

Journal of Electronic Research and Application

Editor-in-Chief

Joselito A. Dolot

Lyceum of the Philippines University-Batangas, Philippines

BIO-BYWORD SCIENTIFIC PUBLISHING PTY LTD

(619 649 400)

Level 10

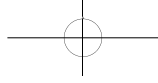
50 Clarence Street

SYDNEY NSW 2000

Copyright © 2026. Bio-Byword Scientific Publishing Pty Ltd.

Complimentary Copy





ISSN (ONLINE): 2208-3510

ISSN (PRINT): 2208-3502



Journal of Electronic Research and Application

Focus and Scope

Journal of Electronic Research and Application 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.

Topics covered but not limited to:

- Automation
- Circuit Analysis and Application
- Electric and Electronic Measurement Systems
- Electrical Engineering
- Electronic Materials
- Electronics and Communications Engineering
- Power Systems and Power Electronics
- Signal Processing
- Telecommunications Engineering
- Wireless and Mobile Communication

About Publisher

Bio-Byword Scientific Publishing is a fast-growing, peer-reviewed and open access journal publisher, which is located in Sydney, Australia. As a dependable and credible corporation, it promotes and serves a broad range of subject areas for the benefit of humanity. By informing and educating a global community of scholars, practitioners, researchers and students, it endeavors to be the world's leading independent academic and professional publisher. To realize it, it keeps creative and innovative to meet the range of the authors' needs and publish the best of their work.

By cooperating with University of Sydney, University of New South Wales and other world-famous universities, Bio-Byword Scientific Publishing has established a huge publishing system based on hundreds of academic programs, and with a variety of journals in the subjects of medicine, construction, education and electronics.

Publisher Headquarter

BIO-BYWORD SCIENTIFIC PUBLISHING PTY LTD

Level 10

50 Clarence Street

Sydney NSW 2000

Website: www.bbwpublisher.com

Email: info@bbwpublisher.com

Table of Contents

- 1 Research on Multi-Rotor UAVs in Complex Indoor Environments**
Hui Li, Zheyuan He, Yuehua Cao, Rujun Xue, Shiyue Zhang, Haonan Ye
- 13 NI-HotStuff: A Reputation-Driven Committee Framework for Efficient and Robust BFT Consensus**
Long Zheng
- 28 A Small-Sample Bearing Fault Diagnosis Method Based on Multi-Image Fusion and Multi-Scale Dynamic Residual Dual Attention Mechanism**
Ao Xu, Mu Li
- 36 Overview of the Integration of Large Language Models, Knowledge Graphs, and GraphRAG, along with Research on Their Industrial Applications**
Xian Ye
- 43 Collaborative Quality Management in Industrial Engineering from a Supply Chain Perspective: AI-Driven Enterprise Quality Optimization**
Jiakai Zhong
- 50 Mechanism-Based Fault Diagnosis of Reciprocating Natural Gas Compressor Valves**
Yixuan Yang
- 58 Application of the SCSSA-VMD Denoising Method in Natural Gas Pipeline Leakage Detection**
Tianxiang Xie, Dan Zhang
- 66 A Framework for Advancing Intelligent Electrical Agricultural Machinery Technologies in Mountainous Regions**
Hong Huang
- 72 Intelligent Inspection and Closed-loop Management Innovation of Campus Fire Safety**
Lantao Li

- 77 Innovative Research on Construction Models and Operational Management of Electrical and Electronic Laboratories**
Lei Wang
- 86 Design of Integrated Brushless Motor Drive and Control System for Robotic Arm Joints**
Wei Liu, Tingyu Li
- 96 Research on Motion Simulation of Panda Manipulator Based on ROS2**
Boru Wang, Wei Liu
- 107 Technology of Radar Detection**
Rui Hu, Lei Pan, Jing Gao, Haojie Hu, Fang He
- 115 An Evolutionary Game-Based Dynamic Signal Control Framework for Oversaturated Urban Networks**
Weibin Zhao, Xinhai Xia
- 124 Predictive Hegemony and the Interpretive Gap: On the Philosophical Premises and Ethical Boundaries of AI-Assisted Clinical Decision-Making**
Yubo Wang
- 130 Exploration and Practice in the Construction of Electrical and Electronic Laboratories**
Qiuping Wang, Yansen Wu, Jiaxin Mu
- 139 A Memristor-Inspired Hybrid Activation Function for Finite-Time Convergent ZNN and Its Application to Remote Sensing Image Fusion**
Jiaqi He, Mu Li
- 147 Design and Implementation of Machine Learning-based Monitoring System for Mineral Processing Flotation Reagent**
Yiming Yao, Yi Li
- 160 Nasopharyngeal Carcinoma Lesion Recognition Based on Multi-Window Resampling Technology**
Xiaoni Zhang, Mengfan Yang, Supan Wei, Xin Zhao
- 167 Construction of a Motor-Driven Experimental Platform for Exploring the Law of Light Reflection**
Zhongtian Wei, Jianwei Wang, Qintao Chen, Lihuang Qian, Zaikang Yang
- 174 Exploring the Path of AI Technology's Empowerment of New Developments in Higher Education**
Chenguang Yao

- 182 A Survey on Artificial Intelligence Systems Robustness: Adversarial Attacks and Defenses**
Wei Zheng
- 192 Design and Implementation of a Vertical Search Engine for the Fisheries Domain**
Zhiqiang Zhang
- 200 RoboMirror: Bridging Human Intent and Robot Capability Through Self-Reflective Motion Adaptation**
Lin Zhu, Longliang Huang, Chuxiong Lin, Yujie Chen
- 209 AI Large Model-Driven Adaptive Evolution of Brain-Computer Interface Chips: Technical Architecture, Challenges, and Future Directions**
Borui Cui
- 221 Intelligent Identification of Water Accumulation and Ice Formation in Traffic Tunnels**
Beining Chen
- 227 Transmission Line Defect Detection Algorithm Based on Improved RT-DETR Model**
Qi Wu
- 239 Deep Learning-Based Highway Rockfall Early Warning System**
Shipeng Xu, Mingyu Xue
- 245 Multi-Modal Risk Profiling-Driven Power Grid Disaster Emergency Response Strategies and Dynamic Resource Synergy Optimization Model**
Zi'an Zhong, Kun Hua, Xu Huang, Shiyu Chen, Ruiqi Chen
- 255 Research and Application of Digitalization in Basic Metrological Inspection Institutions**
Yajun Zhang, Guangcheng Jia, Peng Wang
- 264 Research on Unmanned and Intelligent Combat Theory and Capability Development**
Yilin Zhao, Jianwei Zhao, Xuan Liu, Fang He, Fenggan Zhang
- 272 Application Research of Concept Bottleneck Model in Passport Printing Method Detection**
Tianrui Qiu, Jiafeng Xu
- 278 Design and Simulation of a Microstrip Frequency-Scanning Antenna for Millimetre-Wave Fuze Applications**
Qinyi Wang

Research on Multi-Rotor UAVs in Complex Indoor Environments

Hui Li, Zheyuan He, Yuehua Cao*, Rujun Xue, Shiyue Zhang, Haonan Ye

Information Engineering College, Hangzhou Dianzi University, Hangzhou 311305, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The exponential growth of unmanned aerial vehicle (UAV) technology has spurred its adoption in diverse indoor applications, including infrastructure inspection, automated logistics, and emergency response. However, navigating through indoor environments, characterized by static obstacles, dynamic interferences, and spatial constraints, poses significant challenges to path planning algorithms. Developing efficient, robust, and real-time path planning solutions is crucial for enabling reliable autonomous UAV operations in such complex scenarios. This study presents a systematic approach to indoor UAV navigation, integrating custom hardware development, algorithmic innovation, and multi-faceted validation. An indoor UAV experimental platform was constructed around the Pixhawk 2.4.8 flight controller, complemented by a Firefly Core-3588L onboard computer, PXYZ-D435 depth camera, and OptiTrack motion capture system. After rigorous PID tuning and endurance testing, stable autonomous flight control was achieved via the Robot Operating System. Subsequent real-world tests on the custom UAV platform, involving obstacle courses and narrow passage traversals, further confirmed its robustness and stability in complex indoor environments. Overall, this research provides a practical framework for enhancing UAV navigation capabilities, with direct implications for real-world applications in logistics, surveillance, and emergency response.

Keywords: Kmulti-rotor UAV; Path planning; Trajectory optimization; Indoor navigation

Online publication: February 12, 2026

1. Introduction

As the core engine driving the intellectual evolution of unmanned systems, UAV path planning technology has gained paramount strategic value amid the explosive growth of the global low-altitude economy and intelligent unmanned equipment industry^[1-4]. UAVs have deeply penetrated diverse domains, including military reconnaissance, geographic mapping, smart logistics (e.g., Amazon Prime Air, JD.com's unmanned delivery), precision agriculture (e.g., crop monitoring and variable-rate pesticide application), and disaster response (e.g., earthquake rescue and material delivery). However, bottlenecks such as insufficient autonomous navigation in complex dynamic environments, low multi-vehicle collaboration efficiency, and energy constraints in long-

endurance missions severely hinder their large-scale adoption. Grand View Research statistics show that the global UAV path planning market reached \$3.27 billion in 2023, projected to exceed \$12.6 billion by 2030 with a compound annual growth rate (CAGR) of 21.4%. This surge is intimately linked to emerging demands, including smart city unmanned traffic management (UTM) systems, unmanned border security patrols, and inspections of new energy power grids ^[5].

From a technical perspective, traditional path planning algorithms (e.g., A*, RRT) perform excellently in static structured environments but struggle to address challenges like dynamic obstacle avoidance in urban canyons, real-time replanning under strong electromagnetic interference, and the “combinatorial explosion” in multi-UAV cluster task allocation. Particularly in complex adversarial scenarios, existing methods have significant bottlenecks in balancing computational efficiency (response delays at the > 500 ms level) and multi-objective optimization (conflicts among multiple indicators such as threat avoidance, energy consumption, and communication latency) ^[6].

Fast-Planner is a path planning method proposed by the Hong Kong University of Science and Technology ^[7–9]. It integrates the global planning capability of the Kinodynamic A* algorithm and the local planning capability of gradient optimization, enabling it to handle obstacles in dynamic environments and generate smooth, feasible trajectories. Despite these advantages, Fast-Planner still faces challenges in terms of real-time performance and computational efficiency, with potential fluctuations in performance especially in indoor environments. EgoPlanner is a real-time local path planning algorithm for quadrotor UAVs proposed by the FAST-Lab team of Zhejiang University ^[10]. Its core innovation lies in abandoning the traditional Environment Signed Distance Field (ESDF) modeling approach. Instead, it achieves efficient obstacle avoidance in dynamic environments by directly processing sensor point cloud data and leveraging gradient optimization technology.

