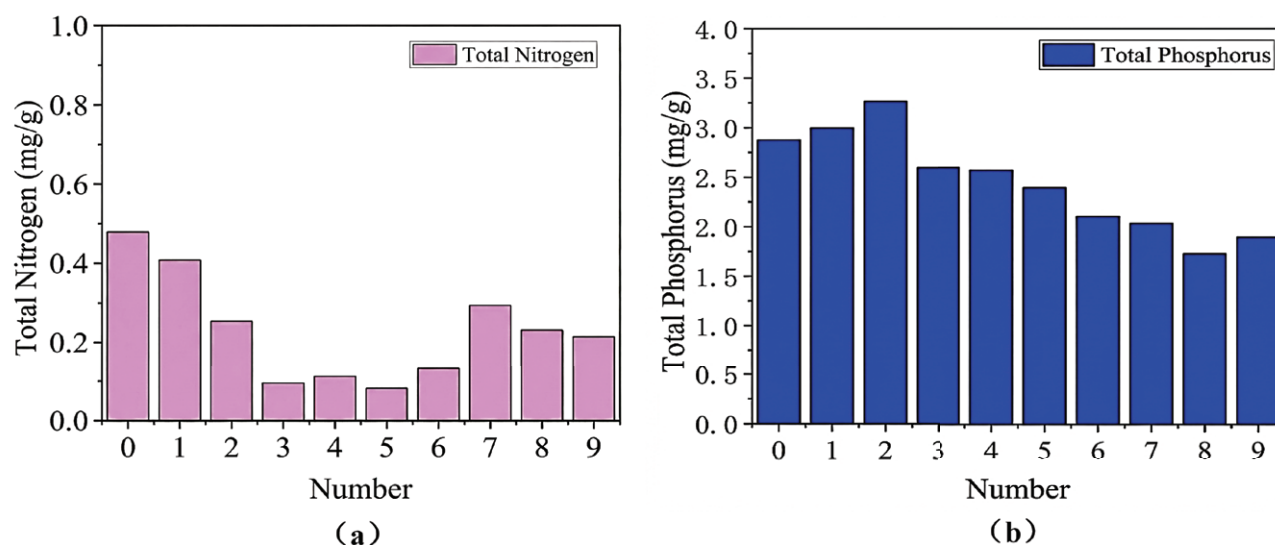


Figure 5. Total nitrogen content in solid sludge and Total phosphorus content in solid sludge



3.3 Microbial Community Diversity, Succession, and Association Analysis

The high-throughput sequencing yielded 49,252 valid sequences per sample. As shown in the Pan and Core OTU analyses (Figure 6), the total species richness increased with sample accumulation, while the number of core OTUs shared between A0 and A9 decreased by approximately one-third, indicating a turnover of microbial populations to adapt to the alternating environment. The adequacy of sequencing depth was confirmed by the rarefaction curves (Figure 7), which reached a clear plateau for both the Sobs and Shannon indices, indicating that the majority of microbial phylotypes were captured. Alpha diversity indices (Figure 8) further quantified the shifts in community complexity. The Goods coverage for all samples exceeded 0.99, reinforcing the reliability of the sequencing data^[23,24]. While the richness indices (Chao and ACE) showed no statistically significant differences between A0 and A9, a notable decline was observed in the diversity indices (Shannon and Simpson). This reduction in evenness implies that the anoxic-aerobic digestion process exerted selective pressure, leading to the dominance of specific functional microbial groups^[25,26].

Figure 6. OTU analysis of sludge samples. (a) Pan OTU curve representing total species richness; (b) Core OTU curve representing shared species

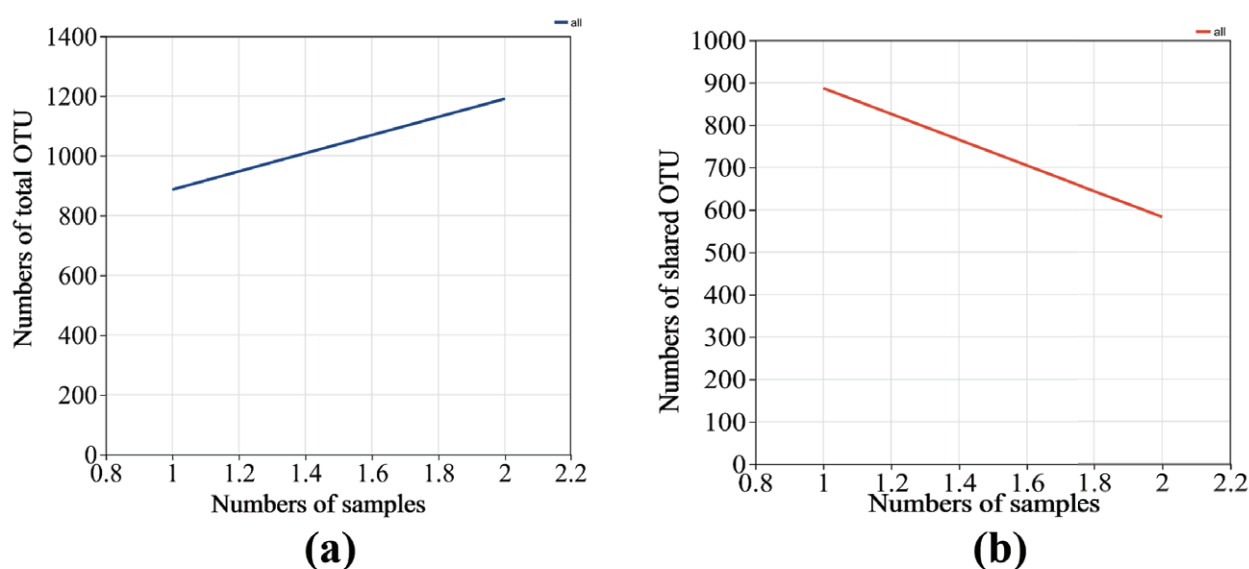


Figure 7. Rarefaction curves verifying sequencing depth suitability. (a) Sobs index; (b) Shannon index

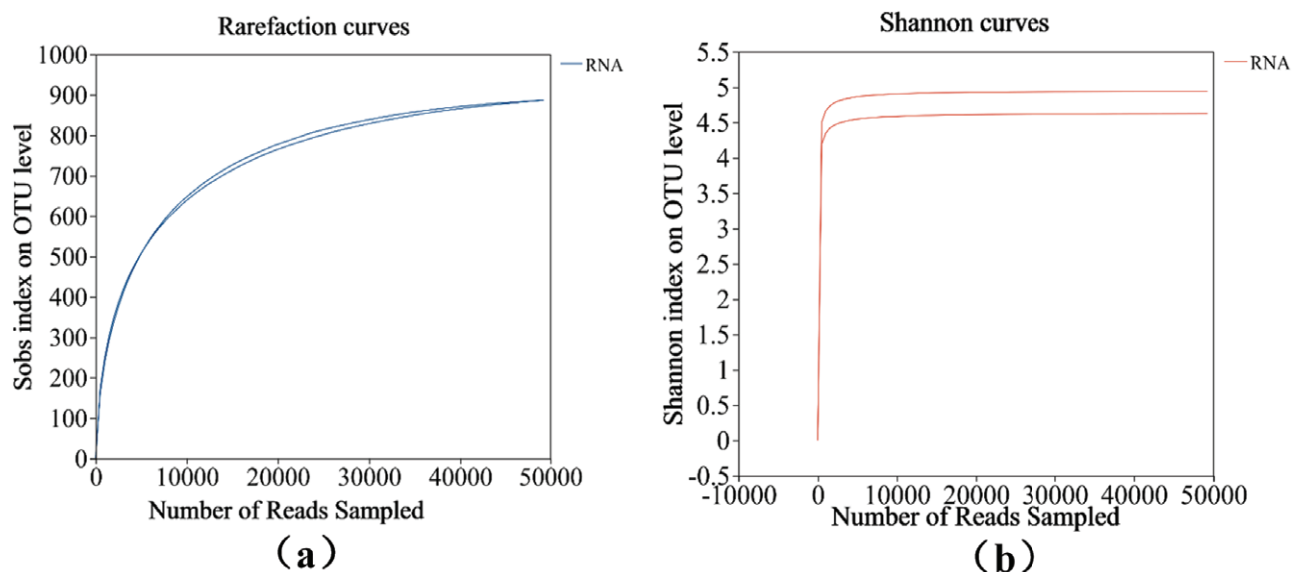
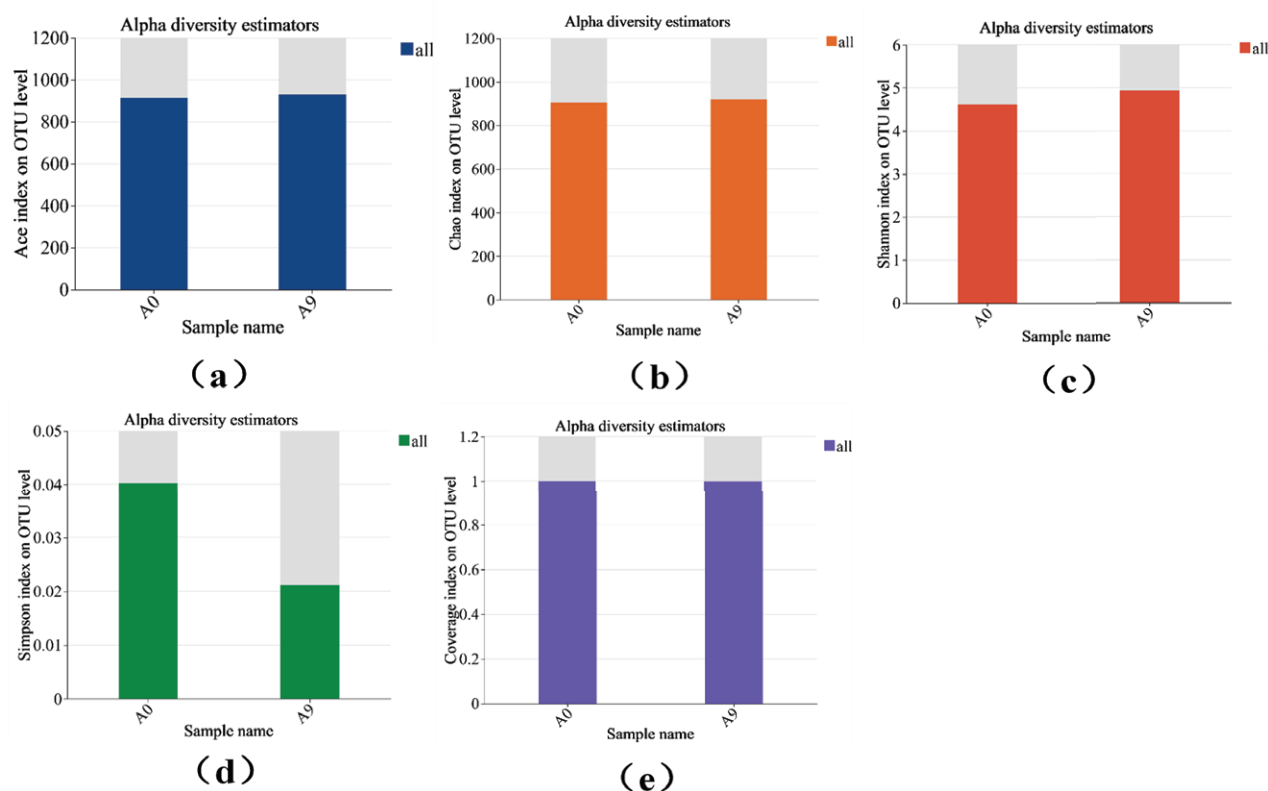