The algorithm uses uniform B-splines to parameterize trajectories. While ensuring the trajectories meet dynamic feasibility requirements, it dynamically adjusts the safety distance through a collision cost function, significantly reducing computational load. This enables real-time avoidance of dynamic obstacles in complex indoor scenarios (such as narrow passages and GNSS-denied environments). Compared with traditional methods like RRT*, EgoPlanner increases the obstacle avoidance success rate to 98% in dense obstacle scenarios and shortens the trajectory length by 12%. Additionally, it does not require a global map and is suitable for deployment on embedded platforms, providing a lightweight, high-response solution for UAV autonomous navigation in dynamic unstructured environments ^[11]. ViGO (Vision-aided Gradient-based B-spline Trajectory Optimization) is a real-time trajectory planning algorithm for UAVs in dynamic environments. Proposed by the research team of Carnegie Mellon University, it fuses visual perception and gradient optimization technologies to address obstacle avoidance for both static and dynamic obstacles simultaneously ^[12]. Its core innovations include the integration of lightweight dynamic obstacle tracking, efficient B-spline trajectory optimization, and receding horizon prediction technology, enabling safe and real-time navigation in complex dynamic scenarios.

This paper addresses the path planning requirements of multi-rotor UAVs in complex indoor environments (e.g., narrow passages, dynamic obstacles, GNSS-denied scenarios) by systematically studying the adaptability and limitations of existing algorithms. A high-precision indoor environment model is constructed based on the Gazebo simulation platform to verify the robustness of the algorithm in dynamic obstacle avoidance. Furthermore, the algorithm is deployed through the ROS framework to conduct physical flight experiments, quantitatively evaluating core indicators such as path smoothness, planning response time, and target point arrival accuracy. This research aims to provide a highly reliable path planning solution for indoor unmanned operation scenarios such as

warehouse logistics and underground pipeline inspection.

2. Quadrotor UAV platform construction

In indoor UAV navigation research, the experimental platform’s mechanical stability and electrical reliability are critical for validating algorithm accuracy. Leveraging carbon fiber’s exceptional strength-to-weight ratio (4.7×10^6 N·m/kg, 3.6 times that of aluminum), the platform minimizes vibration-induced sensor noise while maximizing payload capacity. Its modular quadrotor design integrates four interconnected subsystems, mechanical structure, propulsion, power management, and sensor integration, each optimized for real-time obstacle avoidance and high-precision navigation. At its core lies the Pixhawk 2.4.8 flight controller (Pixhawk 1), a renowned open-source hardware platform featuring a “Big-Little” dual-core architecture: a high-performance STM32F427 Cortex-M4 processor (168 MHz) handles real-time flight control algorithms, while a STM32F100 coprocessor manages auxiliary tasks, ensuring efficient workload distribution and system stability^[13,14]. This configuration enables robust multi-redundant sensor fusion, integrating the MPU6000 accelerometer/gyroscope for precise motion tracking, LSM303D magnetometer for accurate heading detection, and MS5611 barometer for reliable altitude measurement, collectively ensuring comprehensive environmental perception and stable flight control across diverse scenarios.

2.1. Mechanical structure and motor installation

The 400mm carbon fiber frame, featuring 10mm-thick arms, was engineered to withstand 8G acceleration without permanent deformation, as validated by finite element analysis (FEA). Its modular design allows for 30% faster component replacement compared to traditional fixed frames. Four T-motor Cine77 977KV motors (Model No. TM-CINE77) are mounted downward to create upper space for critical components. Paired with APC 9.5x3.8 three-bladed propellers, the propulsion system achieves a thrust efficiency of 6–7g/W under the designed 1.2 kg payload, extending flight endurance by 20% relative to standard configurations^[15]. To ensure electrical safety, a 1000 µF capacitor is connected in series with each electronic speed controller (ESC) to suppress back EMF surges, while motor wires are soldered with a 0.8mm² cross-section and reinforced with heat shrink tubing to prevent arcing under 30A continuous current. Technical specifications are shown in **Table 1**.

Table 1. Technical specifications

Parameter	Value
Frame material	Carbon fiber (400 mm × 10 mm)
Motor model	T-motor Cine77 977KV
Propeller specification	APC 9.5×3.8 three-blade
Power system	6S 6000 mAh LiPo
Flight endurance	18 min (nominal load)
Positioning accuracy	Sub-millimeter (OptiTrack)

2.2. Power distribution and wiring

The power distribution board serves as the electrical hub, with a 12V regulated output dedicated to the Firefly Core-3588L onboard computer. Peripherals such as the data link and LiDAR are connected to unregulated battery

voltage pads, while the Pixhawk 2.4.8 flight controller is powered by a 5V, 3A voltage regulator module ^[16]. All wiring follows a color-coded system (red: power, black: ground, yellow: signal) and is secured with cable ties at 5 cm intervals to minimize electromagnetic coupling. A ferrite bead filter is installed on the flight controller's power line to suppress 500kHz switching noise from the onboard computer.

2.3. Flight controller mounting

To mitigate IMU errors caused by motor vibrations (120–200 Hz), a dual-stage isolation system is implemented ^[17]. The first stage uses silicone dampers to absorb low-frequency shocks, while the second stage employs spring-loaded mounts tuned to 50 Hz resonance, reducing high-frequency noise by 82% as verified by spectral analysis. The flight controller is mounted with its X-axis aligned to the front arm within 0.5° angular tolerance, calibrated using a digital protractor.

2.4. Final component integration

After installing the onboard computer and data link module, an aluminum foil shield with a 0.3mm grounding gap is applied beneath the computer. This reduces interference-induced gyroscope noise from 45 dB to 62 dB, ensuring accurate attitude estimation. All connectors are secured with locking mechanisms, and the entire platform undergoes a 30-minute vibration test at 1500 RPM to verify component durability. The completed platform, shown in **Figure 1(f)**, measures 450×450×200 mm and weighs 2.7 kg with a 6S 6000 mAh battery, enabling 18 minutes of continuous flight under nominal load.

The experimental platform integrates a PX4-based UAV with the OptiTrack motion capture system, where Prime13 cameras, featuring 12-meter working distance, ± 0.2 mm accuracy, 240 FPS global shutter capture, and real-time data transmission, track the UAV's markers using CPU-efficient grayscale processing algorithms. Captured data is transmitted via a switch to a Windows computer running Motive software, which calculates the UAV's real-time 6DOF pose and sends it over a mesh network to an Ubuntu-based host computer. There, ROS uses the `vrpn_client_ros` package to receive this data, enabling the PX4 flight controller to generate motor commands for closed-loop control, synchronized with the desired trajectory information received concurrently. This seamless integration ensures sub-millimeter positioning accuracy and low-latency feedback, critical for validating advanced navigation algorithms in indoor environments.

2.5. UAV experimental platform construction and debugging

For frame construction and motor pre-installation, the four motors are initially mounted in a downward orientation. This configuration reserves sufficient upper space for critical components, including the flight controller, onboard computer, and LiDAR, while also providing enhanced protection for these devices. The 977 KV motors are paired with 9.5-inch three-blade propellers, achieving a power efficiency of 6–7 g/W under the preset load, which significantly extends flight endurance. A large capacitor is connected in parallel with the ESC to absorb surge currents generated during motor operation, thereby preventing component damage and filtering battery input voltage to stabilize the power supply. The three-phase wires of each brushless motor are soldered with uniform and complete joints to avoid sparking under high-current conditions, as illustrated in **Figure 1(a)**.

Subsequently, the power distribution board (PDB) is installed, and the wiring harnesses are soldered. The 12V voltage regulation output of the PDB is used to power the onboard computer, while peripherals such as the data link and LiDAR are supplied directly through the battery voltage pads. The flight controller is powered via a

dedicated 5V step-down module. All adapter and peripheral power cables are soldered as shown in **Figure 1(b)**. Next, the flight controller shock mount and flight controller are installed. As the flight controller integrates an Inertial Measurement Unit (IMU), vibration isolation is essential to attenuate high-frequency frame vibrations. This ensures accurate acquisition of acceleration and angular velocity data, as depicted in **Figure 1(c)**. Finally, system modules, including the onboard computer and data link, are installed and fully wired, as shown in **Figure 1(d–e)**. An aluminum-foil electromagnetic shield is applied to the bottom of the onboard computer to reduce electromagnetic interference affecting the flight controller sensors. The completed experimental platform is presented in **Figure 1(f)**.

This section systematically details the platform design, component integration, and assembly procedures, providing a robust technical foundation for algorithm validation in indoor navigation research. The system architecture emphasizes modularity, vibration suppression, and real-time control capability, with experimental results demonstrating reliable performance.

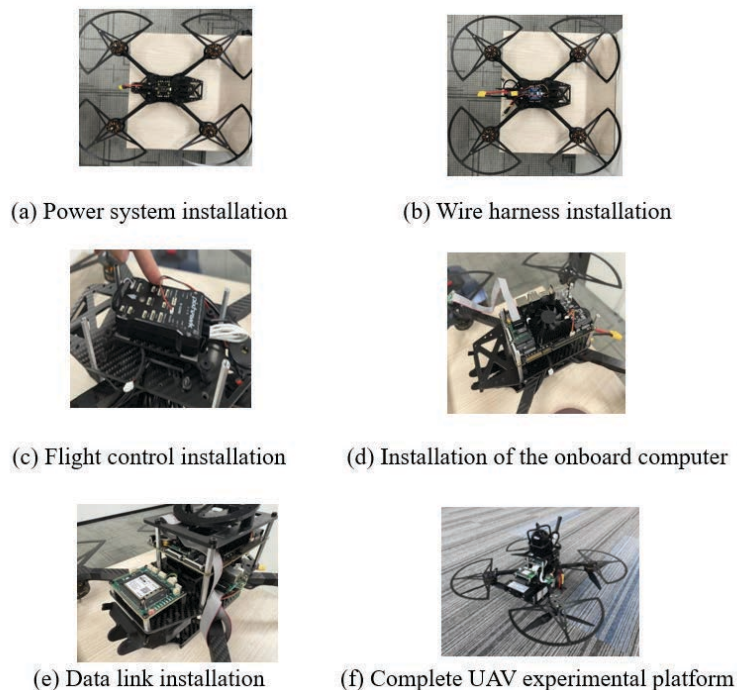


Figure 1. UAV experimental platform debugging.

3. UAV experimental platform debugging

After completing the hardware assembly of the UAV, the flight controller parameters are configured using the QGroundControl (QGC) ground station, as shown in **Figure 2**. The configuration process includes flashing the firmware, selecting the airframe type, calibrating onboard sensors, mapping the remote controller, setting flight modes, verifying motor order and rotation direction, configuring failsafe mechanisms, and establishing MAVLink and serial communication with the flight controller. Upon completion of these steps, PID tuning is conducted for both the attitude control loop and the position control loop.