Figure 8. Alpha diversity indices: (a) Ace index, (b) Chao index, (c) Shannon index, (d) Simpson index, (e) Coverage index



The taxonomic composition at the phylum level was visualized using a community bar chart (Figure 9). Chloroflexi and Proteobacteria remained the predominant phyla throughout the process. However, the relative abundance of Chloroflexi decreased from 38% to 34%, and Proteobacteria dropped from 33% to 19% after the 30-day cycle. Conversely, Bacteroidetes and Patescibacteria exhibited an upward trend, with Bacteroidetes increasing from 9.4% to 15.6%. These shifts indicate a significant restructuring of the microbial backbone in response to the alternating redox environment ^[27,28]. At the genus level, the community succession and the distribution of key functional groups were further elucidated via Heatmap analysis (Figure 10). The hierarchical clustering clearly separated the initial sludge (A0) from the digested sludge (A9), indicating a fundamental shift in the dominant taxa. Specifically, the A0 sample was characterized by a high intensity of *Acinetobacter* (approx. 15% relative abundance), a genus often associated with opportunistic pathogens. Following the anoxic-aerobic

treatment, the signal for *Acinetobacter* in the A9 sample was significantly suppressed to negligible levels (<1%). Meanwhile, the heatmap revealed the concurrent enrichment of functional genera such as *Dechloromonas* and various members of the Saprospiraceae family, which are instrumental in nutrient cycling and organic matter stabilization [29,30].

Figure 9. Community bar chart

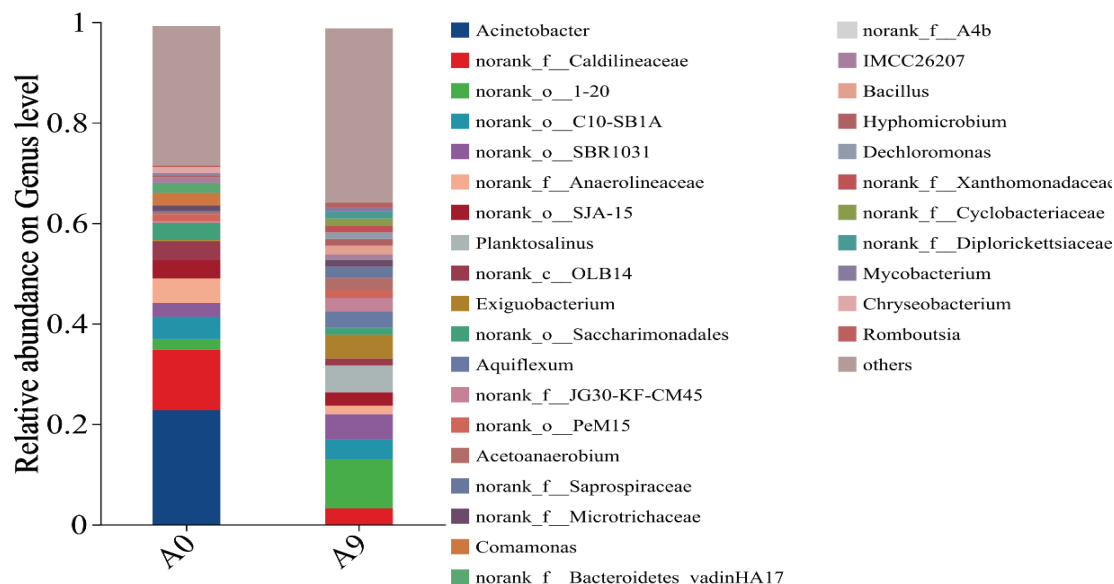
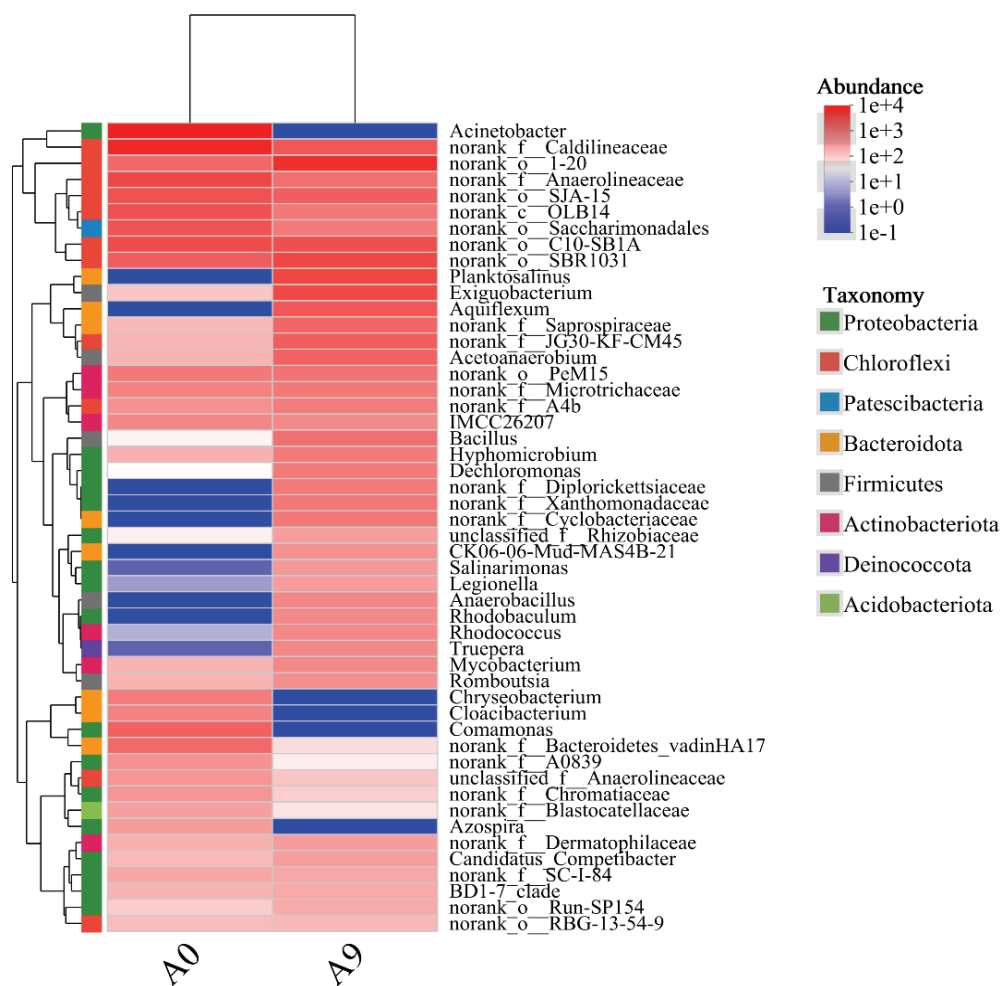
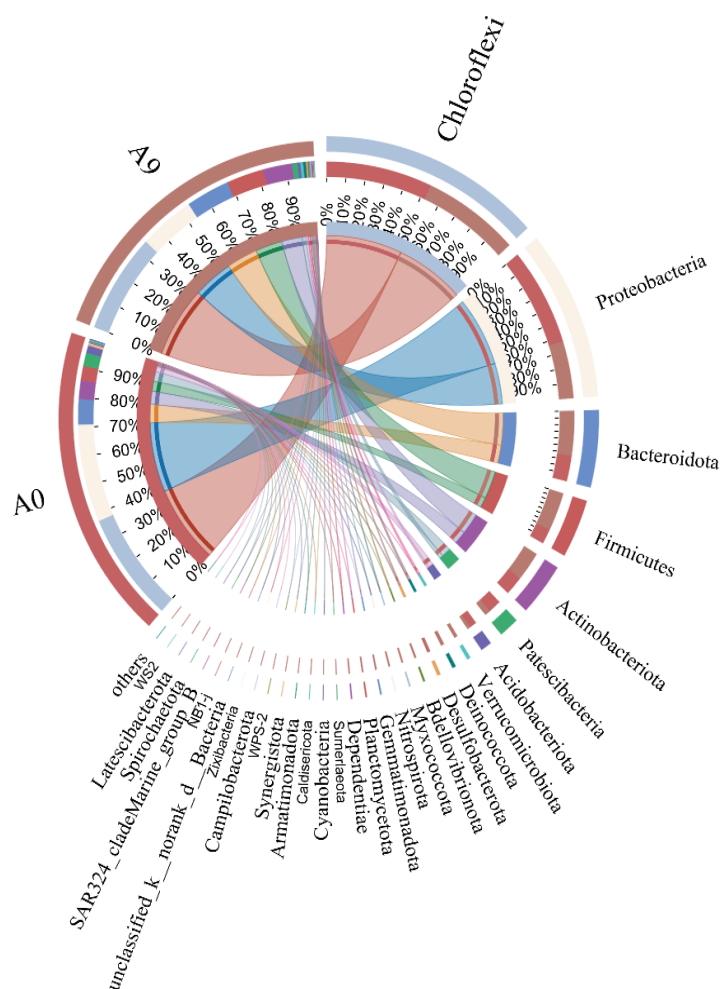


Figure 10. Genus-level heatmap showing the taxonomic succession and relative abundance of microbial communities before (A0) and after (A9) anoxic-aerobic digestion



To further elucidate the internal relationships within the microbial community, hierarchical clustering and association visualizations were performed. The heatmap analysis (Figure 10) provided a high-resolution view of the distribution patterns of the top 35 microbial genera. A clear bifurcation in the clustering tree was observed between A0 and A9, indicating that the anoxic-aerobic treatment significantly reshaped the community hierarchy at the genus level. Notably, the heatmap confirmed the drastic depletion of acinetobacter and the concurrent enrichment of functional groups such as dechloromonas and various members of the saprospiraceae family in the A9 sample, suggesting a coordinated response to the digestion process. The circos diagram (Figure 11) visualized the distribution of major phyla across the experimental samples and the degree of contribution from each taxon. The diagram illustrated that chloroflexi and proteobacteria maintained the dominant representation (visualized as broad ribbons) toward both A0 and A9, reinforcing their roles as the core functional backbone. Consistent with the taxonomic shifts, the ribbon thickness for Bacteroidetes widened significantly toward the A9 sample, visually confirming its proliferation. This holistic visualization integrates the sample-species relationships, demonstrating that while the core phyla remained stable, the fine-tuning of microbial proportions at the genus level was the key driver for successful sludge stabilization and sanitization.

Figure 11. Circos diagram illustrating the distribution and relative abundance of major microbial taxa across sludge samples A0 and A9. The ribbons (flows) represent the correspondence and contribution of each phylum to the respective samples, with the width of the bands indicating relative abundance



4. Discussion

4.1 Mechanisms of Heavy Metal Passivation and Potential Role of Organic Transformation

The observed “release-passivation” pattern of heavy metals in this study (Section 3.1) underscores the dynamic phase transfer driven by the digestion process. The initial surge in liquid-phase metals is attributed to the biodegradation of extracellular polymeric substances (EPS) and the hydrolysis of the organic matrix, which releases bound metal ions. However, the

subsequent decline and passivation in the solid phase are strongly linked to the stabilization of organic matter. The migration of Zn and Cu to the solid phase (Figure 3) implies the formation of stable complexes. Although the chemical structure of organic matter (e.g., humic substances) was not directly characterized in this study, the biological data offers insight into the underlying mechanism. The dominant phylum Chloroflexi, which maintained high abundance (34.08%), has been widely reported as a key player in the hydrolysis of organic matter and the synthesis of humic acid precursors^[31,32]. The enzymatic polymerization mediated by such functional microbes can generate carboxyl and phenolic groups, which act as powerful chelators for divalent cations^[33-35]. Therefore, it is inferred that the microbial succession promoted the transformation of organic substrates into stable complexes, likely humic-like substances, which immobilized the heavy metals [36]. This suggests that biologically mediated organic transformation is a crucial pathway for detoxification in this system^[37,38].