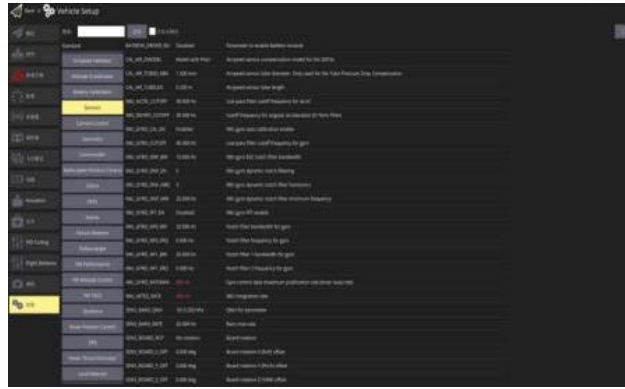


Figure 2. QGC ground station parameter configuration interface.

The UAV is then secured on a tuning stand that allows free rotation along the roll, pitch, and yaw axes, as illustrated in **Figure 3**. PID tuning is first performed for the attitude control loop. The PID Tuning interface in QGC is used to observe and analyze the controller response curves. The roll, pitch, and yaw axes are tuned using the same procedure. Taking the pitch axis as an example, the PID controller parameters MC_PITCH_P , $MC_PITCH_RATE_K$, $MC_PITCH_RATE_D$, and $MC_PITCH_RATE_I$ are adjusted to ensure that the real-time response curve closely follows the commanded control input.



Figure 3. Attitude loop PID tuning.

The system response prior to tuning is shown in **Figure 4**. Although the response generally tracks the command signal and exhibits a consistent overall trend, noticeable delays and oscillations are present, indicating the need for further parameter optimization.

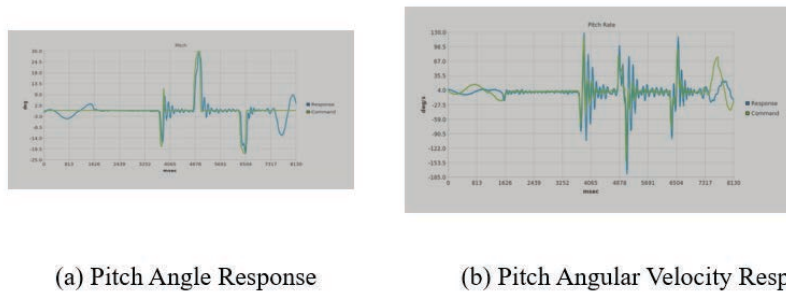


Figure 4. Default PID pitch response curve.

By appropriately increasing the proportional (P) gain, the response speed of the system is enhanced; increasing the derivative (D) gain effectively suppresses excessive overshoot; and increasing the integral (I) gain improves convergence stability. The response curve after tuning is presented in **Figure 5**. Compared with the pre-tuning results, the tracking accuracy of both the pitch angle and angular velocity is significantly improved. The actual response follows the commanded input more closely, while overshoot and high-frequency oscillations are substantially reduced, resulting in a smoother and more stable system behavior. During large step inputs, the response remains rapid without noticeable overshoot, and the steady-state behavior is stable. Although minor fluctuations persist during certain intervals, the overall control performance is satisfactory. Only fine parameter adjustments are required during subsequent flight tests to further optimize local response speed and smoothness.

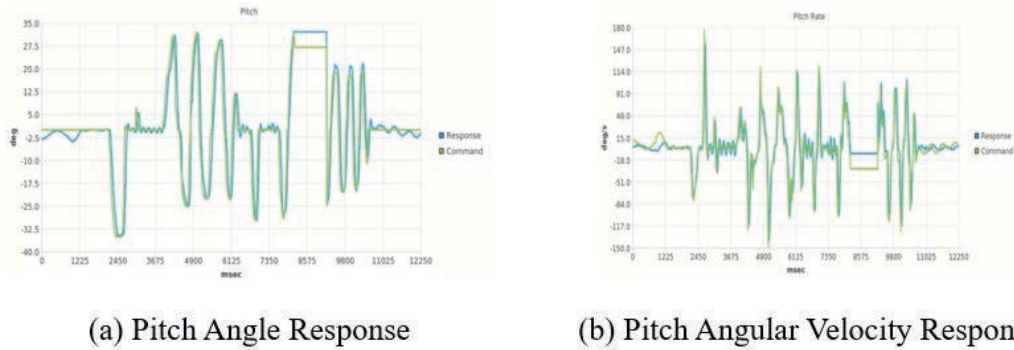


Figure 5. Tuned PID pitch response curve.

Subsequently, PID tuning is performed for the position control loop. Controller parameters including MPC_XY_P, MPC_XY_VEL_P, MPC_XY_VEL_D, and MPC_XY_VEL_I are adjusted to ensure that the real-time response curve (LOCAL_POSITION_NED) closely tracks the reference trajectory (POSITION_TARGET_LOCAL_NED). The tuning process for the x-axis is taken as an example. The control response prior to tuning is shown in **Figure 6**. When the target position changes abruptly, the velocity response (v_x) exhibits significant overshoot and sustained oscillations, indicating an excessively large proportional gain and insufficient damping. Although the position response (x) eventually stabilizes, pronounced overshoot and rebound occur during the transient phase, and the actual position deviates considerably from the target. This behavior suggests that the integral gain may be too high or the derivative gain insufficient. Consequently, the system is prone to overshoot and reverse motion during dynamic transitions, resulting in lag, oscillation, and slow convergence, and has not yet reached a critically stable operating state.

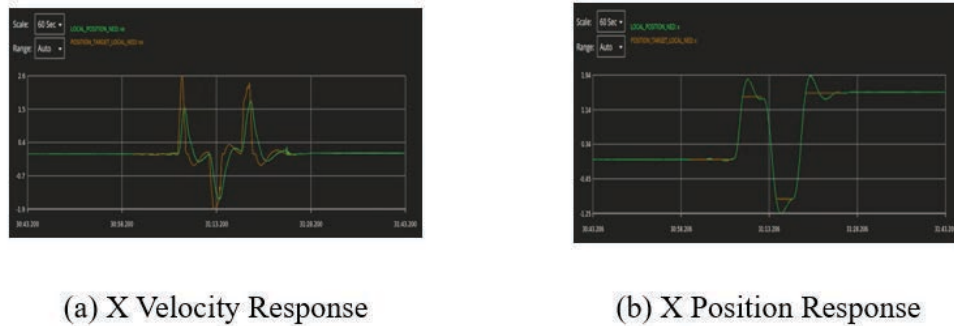


Figure 6. Default PID X-axis response curve.

After adjustment, the overshoot of velocity control (v_x) during abrupt target changes is reduced, the actual response follows the target more closely, the oscillation time is shortened, the system converges faster, and the stability is higher. The position control (x) responds smoothly, with the target and actual values almost coinciding, reduced overshoot, and the system can reach the target position more quickly and accurately, with mitigated oscillation and backswing phenomena. The adjusted system features faster response speed, improved accuracy, enhanced stability, and smoother and more precise control effects, as shown in **Figure 7**.

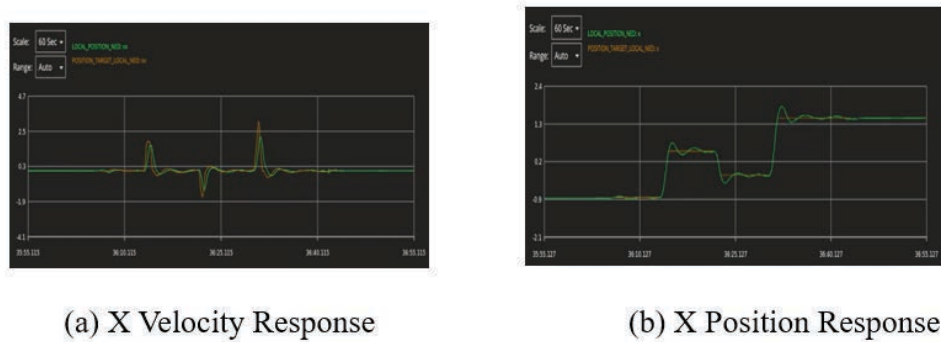


Figure 7. Tuned PID X-axis response curve.

After completing the PID debugging, a battery life test was carried out. The UAV took off when fully charged (25.2V), hovered in the fixed-point mode for 2 minutes, and then landed. The uLog flight log file was exported, imported into the PlotJuggler software for analysis. The average value of the throttle channel output data (actuator_outputs.00/output.02) was calculated when the altitude data (distance_sensor/current_distance) was stable, and the percentage was calculated. The test situation is shown in **Figure 8**. From the test results, it can be observed that the hover throttle of the UAV on this experimental platform is 35.9%. Under the same full-charge condition, the battery life test of the UAV was carried out. The time from takeoff to the UAV automatically landing due to low power was about 15 minutes, meeting the battery life requirement.

To sum up, the UAV experimental platform uses a 400 mm wheelbase full carbon fiber frame, T-motor cine77 KV977 motor, equipped with 9.5-inch three-blade propeller, 80A four-in-one electronic speed controller, and 6000 mah 6s Grep battery. The total weight after assembly is 2.7 kg, the hover throttle is 35.9%, and the battery life is 15 minutes. The flight effect can fully meet the experimental needs.

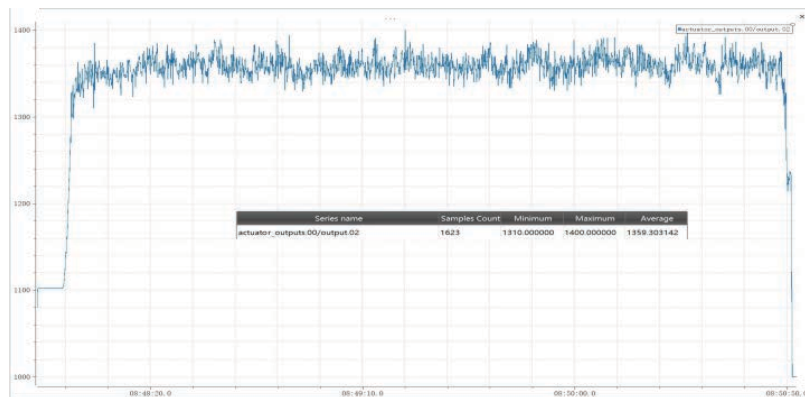


Figure 8. Throttle curve in position mode.