4.2 Nutrient Recovery Potential

The synchronous variations of nitrogen and phosphorus (Section 3.2) reveal migration mechanisms modulated by microbial metabolic activities under alternating redox conditions. The significant accumulation of soluble phosphate in the liquid phase (peaking at $\sim 2.9 \text{ mg}\cdot\text{L}^{-1}$) creates a highly concentrated stream ideal for resource recovery. Advanced technologies can leverage these enriched supernatants to reclaim phosphorus in the form of struvite or vivianite, thereby realizing the high-value utilization of sewage sludge^[39-41]. Furthermore, the stabilization of nitrogen in the solid residues ensures that the dewatered sludge retains nutrient value for potential land application as a soil conditioner.

4.3 Microbial Functional Succession and Pathogen Removal

The reshaping of the microbial community (Section 3.3) provides the biological basis for the observed chemical changes^[42]. The alternating anoxic-aerobic environment imposes a severe “metabolic selection pressure,” acting as a biological filter. This is evidenced by the drastic reduction of the opportunistic pathogen *Acinetobacter* (Section 3.3), likely due to its inability to adapt to the periodic anoxic stressors and competition from adapted taxa. While some *Acinetobacter* species are known Polyphosphate Accumulating Organisms (PAOs), the *Acinetobacter* detected here were predominantly associated with pathogenic groups, and their removal outweighs the potential loss in biological phosphorus removal capacity. Furthermore, the loss of biological P-removal is likely compensated by chemical precipitation and organic chelation mechanisms contributing to P retention in the solid phase. The succession favors versatile functional phyla like Chloroflexi and Proteobacteria^[8]. Chloroflexi, serving as the skeletal structure of sludge flocs, plays a pivotal role in hydrolyzing complex organic substrates. Meanwhile, specific genera within Proteobacteria (e.g., *Dechloromonas*) facilitate nutrient removal and denitrification^[43]. Moreover, the accumulated stable organic substances can act as electron shuttles to facilitate interspecies electron transfer, enhancing the overall stability of the digestion system. From a life cycle perspective, this biological selection not only improves sludge safety by degrading antibiotic resistance genes but also supports carbon footprint reduction through soil carbon sequestration when the treated sludge is applied to land^[44].

5. Conclusions

This study establishes a quantitative relationship between process parameters, microbial community succession, and heavy metal fate within an anoxic-aerobic SBR system. The results demonstrate that the proposed process effectively drives the migration of typical heavy metals, particularly Zn, from the bioavailable liquid phase to the stable solid phase, thereby achieving efficient heavy metal passivation and mitigating the environmental risks associated with sludge. Simultaneously, the substantial accumulation of liquid-phase phosphate offers a promising pathway for high-value nutrient recovery. The alternating redox environment exerts significant selective stress, which reduces overall microbial diversity but successfully suppresses opportunistic pathogenic genera such as *Acinetobacter*, thereby enhancing the biosafety of the digested sludge. These performance improvements are primarily attributed to the enrichment of functional phyla, notably Chloroflexi and Proteobacteria, which facilitate the humification of organic matter and further promote the chelation and immobilization of heavy metals. Overall, anoxic-aerobic digestion represents a highly reliable technical strategy for the integrated detoxification, stabilization, and resource recovery of municipal sludge. Future research should prioritize parameter optimization for pilot-scale operations and conduct long-term assessments regarding the agricultural and environmental risks of applying treated sludge to land.

Funding

This research was funded by the National Natural Science Foundation of China Youth Program (82304889), Shaanxi Fundamental Science Research Project for Mathematics and Physics (Grant No. 25JSQ029), and Shaanxi College Students' Innovation Training Program (Project No. S202610709106).