4. Experimental environment setup

4.1. Experimental system overall framework diagram

In this experiment, an indoor positioning system was first constructed, where an infrared high-speed camera transmits captured data to a Windows computer installed with Motive software, and the data is then sent in real time via a data link and mesh networking to the Robot Operating System platform running on the Ubuntu system. Within ROS, a trajectory planning algorithm is programmed and implemented, and the generated control commands are transmitted to the flight control system through serial communication using the MAVLink protocol, autonomous flight is thus achieved through the execution of control algorithms, while the ROS node publishes the UAV's pose and velocity commands to the corresponding pose and velocity topics, completing the flight control closed loop via inter-node communication, as illustrated in **Figure 9**. For this experiment, the environmental space was built using the experimental system framework of a quadrotor UAV. **Figure10(a)** depicts the motion capture system, within whose capture area the UAV's actual flight space is confined. Reflective marker points were affixed to the UAV, a rigid body was created in Motive software, and the UAV's position data was then streamed.

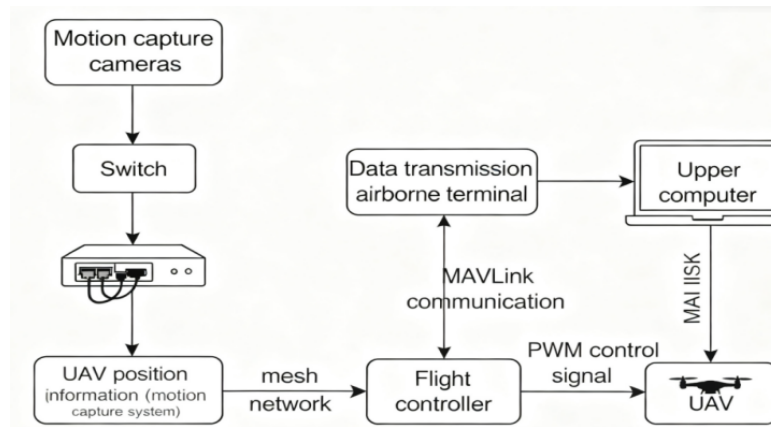


Figure 9. Experimental system overall framework diagram

Obstacles were set up to simulate an indoor environment. As shown in **Figure 10(b)**, Scene 1 replicates a conventional indoor scenario: the UAV takes off from Point A, autonomously plans a trajectory to bypass two pillar-shaped obstacles, and finally reaches the designated Point B, this setup is designed to test the algorithm's obstacle avoidance capability in typical indoor environments. As shown in **Figure 10(c)**, Scene 2 simulates a narrow passage scenario (e.g., indoor windows); the UAV is required to take off from the front of the obstacle, pass through the small gap illustrated in **Figure 10(c)**, and land at the designated point behind the obstacle, which is intended to test the algorithm's ability to navigate through narrow spaces. Power on the UAV and wait for the onboard equipment to initialize. Terminus software is used to establish an SSH connection with the onboard computer to verify the publishing frequency of the `/local_position/pose` topic. The UAV is then manually moved to observe whether the X-, Y-, and Z-axis data change correctly, thereby validating the coordinate system consistency, as shown in **Figure 10(d)**. Finally, a takeoff test is conducted in position-hold mode to evaluate whether the UAV can maintain stable fixed-point hovering.

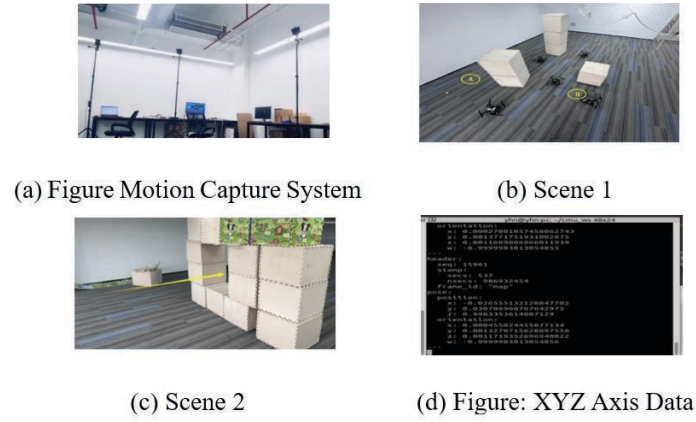


Figure 10. Test environment.

4.2. Indoor scene obstacle avoidance experiment

To verify the obstacle avoidance performance and stability of the path planning algorithm in real-world environments, the experiment conducted 10 repeated flight tests on the premise of ensuring the UAV take-off point was strictly aligned with the coordinate system. In each experiment, the UAV was required to independently traverse obstacles and finally reach the preset target point. The system dynamically updated the path and map based on real-time environmental perception data, and triggered the path replanning mechanism when necessary to adapt to sudden obstacles or environmental changes in the path. As shown in **Table 2**, the UAV successfully completed the task in all 10 experiments, accurately reaching the set destination each time, demonstrating a high task completion rate and path execution stability. The experiment duration ranged from 23.5s to 30.6s, with an average task completion time of 26.48s, indicating that the overall planning efficiency of the system was acceptable, though there were still cases of prolonged task execution time under specific conditions.

In all experiments, path replanning was triggered only once during the 6th flight: the system successfully detected obstacle risks in the path, constructed a new feasible trajectory based on the current environment, and the UAV completed the remaining flight tasks according to the new path without task failure or flight interruption. This indicates that the algorithm has certain robustness in coping with local obstacles or dynamic path changes. No path replanning occurred in the remaining 9 experiments, proving the high rationality of the initial planned path.

Table 2. Experimental results

Number of experiments	Experimental duration	Number of re-plannings during obstacle traversal	Experimental results
1	27.8s	0	Reach the preset point
2	25.3s	0	Reach the preset point
3	27.5s	0	Reach the preset point
4	26.3s	0	Reach the preset point
5	26.8s	0	Reach the preset point
6	30.6s	1	Reach the preset point
7	25.4s	0	Reach the preset point
8	26.3s	0	Reach the preset point
9	23.5s	0	Reach the preset point
10	25.3s	0	Reach the preset point

5. Conclusion

For real-world experimental validation and performance evaluation, a complete UAV experimental platform was built. Leveraging an upper computer based on RK3588, a depth camera, and a high-precision motion capture system, the algorithm was deployed on the onboard computer to realize autonomous flight missions. Two scenario tests, indoor planning simulation and window-traversal planning, were conducted, and the experimental results demonstrated that the algorithm can effectively handle obstacles and narrow passage environments, meeting the practical application requirements in terms of real-time performance and robustness. The experimental data further confirmed the effectiveness of the integrated scheme (combining vision-aided dynamic mapping and receding horizon prediction) in enhancing the safety of UAV autonomous navigation and improving planning efficiency. In the broader context of UAV indoor intelligent navigation research, this study's exploration of lightweight and real-time algorithms provides a valuable summary for multi-rotor UAV applications, offering guidance for algorithm selection in scenarios such as warehouse logistics, underground pipeline inspection, and disaster rescue in complex buildings. Despite the algorithm's proven advantages in indoor path planning, limitations remain, particularly in the in-depth understanding of more advanced path planning methodologies. Future research will focus on three directions as follows:

- (1) The organic integration of data-driven approaches (e.g., deep learning and reinforcement learning) with the current gradient-based optimization method;
- (2) In-depth investigation of multi-UAV collaboration and dynamic obstacle prediction;
- (3) Research on real-time scheduling mechanisms in edge computing environments.

These efforts aim to enable UAVs to achieve autonomous and stable flight in even more complex indoor environments.

Funding

General Research Project of Education Department of Zhejiang Province (Project No.: Y202558184, Y202558181); Scientific Research Fund of Hangzhou Dianzi University Information Engineering College (Project No.: KYP0324006); National Training Program of Innovation and Entrepreneurship for Undergraduates (Project No.: 202513279019)

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Wu H, 2021, Research on Trajectory Planning Algorithm of Quadrotor UAV in Complex Environment, thesis, Northeastern University.
- [2] Liao W, 2021, Research on UAV Path Planning Algorithm in Complex Environment, thesis, University of Electronic Science and Technology of China.
- [3] Liu H, 2023, Research on Path Planning Algorithm of Quadrotor UAV in Complex Environment, thesis, Central

South University.

- [4] Wang Y, 2023, Research on UAV Path Planning Algorithm in Complex Environment, thesis, Xidian University.
- [5] Li J, Xiong X, Yang K, et al., 2024, Research on Indoor UAV Path Planning Based on ROS and Fusion Algorithm. *Foreign Electron. Meas. Technol.* 2024(43): 173–181.
- [6] Lü Z, 2017, Research on Path Planning Method of Multi-Rotor UAV in Indoor Environment, thesis, Shenyang Jianzhu University.
- [7] Xu Z, Zhan X, Chen B, et al., 2023, A Real-Time Dynamic Obstacle Tracking and Mapping System for UAV Navigation and Collision Avoidance with an RGB-D Camera. In *Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA)*, 10645–10651.
- [8] Xu Z, Zhan X, Xiu Y, et al., 2024, Onboard Dynamic-Object Detection and Tracking for Autonomous Robot Navigation With RGB-D Camera. *IEEE Robot. Autom. Lett.* 2024(9): 651–658.
- [9] Zhou B, Gao F, Wang L, et al., 2019, Robust and Efficient Quadrotor Trajectory Generation for Fast Autonomous Flight. *IEEE Robot. Autom. Lett.* 2019(4): 3529–3536.
- [10] Zhou X, Wang Z, Ye H, et al., 2021, EGO-Planner: An ESDF-Free Gradient-Based Local Planner for Quadrotors. *IEEE Robot. Autom. Lett.* 2021(6): 478–485.
- [11] Luo J, Tian Y, Wang Z, 2024, Research on Unmanned Aerial Vehicle Path Planning. *Drones*, 2024(8): 51.
- [12] Xu Z, Xiu Y, Zhan X, et al., 2023, Vision-Aided UAV Navigation and Dynamic Obstacle Avoidance Using Gradient-Based B-Spline Trajectory Optimization. In *Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA)*, 1214–1220.
- [13] Liu X, Long Y, 2024, Research on UAV Path Planning in Complex Environment, thesis, University of Electronic Science and Technology of China.
- [14] Lin S, 2023, Research on UAV Path Planning and Control in Complex Environment, thesis, Nanjing University of Information Science and Technology.
- [15] Zhang L, 2018, Design and Implementation of Quadrotor UAV Control System Based on Intelligent Optimization Algorithm, thesis, North China University of Water Resources and Electric Power.
- [16] Zhang F, Bai W, Qiao Y, et al., 2019, UAV Indoor Path Planning Based on Improved D* Algorithm. *J. Intell. Syst.* 2019(14): 662–669.
- [17] Feng X, 2019, Research on Dynamic Path Planning of Multi-Rotor UAV in Indoor Environment, thesis, Shenyang Jianzhu University.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

NI-HotStuff: A Reputation-Driven Committee Framework for Efficient and Robust BFT Consensus

Long Zheng*

School of Computer Science and Technology, Taiyuan Normal University, No.319 Daxue Street, Yuci, 030619, China

*Corresponding author: Long Zheng, 1649635464@qq.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Consensus mechanisms are fundamental to blockchain systems, ensuring that distributed nodes agree on the validity of transactions and data. However, performance bottlenecks, particularly those related to throughput, latency, and node selection, have increasingly constrained the scalability of modern blockchain deployments. To address these issues, this paper proposes NI-HotStuff, a reputation-driven committee-based BFT consensus framework built upon the HotStuff protocol. A CatBoost-based reputation model is introduced to learn and evaluate historical behavioral features of nodes, enabling quantitative reputation scoring. A hardware-aware bidding mechanism is further incorporated to dynamically compute each node's bid value and integrate it with its reputation score, thereby prioritizing stable and high-performance nodes for consensus participation. Moreover, a committee mechanism is established in which a set of committee nodes were selected from the candidate pool, and only committee members participate in the consensus process, reducing redundant communication and mitigating the performance drag caused by weak nodes. On top of that, a leader-selection strategy based on reputation values and inter-view time intervals is designed to prevent low-reputation or potentially malicious nodes from frequently becoming leaders. Experimental results demonstrate that NI-HotStuff significantly outperforms traditional PBFT and HotStuff in terms of communication overhead, consensus latency, and system throughput, with particularly notable improvements in small- and medium-scale node environments.