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

- [1] Yue, W., Yu, W., Cao, Z., Wang, K., Chen, F., & Si, X. (2024). Current status and trends of municipal sludge treatment and disposal in China. *Global NEST Journal*, 26. <https://doi.org/10.30955/gnj.06161>
- [2] Zhang, J., Shao, Y., Wang, H., Liu, G., Qi, L., Xu, X., ... & Liu, S. (2021). Current operation state of wastewater treatment plants in urban China. *Environmental Research*, 195, 110843. <https://doi.org/10.1016/j.envres.2021.110843>
- [3] Jin, F., Lu, J., Sun, F., Yang, F., & Li, Z. (2025). Application and development of sludge-based materials for environmental pollution remediation: a bibliometric review from 2004 to 2024. *RSC Advances*, 15, 8072-8087
- [4] Han, W., Jin, P., Chen, D., Liu, X., Jin, H., Wang, R., & Liu, Y. (2021). Resource reclamation of municipal sewage sludge based on local conditions: A case study in Xi'an, China. *Journal of Cleaner Production*, 316, 128189. <https://doi.org/10.1016/j.jclepro.2021.128189>
- [5] Yu, H., Li, P., Bo, G., & Shen, G. (2024). Studies on the humic acid structure and microbial nutrient restriction mechanism during organic-inorganic co-composting. *Journal of Environmental Management*, 353, 120186. <https://doi.org/10.1016/j.jenvman.2024.120186>
- [6] Wang, H., Chen, S., Gu, J., Liu, Y., Yang, G., Su, W., ... & Yu, R. (2023). Biolysed Sludge Composting for Nitrogen Conservation and Humification Improvements and Mechanisms. *Sustainability*, 15, 10119
- [7] Liu, R., Hao, X., van Loosdrecht, M. C., Zhou, P., & Li, J. (2019). Dynamics of humic substance composition during anaerobic digestion of excess activated sludge. *International Biodeterioration & Biodegradation*, 145, 104771. <https://doi.org/10.1016/j.ibiod.2019.104771>
- [8] Pan, X., Luo, L., Wang, H., Chen, X., Zhang, Y., Dai, Y., & Luo, F. (2025). Developments in Microbial Communities and Interaction Networks in Sludge Treatment Ecosystems During the Transition from Anaerobic to Aerobic Conditions. *Microorganisms*, 13, 2178. <https://doi.org/10.3390/microorganisms13092178>
- [9] Zhong, Y., Yang, W., Zhuo, Q., Cao, Z., Chen, Q., & Xiao, L. (2024). Research Progress on Heavy Metal Passivators and Passivation Mechanisms of Organic Solid Waste Compost: A Review. *Fermentation*, 10, 88. <https://doi.org/10.3390/fermentation10020088>
- [10] Shamsollahi, H. R., Alimohammadi, M., Momeni, S., Naddafi, K., Nabizadeh, R., ... & Yousefi, M. (2019). Assessment of the health risk induced by accumulated heavy metals from anaerobic digestion of biological sludge of the lettuce. *Biological Trace Element Research*, 188, 514-520. <https://doi.org/10.1007/s12011-018-1422-y>
- [11] Dong, B., Liu, X., Dai, L., & Dai, X. (2013). Changes of heavy metal speciation during high-solid anaerobic digestion of sewage sludge. *Bioresource Technology*, 131, 152-158. <https://doi.org/10.1016/j.biortech.2012.12.112>
- [12] Zdeb, M., Pawłowska, M., & Pacan, J. (2020). The Influence of anaerobic digestion on selected heavy metals fractionation in sewage sludge. *Journal of Ecological Engineering*, 21
- [13] Wang, Y., Tang, T., He, Q., & Huangfu, X. (2025). Insights into the cold adaptability, nitrogen, and phosphorus metabolism mechanism of active sludge in response to low temperature (4° C). *Journal of Cleaner Production*, 537, 147225. <https://doi.org/10.1016/j.jclepro.2025.147225>
- [14] Lu, D., Yu, B., He, J., Dai, H., Wang, Z., & Li, H. (2025). Novel insight on humic acid addition into waste activated sludge to enhance the production of short-chain fatty acids during anaerobic fermentation. *Fuel*, 397, 135372. <https://doi.org/10.1016/j.fuel.2025.135372>
- [15] Liu, C., Tong, Q., Li, Y., Wang, N., Liu, B., & Zhang, X. (2019). Biogas production and metal passivation analysis

- during anaerobic digestion of pig manure: Effects of a magnetic Fe₃O₄/FA composite supplement. *RSC Advances*, 9, 4488-4498
- [16] Han, W., Chen, S., Tan, X., Li, X., Pan, H., Ma, P., ... & Xie, Q. (2022). Microbial community succession in response to sludge composting efficiency and heavy metal detoxification during municipal sludge composting. *Frontiers in Microbiology*, 13, 1015949. <https://doi.org/10.3389/fmicb.2022.1015949>
- [17] Tang, H., Xiang, G., Xiao, W., Yang, Z., & Zhao, B. (2024). Microbial mediated remediation of heavy metals toxicity: mechanisms and future prospects. *Frontiers in Plant Science*, 15, 1420408. <https://doi.org/10.3389/fpls.2024.1420408>
- [18] Guan, W., Ansari, A. J., Yin, R., Qi, C., & Song, X. (2024). Optimizing feedstock organic composition to regulate humification and heavy metal passivation during solid-state anaerobic digestion. *Chemical Engineering Journal*, 499, 156071. <https://doi.org/10.1016/j.cej.2024.156071>
- [19] Zhao, Y., An, P., Wan, J., & Zhang, X. (2024). Enhanced Simultaneous Nitrogen and Phosphorus Removal in a Continuous-Flow Granular Sludge System under Gradient-Controlled Hydraulic Loading. *Water*, 16, 1510. <https://doi.org/10.3390/w16111510>
- [20] Yu, H., Yang, Y., Wu, X., Yuan, W., Wang, H., Deng, H., ... & Li, R. (2022). Phosphorus migration and transformation in activated sludge by ultrasonic treatment. *Sustainability*, 14, 11224. <https://doi.org/10.3390/su141811224>
- [21] Chen, H., Deng, F., & Zhang, L. (2022). Phosphorus recovery from sludge by rusty scrap iron enhanced anaerobic digestion and vivianite crystallization. *Journal of Water Process Engineering*, 47, 102697. <https://doi.org/10.1016/j.jwpe.2022.102697>
- [22] Sun, H., Yuan, Y., Zhang, L., & Liu, Y. (2025). Phosphorus Recovery From Wastewater Through Anaerobic Digestion Under Sub-Supersaturation Conditions. *Water Environment Research*, 97, e70161. <https://doi.org/10.1002/wer.70161>
- [23] Zou, Y., & Axmacher, J. C. (2020). The Chord-normalized expected species shared (CNESS)-distance represents a superior measure of species turnover patterns. *Methods in Ecology and Evolution*, 11, 273-280. <https://doi.org/10.1111/2041-210X.13333>
- [24] Dorazio, R. M., Li, P., Xiao, L., & Kachel, S. M. (2025). An evaluation of multi-species occupancy models with correlated species occurrences. *Methods in Ecology and Evolution*, 16, 2681-2693. <https://doi.org/10.1111/2041-210x.70168>
- [25] Ning, X., Hu, J., Yue, J., Tang, T., & Zhang, B. (2024). Microbial community structure of an anaerobic side-stream coupled anoxic-aerobic membrane bioreactor (AOMBR-ASSR) for an in-situ sludge reduction process. *Bioprocess and Biosystems Engineering*, 47, 1027-1037. <https://doi.org/10.1007/s00449-024-03019-y>
- [26] Fu, X., Zuo, H., Weng, Y., Wang, Z., Kou, Y., ... & El-Din, M. G. (2024). Performance evaluation and microbial community succession analysis of co-composting treatment of refinery waste activated sludge. *Journal of Environmental Management*, 370, 122872. <https://doi.org/10.1016/j.jenvman.2024.122872>
- [27] Bovio-Winkler, P., Guerrero, L. D., Erijman, L., Oyarzúa, P., Suárez-Ojeda, M. E., Cabezas, A., & Etchebehere, C. (2023). Genome-centric metagenomic insights into the role of Chloroflexi in anammox, activated sludge and methanogenic reactors. *BMC Microbiology*, 23, 45. <https://doi.org/10.1186/s12866-023-02765-5>
- [28] Speirs, L. B., Rice, D. T., Petrovski, S., & Seviour, R. J. (2019). The phylogeny, biodiversity, and ecology of the Chloroflexi in activated sludge. *Frontiers in Microbiology*, 10, 2015
- [29] Guo, N., Wang, Y., Tong, T., & Wang, S. (2018). The fate of antibiotic resistance genes and their potential hosts during bio-electrochemical treatment of high-salinity pharmaceutical wastewater. *Water Research*, 133, 79-86. <https://doi.org/10.1016/j.watres.2018.01.020>
- [30] Zheng, L., Ren, M., Xie, E., Ding, A., Liu, Y., Deng, S., & Zhang, D. (2019). Roles of phosphorus sources in microbial community assembly for the removal of organic matters and ammonia in activated sludge. *Frontiers in Microbiology*, 10, 1023. <https://doi.org/10.3389/fmicb.2019.1023>
- [31] Cai, S., Zhang, Y., Hu, A., Liu, M., Wu, H., Wang, D., & Zhang, W. (2023). Dissolved organic matter transformation mechanisms and process optimization of wastewater sludge hydrothermal humification treatment for producing plant