Keywords: Blockchain; Consensus algorithm; HotStuff; Reputation model; Bidding mechanism

Online publication: February 12, 2026

1. Introduction

In 2008, Satoshi Nakamoto launched the era of digital currencies driven by blockchain technology through the seminal work “Bitcoin: A Peer-to-Peer Electronic Cash System”^[1]. As a decentralized and tamper-resistant distributed ledger based on chained data storage, blockchain integrates multiple fundamental technologies, including distributed storage, peer-to-peer communication, consensus mechanisms, and cryptography^[2]. By continuously appending blocks that record transactions and information, blockchain ensures data security, transparency, and trustworthiness. Today, it has become an essential component of the modern digital economy,

with applications spanning finance, healthcare, logistics, supply chain management, and many other domains. A schematic diagram of blockchain-based traceability is shown in **Figure 1**. However, as blockchain is increasingly deployed in practical scenarios, improving system performance, particularly the efficiency of consensus, has emerged as one of the core challenges in its technological development ^[3].

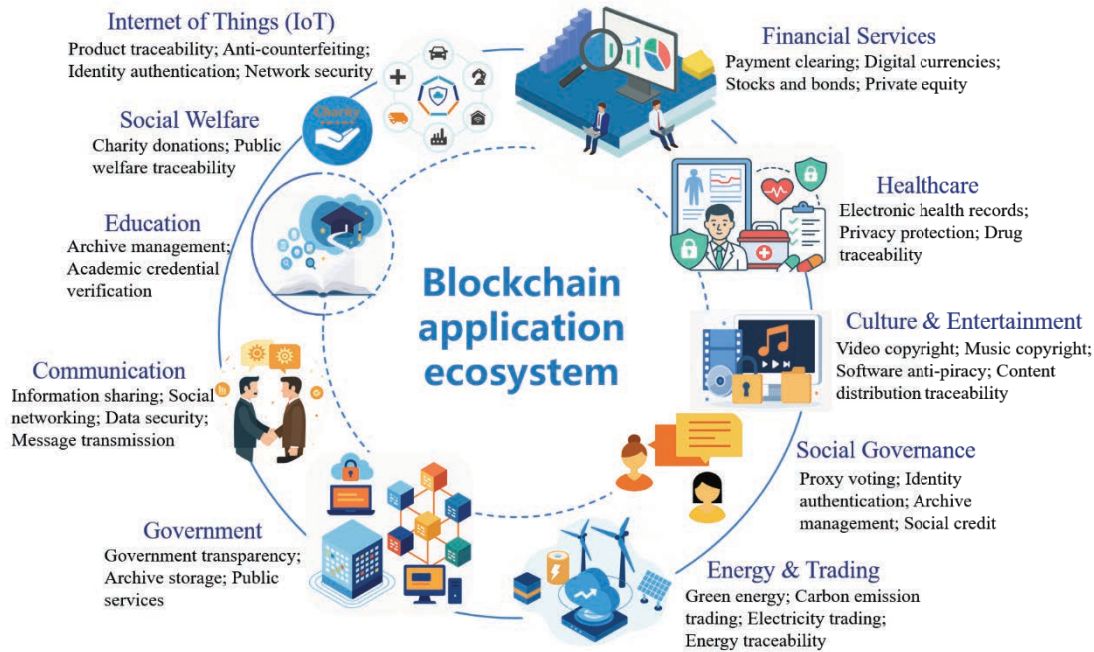


Figure 1. Diagram of blockchain-based traceability scenarios.

Consensus mechanisms constitute the core of blockchain systems, ensuring that all distributed nodes reach agreement on the validity of transactions and data. Commonly used consensus algorithms include Proof of Work (PoW), Proof of Stake (PoS), and Practical Byzantine Fault Tolerance (PBFT) ^[4]. PBFT has been widely adopted in private and consortium blockchains due to its strong fault tolerance and high security ^[5]. However, as the number of participating nodes increases, the communication overhead and latency of PBFT grow rapidly, limiting its scalability and applicability in large blockchain networks. To address these challenges, the HotStuff protocol introduces an optimized design. HotStuff employs pipelined execution and simplified signature aggregation to reduce communication rounds among nodes, thereby significantly improving consensus efficiency ^[6].

Nevertheless, HotStuff still faces bottlenecks in node selection, leader election, and its overall consistency protocol. To mitigate these issues, a variety of enhancements have been proposed in the literature ^[7]. Some work introduced a novel view-change mechanism that optimizes leader rotation, reducing both communication overhead and delay during leader transitions, and enabling rapid and effective leader replacement when failures occur ^[8]. This design improves consensus efficiency and enhances overall system performance. Another study proposed the Sync HotStuff protocol, which removes the traditional lock-step execution constraints in synchronous BFT systems and adopts a two-phase leader-driven structure that enables replicas to advance the next round of consensus without waiting for the previous commit ^[9]. This significantly reduces latency bottlenecks and improves throughput in synchronous BFT settings. Furthermore, it was introduced that weighted voting and an enhanced leader-rotation strategy, yielding substantial improvements from vote distribution to leader replacement within the HotStuff framework ^[10].

To address the limitations of HotStuff in node selection and leader election, this paper proposes NI-HotStuff, a reputation-driven committee-based BFT consensus framework. The main contributions are as follows:

- (1) Construct a reputation evaluation and node classification module that generates dynamic reputation scores for all nodes based on their historical behavior features using the CatBoost model;
- (2) A node bidding mechanism is introduced to dynamically compute each node's bid value by combining its hardware resources with its reputation score. This mechanism enables the system to select well-provisioned nodes for consensus participation, thereby reducing the involvement of low-performance nodes and improving both fault tolerance and overall consensus efficiency;
- (3) A committee formation and leader-selection strategy is designed. From the candidate set with reputation scores above a minimum threshold, committee members are selected through weighted random sampling, and only committee nodes participate in the consensus process. Within the committee, the leader is dynamically chosen based on both node reputation and inter-view time intervals, preventing malicious or low-reputation nodes from frequently serving as leaders and enhancing the fairness, stability, and overall efficiency of the consensus protocol.

2. Related work

2.1. PBFT algorithm

The Byzantine Generals Problem was introduced by Lamport *et al.* to describe the consensus challenge faced by distributed systems in the presence of malicious nodes^[11]. To address this issue, Castro *et al.* proposed the PBFT algorithm, which aims to ensure that all nodes can reach agreement even in Byzantine environments. The core objective of PBFT is to guarantee consistent consensus despite the presence of faulty or adversarial nodes. Specifically, in a blockchain network with n nodes, the system can tolerate up to f Byzantine nodes, where $f = (n-1)/3$. The PBFT algorithm proceeds through several phases, and the overall consensus workflow is illustrated in

Figure 2 and as follows:

- (1) Request phase: The client sends a request to the primary node and waits for a response;
- (2) Pre-prepare phase: Once the primary node receives the client request, it generates a unique identifier and creates a proposal for the request. The proposal is then broadcast to all replica nodes;
- (3) Prepare phase: Upon receiving the pre-prepare message, each replica generates a prepare message and broadcasts it to all nodes. When a node receives more than $2f$ matching prepare messages that are consistent with the pre-prepare message, it proceeds to the commit phase;
- (4) Commit phase: The node generates a commit message and broadcasts it to all other nodes. When it receives more than $2f$ matching commit messages that correspond to the same pre-prepare message, it sends a response to the client;
- (5) Reply phase: When the client receives more than f matching responses, it concludes that the transaction has been successfully completed.

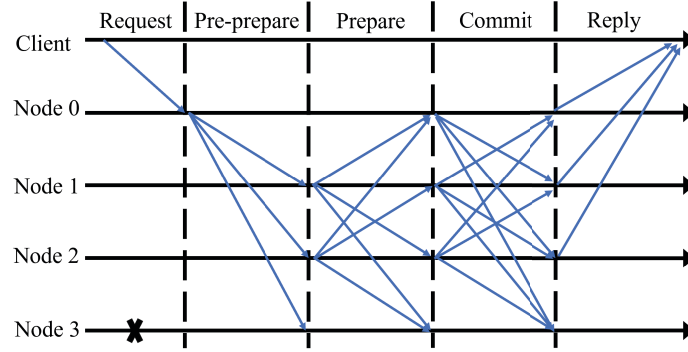


Figure 2. PBFT consensus process.

2.2. HotStuff algorithm

HotStuff is an improved Byzantine Fault Tolerant consensus protocol that operates in an environment similar to PBFT but maintains system state using a tree-structured architecture rather than a traditional log-based structure. Meanwhile, HotStuff extends the traditional two-phase commit procedure into a three-phase commit pipeline, reducing the communication complexity of view change from $O(n^2)$ to $O(n)$ and thereby significantly lowering the communication burden as the system scales. By adopting a one-to-many message dissemination pattern, HotStuff effectively compresses communication overhead and completes consensus within a single view^[12]. Each view is associated with a unique, monotonically increasing view number and is coordinated by a designated leader. Upon completing a view, the system automatically progresses to the next one. If the current leader behaves maliciously or becomes faulty, a new leader takes over after a view timeout and continues the consensus process. Unlike PBFT, HotStuff embeds the view-change mechanism within every round of consensus, rotating the leader after each committed block. The overall workflow proceeds through three phases, prepare, pre-commit, and commit. In each phase, the leader must collect at least $n-f$ replica votes to form a quorum certificate (QC), which serves as proof of intermediate agreement. The execution of the algorithm consists of the following stages, and the complete consensus process is illustrated in **Figure 3** and as follows:

- (1) **Prepare phase:** The new leader first collects $n-f$ NEW-VIEW messages from different replicas and extracts the prepareQC with the highest view number as the base branch (highQC). The leader then constructs a leaf node containing the client command at the end of the current branch and broadcasts a prepare message containing the proposal details to all replicas. Each replica verifies the validity of the proposal, and if it satisfies the branch extension rules, it sends a prepareVote to the leader;
- (2) **Pre-commit phase:** When the leader successfully aggregates $n-f$ prepareVote messages, it forms the prepareQC and broadcasts the pre-commit message. After receiving this message, each replica confirms that more than $2f$ replicas have approved the proposal, updates its local prepareQC bstatus, and sends back a commitVote to advance the protocol;
- (3) **Commit phase:** The leader must collect $n-f$ preCommitVote messages to generate the preCommitQC. After verifying this certificate, each replica marks the branch of the current proposal as a stable state (updating its lockedQC) and sends a commitVote to the leader, indicating its agreement to commit the proposal;
- (4) **Decide phase:** After obtaining $n-f$ commitVote messages, the leader generates the final commitQC and broadcasts the decide message. Replicas confirm that the proposal branch now satisfies commit rules and execute the client command stored in the leaf node. Once the command is executed, all replicas automatically trigger the view transition and proceed to the next consensus round.

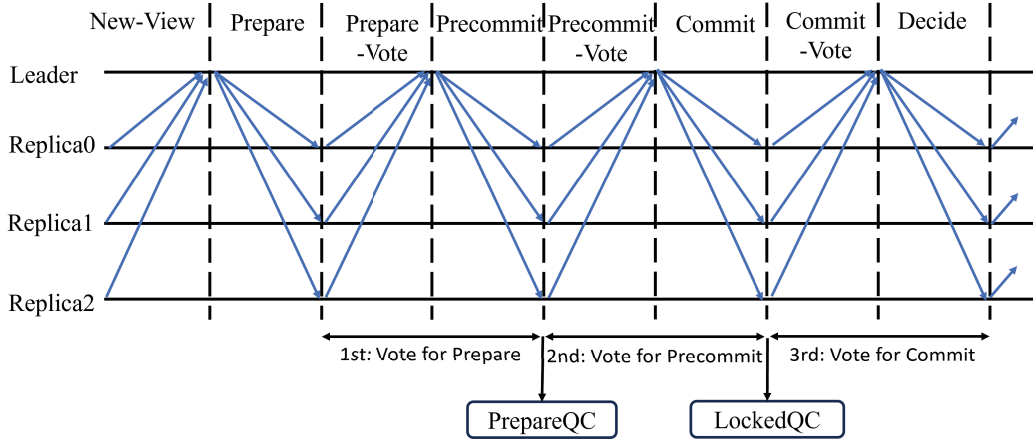


Figure 3. HotStuff consensus process.

2.3. CatBoost model

CatBoost is an ensemble model based on Gradient Boosting Decision Trees, with its most distinctive feature being its method of handling categorical features^[13]. Traditional gradient boosting decision trees typically require categorical features to be converted into numerical features, such as one-hot encoding or label encoding^[14,15]. In contrast, CatBoost adopts a specialized ordered-boosting algorithm that enables direct processing of categorical attributes without the need for explicit transformation, making it particularly suitable for solving classification problems^[16]. By applying the CatBoost algorithm, a node classification model can be constructed to categorize nodes on the blockchain into different classes. The details are as follows:

Given a dataset with m features $\varphi = \langle x_1, x_2, \dots, x_m \rangle$ and label δ , a training set T with n samples can be formed as $T = \{[\varphi_1, \delta_1], [\varphi_2, \delta_2], \dots, [\varphi_n, \delta_n]\}$. The CatBoost algorithm performs iterative optimization, where each iteration constructs a model that minimizes the loss function using gradient boosting decision trees. After k rounds of boosting, the optimal weak learner h_k for the final iteration is obtained, and all weak learners from each iteration are combined to form the final predictor F_k . As shown in **Equation (1)**:

$$\begin{cases} F_k = F_{k-1} + \alpha h_k \\ h_k = \arg \min \sum_{i=1}^m \beta_i L(y_i, F_{k-1}(x_i) + h_k(x_i)) \end{cases} \quad (1)$$

Here, α is the learning rate, which controls the update step size of the model parameters in each boosting iteration. The operator \min denotes the minimization of the objective function, β_i represents the sample weight, and $L()$ is the loss function used to measure the deviation between the true value and the predicted value of a sample. The term $F_{k-1}(x_i) + h_k(x_i)$ denotes the output of the k -th learner for the input x_i .

For the multi-class classification task in this study, the multinomial cross-entropy loss function is adopted, which is derived from the concept of categorical cross-entropy, defined as

$$L(y_i, p_i) = - \sum_{i=1}^N y_i \log(p_i) \quad (2)$$

Here, N denotes the number of classes, y_i represents the probability distribution of the true label, and p_i

represents the probability distribution output by the model.

2.4. Bidding mechanism

The bidding strategy refers to a class of methods in which participants acquire specific resource access rights or benefits by submitting competitive bids based on their own resource valuations and actual demands during the resource allocation and decision-making process^[17]. In traditional bidding mechanisms, participants typically consider factors such as initial investment cost, risk preference, and the expected value fluctuations of target resources to determine a reasonable bid. At the same time, participants may engage in strategic interactions and competition, requiring them to assess and anticipate possible bidding behaviors of others and adjust their own strategies accordingly to improve their chances of obtaining the desired resources or services.

Bidding strategies have been widely applied across various domains^[18–20]. In the energy management sector, power companies and users utilize bidding schemes to complete electricity trading and scheduling, thereby increasing energy utilization efficiency while reducing operational costs. In the transportation domain, vehicles can obtain preferential passage rights or facility usage rights through reliable bidding to optimize traffic flow and improve the allocation efficiency of transportation resources, easing congestion to some extent. In the field of data sharing and privacy protection, bidding allows participants to negotiate data access rights and usage permissions, enabling controlled and secure data exchange. Furthermore, in Internet-of-Things environments, intelligent devices can dynamically allocate bandwidth and storage resources through bidding, thereby enhancing system responsiveness and improving overall performance. Therefore, introducing bidding mechanisms into blockchain consensus provides new perspectives and support for node selection and resource allocation.

3. NI-HotStuff algorithm

To address the issues in the traditional HotStuff protocol, such as the randomness of node selection, the delayed handling of primary node failures, and the reduced consensus efficiency in the presence of malicious nodes, this paper proposes a reputation-driven committee-based BFT consensus framework, named NI-HotStuff. The proposed framework establishes a reputation evaluation and node classification module, which models historical node behaviors to generate dynamic reputation scores, forming a basis for subsequent committee selection and primary node determination. Building on this foundation, the algorithm incorporates hardware resource indicators and reputation scores to design a dynamic bidding mechanism that comprehensively assesses each node's bidding value. This mechanism ensures that only high-performance nodes participate in the consensus process, thereby reducing interference from low-capacity nodes and improving overall system efficiency. Furthermore, NI-HotStuff introduces a primary-node selection strategy based on reputation scores and time intervals. When the primary node fails or exhibits Byzantine behavior, the protocol can quickly determine a new primary node through reputation-based ranking and priority adjustment, ensuring the continuity and stability of consensus. Before consensus begins, NI-HotStuff performs preprocessing on all participating nodes, collecting and organizing behavioral data to train the reputation model. After consensus starts, the protocol follows the consistency procedure described in Section 3.1 to ensure the integrity and correctness of blockchain data.

3.1. Node identification mechanism

In the HotStuff protocol, the primary node is typically selected through a fixed round-robin mechanism based on the view number. This rotation strategy helps mitigate the risk of single-point failure and improves fairness

and robustness to some extent ^[21]. However, as the network scales, the communication overhead among nodes increases significantly, and fundamental factors such as hardware capability and network conditions can further influence the overall performance of HotStuff. Since the primary node is responsible not only for proposing and broadcasting blocks but also for performing additional bookkeeping operations, selecting an unreliable or Byzantine node as the primary may compromise the security of the consensus process. To address these challenges, NI-HotStuff integrates a CatBoost-based model along with a bidding mechanism to optimize primary node selection. By ranking candidate nodes through a comprehensive scoring scheme, the algorithm prioritizes stable and high-performance nodes for participation in consensus, thereby reducing the likelihood of Byzantine nodes being elected and effectively alleviating communication overhead within the network.

3.1.1. Node competitive value assessment

Under similar conditions, nodes equipped with superior computing and storage resources can process transactions, message broadcasting, and other consensus operations more efficiently, exhibiting faster execution speed and higher processing throughput. Meanwhile, their lower probability of hardware failure provides better stability and resilience against risks. Based on this observation, the NI-HotStuff algorithm introduces a node bidding mechanism into the reputation evaluation process, treating basic resource allocation as an important weighting factor. Before entering each consensus round, nodes submit a bidding value according to their computational and storage capabilities. A higher bidding value not only reflects a node's greater willingness to contribute resources to the consensus process but also increases the likelihood of obtaining a higher reputation score, thereby improving the potential of being selected as the primary node.

In this paper, considering node attributes and interaction characteristics, three key indicators, CPU core count, available storage capacity, and network bandwidth, are selected to measure a node's basic resource level. When a node first joins the network, its maximum available resources are recorded as the basis for subsequent reputation evaluation and the bidding process. Since these indicators differ in magnitude, we adopt normalization to ensure comparability. The normalized value is obtained using **Equation (3)**, where x denotes the original metric value and x' represents the normalized result:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (3)$$

After normalization, the bidding value of each node is computed using **Equation (4)**. Specifically, B_i denotes the bidding value of node i ; c_i , s_i , and b_i represent the number of CPU cores, available storage capacity, and network bandwidth of node i in the current consensus round, respectively. The parameters α , β , and γ are the weight coefficients corresponding to CPU cores, available storage, and bandwidth, and they satisfy $\alpha + \beta + \gamma = 1$. These weights can be dynamically adjusted according to practical system requirements.

$$B_i = \alpha \cdot c_i + \beta \cdot s_i + \gamma \cdot b_i \quad (4)$$

3.1.2. Node reputation score assessment

Before constructing the node identification model, a large number of predefined node samples are first introduced into the blockchain system in a balanced manner, during which the system extracts and transforms node features to obtain the corresponding characteristic values. To ensure the completeness of feature extraction, the system assigns an initial reputation value to each node and categorizes them into high-reputation, medium-reputation,

and low-reputation nodes according to their reputation levels. The correspondence between node categories and reputation values is shown in **Table 1**. During the identification phase, high-reputation nodes are generally capable of efficiently completing tasks such as block exchange and broadcasting, thus improving system verification and consensus performance. Medium-reputation nodes behave similarly to high-reputation nodes but may occasionally exhibit slight deviations that remain within an acceptable range. In contrast, low-reputation nodes often exhibit unfavorable behaviors, such as delayed responses or frequent timeout events, which may adversely affect system consensus efficiency and security.