- biostimulants. *Water Research*, 235, 119910. <https://doi.org/10.1016/j.watres.2023.119910>
- [32] Zhang, Z., Duan, C., Liu, Y., Li, A., Hu, X., Chen, J., ... & Li, S. (2023). Green waste and sewage sludge feeding ratio alters co-composting performance: emphasis on the role of bacterial community during humification. *Bioresource Technology*, 380, 129014. <https://doi.org/10.1016/j.biortech.2023.129014>
- [33] Zhang, S., Zhang, Q., Gao, H., Wang, L., Song, C., Tang, G., ... & Hu, X. (2024). Effects of adding steel slag on humification and characteristics of bacterial community during phosphate-amended composting of municipal sludge. *Bioresource Technology*, 394, 130229. <https://doi.org/10.1016/j.biortech.2023.130229>
- [34] Sun, Y., Liu, H., Wen, J., Xiao, H., Shi, M., Lu, X., ... & Cui, M. (2024). Influences of released humic acids during thermal hydrolysis on sludge anaerobic digestion: New insights from the molecular weight of humic acids. *Journal of Environmental Management*, 370, 122555
- [35] He, Z.-W., Wang, F., Zou, Z.-S., Tang, C.-C., Zhou, A.-J., Liu, W., ... & Wang, A. (2023). Recent advances and perspectives in roles of humic acid in anaerobic digestion of waste activated sludge. *Chemical Engineering Journal*, 466, 143081. <https://doi.org/10.1016/j.cej.2023.143081>
- [36] Gao, J., Chang, H., Wu, S., Zhang, X., Shaaban, M., & Shi, Z. (2025). Variations of dissolved organic matter vertically in the scenario of sludge composting acting as nitrogen fertilizers: perspectives of spectral implications. *Journal of Environmental Management*, 393, 126903
- [37] Yang, X., Peng, X., Feng, K., Wang, S., Zou, X., & Deng, Y. (2025). Organic molecular network analysis reveals transformation signatures of dissolved organic matter during anaerobic digestion process. *Water Research*, 123777. <https://doi.org/10.1016/j.watres.2025.123777>
- [38] Chi, Z., Zhang, S., Li, Y., & Li, H. (2025). Ultra-high temperature bacterial agents enhance heavy metal passivation and antibiotic degradation in compost. *Frontiers in Microbiology*, 16, 1708982
- [39] Liu, L., Wang, Y., Wang, J., He, C., & Zhang, J. (2025). Recovery of nitrogen and phosphorus in sewage sludge treatment technologies. *Chinese Chemical Letters*, 111431. <https://doi.org/10.1016/j.cclet.2025.111431>
- [40] Zhu, Y., Wen, X., & Guo, Z. (2025). Research progress on high-value utilization technology of sludge solid waste in China. *Journal of Material Cycles and Waste Management*, 27, 654-665. <https://doi.org/10.1007/s10163-024-02102-6>
- [41] Bi, S., Yang, X., Lan, R., Cheng, X., & Sun, D. (2024). Separation of vivianite from digested sludge by flotation using oleic acid as a collector. *Separation and Purification Technology*, 344, 127257. <https://doi.org/10.1016/j.seppur.2024.127257>
- [42] Domini, M., & Bertanza, G. (2025). The Role of Alternate Oxidic–Anoxic Cycles in Full-Scale Sludge Stabilization for Energy Savings. *Energies*, 18, 2514. <https://doi.org/10.3390/en18102514>
- [43] Zhang, W., Yin, Z., Gu, R., Liu, C., Xu, N., Sun, Z., ... & Niu, X. (2023). Anaerobic digestion of municipal sewage sludge integrated with brewery wastewater treatment: importance of temperature and mixing ratio. *Water*, 15, 2902. <https://doi.org/10.3390/w15162902>
- [44] Xu, C., Chen, W., & Hong, J. (2014). Life-cycle environmental and economic assessment of sewage sludge treatment in China. *Journal of Cleaner Production*, 67, 79-87. <https://doi.org/10.1016/j.jclepro.2013.12.002>

Dear Researchers and Scholars :

Greetings from Asia Pacific Science Press, a beacon of academic and scientific publishing, located in the vibrant city of Hong Kong.