Table 1. Reputation classification criteria for nodes

Node category	Reputation value
High-reputation node	$r_i > 0.7$
Medium-reputation node	$-0.5 < r_i \leq 0.7$
Low-reputation node	$r_i \leq -0.5$

Note: denotes the reputation value of node i .

The definitions of the four feature attributes are given as follows:

(1) Definition 1: Heartbeat detection, which measures whether a node is online and its responsiveness. It is defined as:

$$H(i) = \frac{r_i}{t_i} \quad (5)$$

where r_i denotes the number of successful heartbeat responses received from node i during the observation period, and t_i represents the total number of heartbeat requests sent to node i .

(2) Definition 2: Node contribution, which reflects the node's activity level and assigns higher weights to recent participation through time decay. It is defined as:

$$C(i) = \frac{s_i}{t_i + \alpha} \quad (6)$$

where s_i refers to the number of successful consensus completions by node i during the observation period, t_i is the total number of consensus rounds that node i participated in, and α is a smoothing factor.

(3) Definition 3: Delay index, which reflects the node's delay performance during communication. It is defined as:

$$D(i) = \frac{1}{1 + \frac{\bar{l}_i}{L_{max}}} \quad (7)$$

where \bar{l}_i is the average round-trip latency of node i , and L_{max} denotes the maximum allowable latency in the system.

(4) Definition 4: Network utilization, which reflects the usage condition of the node's network resources. It is defined as:

$$U(i) = \frac{\bar{b}_i}{B_i} \cdot (1 - p_i) \quad (8)$$

where \bar{b}_i is the average effective bandwidth usage of node i , B_i is its available bandwidth capacity, and p_i denotes the average packet loss rate of node i .

After training, the CatBoost model is deployed into the operational environment. At the end of each consensus round, behavioral and performance data are collected for each node, and corresponding feature values are computed and broadcast to all nodes. Based on the accumulated feature entries for each node, the CatBoost algorithm performs classification and identification. Furthermore, these features are not only used as classification inputs but also form the feature set for each consensus round. The updated reputation values from the previous round are incorporated as inputs, enabling correction of historical reputation scores. In this manner, the model reflects both long-term node stability and short-term dynamic changes, thereby providing more accurate references for the overall reputation scoring mechanism.

3.1.3. Node comprehensive scoring mechanism

During node selection, relying solely on performance-based metrics to compute the bidding value can reflect a node's computing power, bandwidth, and storage resources, but it fails to capture the node's overall reliability. Similarly, although reputation values derived from behavioral features can effectively identify anomalous behavior or malicious nodes, they overlook differences in hardware resources. Therefore, bidding values emphasize static computational capability, whereas reputation emphasizes dynamic operational behavior, and neither alone offers a complete evaluation. Thus, based on the two categories of metrics introduced earlier, this paper proposes a hybrid evaluation mechanism that integrates the bidding value with the reputation score, thereby preserving both security and performance to obtain a more representative node evaluation outcome.

To achieve this, we introduce a cross-round dynamic reputation update mechanism. The reputation value of node i in round t is defined as:

$$R_i^t = \delta \cdot R_i^{t-1} + (1 - \delta) \cdot f(X_i^t) \quad (9)$$

where R_i^t is the reputation value from the previous round, and $f(X_i^t)$ is the reputation score computed using this round's behavioral feature set (including heartbeat detection, node contribution, delay index, and network utilization). The parameter $\delta \in (0,1)$ is a smoothing factor that balances long-term and short-term behavior, preventing extreme oscillations caused by temporary performance fluctuations.

Based on this, the comprehensive evaluation score of node i is computed as:

$$S_i^t = \lambda \cdot R_i^t + (1 - \lambda) \cdot B_i \quad (10)$$

where B_i denotes the bidding value derived from the node's CPU cores, storage capacity, and available bandwidth, and λ is the weight factor.

Through this evaluation method, the final score not only accounts for the node's fundamental hardware resources but also incorporates its accumulated reputation across rounds, enabling a more holistic node assessment.

3.2. Consistency protocol optimization

Although the HotStuff protocol significantly reduces communication complexity compared with PBFT, in practical deployments it still requires all nodes to participate in consensus, which leads to additional overhead and increases the risk of communication failures. Therefore, NI-HotStuff introduces a committee-based mechanism at the consistency layer. By computing the comprehensive score of each node using the reputation evaluation model and setting a minimum threshold S_{min} , only nodes whose scores exceed this threshold are included in the candidate set C . To ensure the quality and controllable size of the candidate set, this paper adopts a Top-M selection rule: in the t -th round, nodes are sorted in descending order according to their comprehensive score S_i^t , and the top M nodes are selected to form C_t , where $M=K+m$ ($K=3f+1$, m is the redundancy, fixed as $m=\max(2, \lceil 0.2K \rceil)$). The minimum threshold S_{min} corresponds to the M -th highest score. If system fluctuations or abnormal score distribution cause fewer than M nodes to satisfy the threshold, then all valid nodes are included in C_t . In case of ties, the ordering follows the previous-round ranking or the lexicographical order of node IDs. According to Byzantine fault tolerance theory, in a blockchain system with $n \geq 3f+1$ nodes, only when the committee size contains at least $3f+1$ nodes can it guarantee that, even in the presence of up to f malicious nodes, any two quorums (each larger than $2f+1$) still have non-empty intersection with honest nodes, thus preventing consensus divergence. Hence, the system adopts a weighted random sampling process to select a final committee of size $n'=3f'+1$, denoted as K , and only committee members participate in the consensus procedure.

The probability that node i is selected into the committee K is given by:

$$P_i(i \in K) = \frac{e^{\tau S_i^t}}{\sum_{j \in C} e^{\tau S_j^t}} \quad (11)$$

where S_i^t denotes the comprehensive score of node i in the t -th round, which considers both its reputation value and hardware resource capability. A higher score indicates that the node is more advantageous. The parameter τ is a tuning factor (temperature coefficient) used to control the preference intensity toward high-score nodes; as τ increases, the probability of selecting nodes with higher scores becomes larger. The denominator is the sum of the exponential scores of all nodes in the candidate set C , used to normalize the probability and ensure that the sum of the selection probabilities of all candidate nodes is 1.

Under this mechanism, nodes with higher comprehensive scores are more likely to be selected into the committee, while a certain degree of randomness is preserved to maintain fairness. Meanwhile, by adhering to the safety bound of BFT theory, the system guarantees security and liveness even under potential Byzantine behaviors. Therefore, this mechanism effectively reduces unnecessary communication and computation overhead, improving system reliability and consensus efficiency. The improved consistency protocol is illustrated in **Figure 4**.

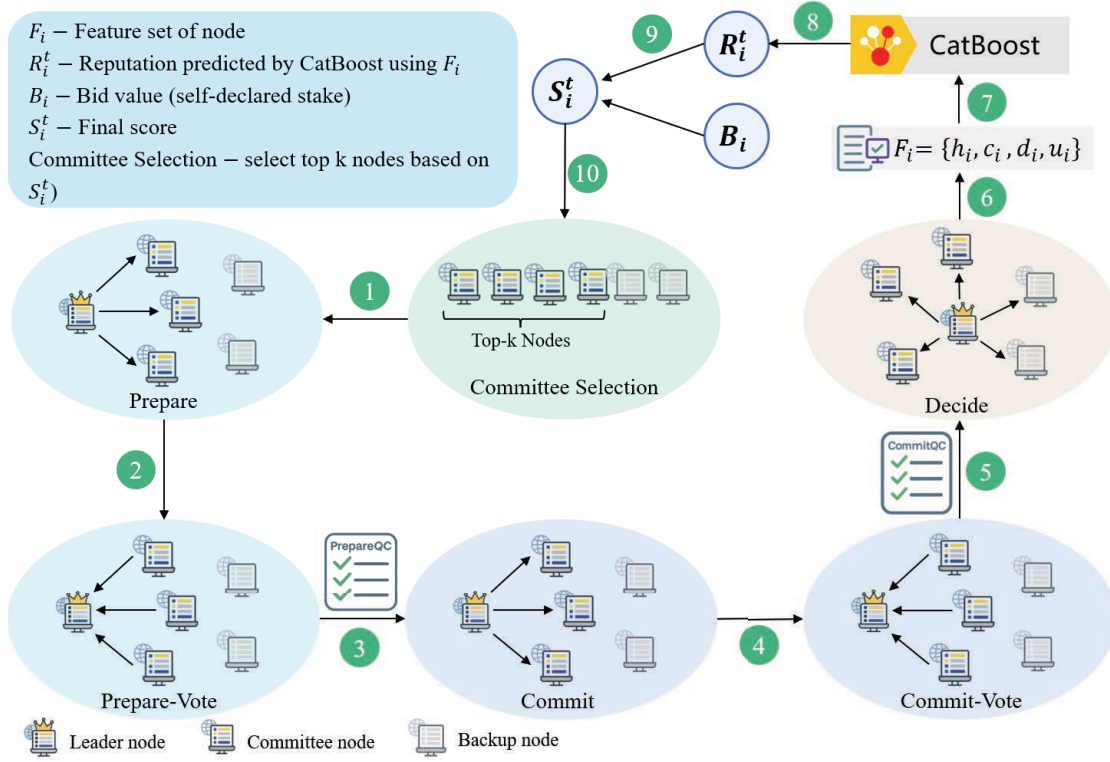


Figure 4. NI-HotStuff consensus process diagram.

3.3. Master node selection

In terms of leader selection, the algorithm recalculates and updates the comprehensive score of each node after every consensus round, thereby refreshing the candidate committee. The priority of a node being selected as the leader is determined by the committee based on its latest comprehensive score. This priority is correlated with the score defined earlier and is also adjusted according to the time interval since the node last served as leader. Nodes with higher recent scores and shorter leader intervals are more likely to obtain higher priority, which effectively reduces the risk that a long-idle node becomes the leader and enhances the fairness and stability of the system. The leader selection priority is defined as follows:

$$M_i = S_i^t \times \frac{\Delta T_i}{T_c} \quad (12)$$

where S_i^t denotes the comprehensive score of node i at round t , as defined in Equation 10. The term ΔT_i represents the time elapsed since node i last served as the leader, and T_c is a leader selection parameter set by the system to adjust the effect of time intervals on the leader selection process.