We extend our heartfelt gratitude for your relentless pursuit of knowledge, and your significant contributions to the advancement of science and society. It is researchers and scholars like you who propel humanity forward, and we at the Asia Pacific Science Press are devoted to ensuring that your groundbreaking works receive the global recognition they rightfully deserve.

In light of our commitment to disseminating pioneering research across various disciplines, such as medicine, architecture, education, and electronics, we are reaching out with two pivotal opportunities to augment our collaboration with the global academic community:

Call for Paper Submissions:

We cordially invite you to submit your original research articles to our fast-growing, peer-reviewed, and open-access journals. Our platform guarantees an extensive, global reach, enabling your work to garner maximum visibility and citation in the academic sphere. Rest assured, your work will be meticulously assessed by experts in the field, ensuring it receives the acknowledgment and exposure it merits.

Join Our Esteemed Team:

We are fervently searching for passionate researchers and scholars interested in joining our burgeoning team at Asia Pacific Science Press. We offer numerous roles, such as peer reviewers, editors, and advisory board members, where your expertise will significantly shape the content and quality of our publications. In return, you will gain invaluable experience, network with preeminent scholars, and play a pivotal role in molding the future of global academic publishing.

Why Choose Asia Pacific Science Press?

Global Reach: Your work will be accessible to a worldwide audience, free from any access barriers.

Collaboration with Renowned Universities: We have established extensive publishing systems in cooperation with world-renowned universities, such as Wuhan University, Hong Kong University, and the University of Malaya.

Diverse Disciplines: Your research will be housed among numerous journals across a multitude of academic projects and disciplines.

As we stride forward in the academic landscape, we envision a future where our collective efforts shape a more enlightened, innovative, and interconnected global society. We sincerely hope that you consider this invitation to join us on this auspicious journey towards knowledge, discovery, and global impact.

Should you wish to submit your work or express interest in joining our team, please do not hesitate to contact us. You can submit your manuscript or personal profile to info@apspublisher.com or visit our website at www.apspublisher.com for more information.

Thank you for considering this opportunity, and we eagerly anticipate the possibility of welcoming you to the Asia Pacific Science Press family. Together, let's forge a future of unparalleled scientific advancement and discovery.

Warm regards

Asia Pacific Science Press

OUR JOURNALS

Asia Pacific Economic and Management Review is an international, peer-reviewed and open access journal which focuses on theoretical and applied studies of corporate and financial behavior. Aiming to promote the research in fields of business economics and management, it covers mainly but not limits to the following areas: accounting and financial management, economics, human resource management and organizational behavior, information management, international business, strategy and innovation, management science and operations management, marketing and retailing, finance.



Critical Humanistic Social Theory is an journal that publishes papers specifically using quantitative or qualitative research methods for social science research. The journal encourages scholars to conduct social science theory research from the perspective of social critical theory and emphasizes research concerned with issues or methods that cut across traditional disciplinary lines.



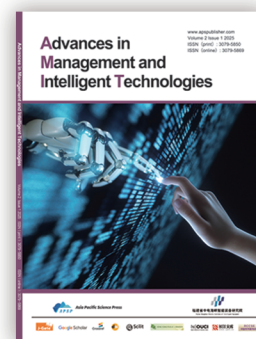
Journal of Educational Theory and Practice is an international, peer-reviewed and open access journal which is to promote the evaluative, integrative, theoretical and methodological research on contemporary education; shape a novel, broader view of issues in contemporary education; enhance the caliber of humanities research through active use of best domestic and foreign practices; and integrate the achievements of various sciences and knowledge areas with unconventional approaches.



Journal of Advances in Engineering and Technology is an international, peer-reviewed and open access journal which publishes original articles, reviews, short communications, case studies and letters in the field of electronic research and application.



Advances in Management and Intelligent Technologies is an international, peer-reviewed, open-access academic journal, hosted by the Fujian Strait Institute of Intelligent Equipment and managed and published by Asia-Pacific Science Press. It focuses on the latest research in the fields of management and intelligent technologies, and aims to advance both theoretical and applied research in management, technological innovation, and intelligent development.



Asia Pacific Journal of Clinical Medical Research is an international, peer-reviewed, open access journal dedicated to advancing clinical medical research across multiple disciplines. The journal serves as a platform for publishing high-quality original research, reviews, and clinical studies that enhance the understanding of medical practices, treatment innovations, and healthcare outcomes, thereby supporting patient care and medical advancements in the Asia Pacific region and beyond.



Asia Pacific Journal of Educational Research is an international, peer-reviewed, open-access academic journal focusing on educational theory and practice. It publishes high-quality research on educational reform, teaching methods, educational equity, and policy studies. The journal addresses practical needs and institutional changes in the education systems of the Asia-Pacific region, advocating a balance between theoretical inquiry and practical experience. It encourages original studies from multicultural, comparative, and interdisciplinary perspectives, aiming to support educational innovation and policy development across the region.



Asia Pacific Economic and Social Development is an international, peer-reviewed, open-access academic journal openly distributed to the global academic community. The journal is committed to publishing original research with theoretical depth and practical value in the fields of economic and social development. It focuses on issues such as economic behavior, social structure transformation, policy innovation, and regional coordinated development in the Asia-Pacific region. The journal encourages interdisciplinary perspectives and promotes the integration of economics, sociology, management, and related disciplines.