Before leader selection, all nodes in the blockchain update their priority values. That is, after the previous consensus round, each node's comprehensive score is recalculated according to the reputation model introduced earlier, and the new leader is selected based on the updated priority ranking.

4. Experiment and results analysis

This study develops a lightweight blockchain system using the Python programming language. The experimental

environment is configured as follows: the hardware platform is equipped with an Intel® Core™ i7-14700HX CPU and 16 GB of RAM; the operating system is Windows 11 (64-bit); and the software environment uses PyCharm. The PBFT, HotStuff, and NI-HotStuff algorithms are simulated and evaluated within this setup.

4.1. Communication overhead analysis

The communication overhead refers to the total number of messages generated during the consensus process. NI-HotStuff adopts a committee-based two-phase pipeline (Prepare→Commit), where only a committee of size $M=3f+1$ participates in intra-phase interactions. According to the phase design, the leader broadcasts once to the committee in each phase, and each replica broadcasts one confirmation message. Therefore, the number of messages in a single phase is $2(M-1)$, and the two phases generate a total of $4M-4$ messages. When a view change occurs, $M-1$ NEW-VIEW messages are introduced in each switch. After v view changes, an additional $v(M-1)$ messages are produced. Hence, the total communication overhead can be approximated as: $4M-4+v(M-1)$. Compared with the original HotStuff, in which all N nodes participate in the three-phase process M , NI-HotStuff reduces the number of participating nodes from N to committee size M , and reduces the number of phases from three to two, thereby maintaining linear communication complexity $O(M)$ while significantly decreasing constant factors and practical communication overhead.

4.2. Latency analysis

In blockchain systems, consensus latency refers to the time interval from when a client initiates a request to when the request is finalized. Under the same experimental environment, we conducted one-round consensus delay tests for the NI-HotStuff algorithm and several other algorithms. The number of nodes was set to 80, 100, 120, 140, 160, and 180, respectively. Multiple experiments were repeated and averaged. The comparison results of PBFT, HotStuff, and NI-HotStuff are shown in **Figure 5**. It can be observed that as the network scale increases, the consensus latency of all three algorithms shows an upward trend, but to different extents. PBFT, due to its communication complexity of $O(n^2)$, experiences rapidly increasing latency as the number of nodes grows, and at 180 nodes, the delay exceeds 2500 ms, indicating poor scalability. HotStuff reduces communication complexity to $O(n)$ through its three-phase chain-locking and threshold signature optimization, significantly decreasing latency

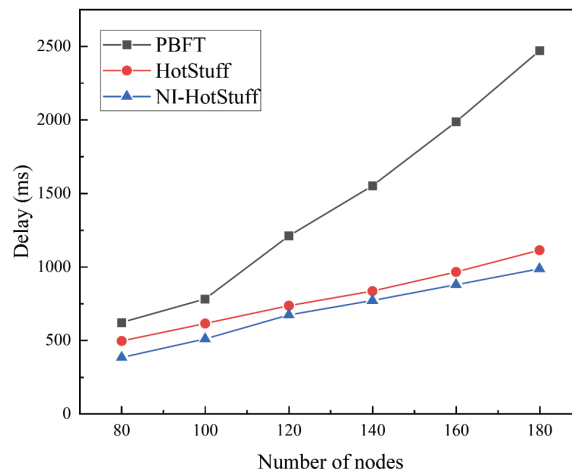


Figure 5. Comparison of consensus latency.

compared with PBFT and exhibiting a more gradual increasing trend overall. Building upon HotStuff, the NI-HotStuff algorithm proposed in this work further reduces the impact of low-performance nodes and frequent view changes through reputation evaluation and dynamic leader selection mechanisms. Therefore, NI-HotStuff maintains the lowest delay across all network sizes, achieving more than a 50% reduction compared with PBFT and a further 15–20% improvement compared with HotStuff, demonstrating superior performance and stability in complex network environments.

4.3. Throughput analysis

The throughput refers to the number of tasks or requests that the system can process per unit of time, which is generally used to measure the speed and transaction-handling capability of blockchain systems. It is usually expressed in TPS (Transactions Per Second), and its calculation formula is:

$$TPS = Transaction_{\Delta T} / \Delta T \quad (13)$$

where $Transaction_{\Delta T}$ denotes the number of transactions, and ΔT represents the time required to complete these transactions.

Under the same experimental settings with equal numbers of nodes and equal transaction loads, the throughput of the NI-HotStuff algorithm and several other algorithms is tested. The number of nodes is selected as 80, 100, 120, 140, 160, and 180, with the number of transactions set to 10. The throughput comparison results are shown in **Figure 6**. From the experimental results, it can be observed that with the increase in network size, the throughput of the three algorithms declines, but the rate of decline differs. PBFT suffers from communication complexity $O(n^2)$, causing throughput to drop sharply when the number of nodes exceeds 100 TPS, revealing a severe scalability bottleneck. HotStuff improves throughput by reducing communication complexity and outperforms PBFT across all scales. Meanwhile, the proposed NI-HotStuff algorithm achieves higher processing capability than HotStuff, with an improvement of over 25% in small-scale networks and potentially maintaining an advantage of about 10% in large-scale settings, demonstrating its efficiency and adaptability.

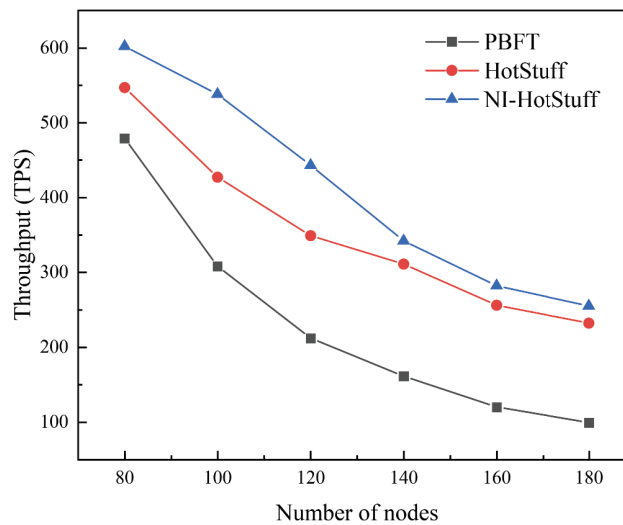


Figure 6. Throughput comparison.

5. Conclusion

To address the issues of arbitrary node selection, insufficient handling of faulty primary nodes, and performance degradation in the presence of Byzantine nodes in the traditional HotStuff consensus algorithm, this paper proposes a reputation-driven committee-based BFT consensus framework, termed NI-HotStuff. The proposed method first evaluates node reputation using a CatBoost model and incorporates competitiveness-based quantification of hardware resources to derive a weighted composite score for node selection. Meanwhile, a committee mechanism is introduced, where nodes whose reputation values exceed the minimum threshold form the candidate set, and a weighted random sampling strategy is employed to select the committee members. Only committee nodes participate in the consensus process, effectively reducing redundant communication and lowering system complexity. Furthermore, the algorithm designs a dynamic primary node selection strategy based on node reputation and time intervals, prioritizing committee members with high reputation and less frequent prior leadership, thereby significantly reducing the probability of low-performance or malicious nodes taking the primary role. Experimental results demonstrate that NI-HotStuff outperforms PBFT and HotStuff in terms of communication overhead, consensus latency, and system throughput, validating its higher efficiency and scalability while ensuring security and fairness.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Nakamoto S, Bit B, 2008, Bitcoin: A Peer-to-Peer Electronic Cash System. 2008, 2007.
- [2] Yuan F, Zuo Z, Jiang Y, et al., 2015, AI-Driven Optimization of Blockchain Scalability, Security, and Privacy Protection. *Algorithms*, 18(5): 263.
- [3] Amiri M, Wu C, Agrawal D, et al., 2024, The Bedrock of Byzantine Fault Tolerance: A Unified Platform for BFT Protocols Analysis, Implementation, and Experimentation. *21st USENIX Symposium on Networked Systems Design and Implementation (NSDI 24)*, 371–400.
- [4] Yuan F, Huang X, Zheng L, et al., 2025, The Evolution and Optimization Strategies of a PBFT Consensus Algorithm for Consortium Blockchains. *Information*, 16(4): 268.
- [5] Castro M, Liskov B, 1999, Practical Byzantine Fault Tolerance. *OsDI*, 99(1999): 173–186.
- [6] Yin M, Malkhi D, Reiter M, et al., 2019, HotStuff: BFT Consensus with Linearity and Responsiveness. *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing*, 2019: 347–356.
- [7] Malkhi D, Yin M, 2023, Lessons from HotStuff. *Proceedings of the 5th Workshop on Advanced Tools, Programming Languages, and PLatforms for Implementing and Evaluating Algorithms for Distributed Systems*, 2023: 1–8.
- [8] Jalalzai M, Niu J, Feng C, et al., 2023, Fast-Hotstuff: A Fast and Robust BFT Protocol for Blockchains. *IEEE Transactions on Dependable and Secure Computing*, 21(4): 2478–2493.
- [9] Abraham I, Malkhi D, Nayak K, et al., 2020, Sync Hotstuff: Simple and Practical Synchronous State Machine Replication. *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2020: 106–118.
- [10] Liang B, Yuan F, Deng J, et al., 2025, Cs-pbft: A Comprehensive Scoring-Based Practical Byzantine Fault Tolerance Consensus Algorithm. *The Journal of Supercomputing*, 81(7): 859.
- [11] Lamport L, Shostak R, Pease M, 1919, The Byzantine Generals Problem, *Concurrency: The Works of Leslie*

Lamport, 203–226.

- [12] Zhang Z, Hu B, Tian L, et al., 2025, Efficient Dynamic-Committee BFT Consensus Based on HotStuff. *Peer-to-Peer Networking and Applications*, 18(3): 111.
- [13] Prokhorenkova L, Gusev G, Vorobev A, et al., 2018, CatBoost: Unbiased Boosting with Categorical Features. *Advances in Neural Information Processing Systems*, 2018: 31.
- [14] Breskuvienė D, Dzemyda G, 2023, Categorical Feature Encoding Techniques for Improved Classifier Performance When Dealing with Imbalanced Data of Fraudulent Transactions. *International Journal of Computers Communications & Control*, 18(3).
- [15] Wang Z, Chen L, Wang F, 2023, Fuzzy Inference Attention Module for Unsupervised Domain Adaptation. *IEEE Transactions on Fuzzy Systems*, 32(4): 1706–1718.
- [16] Wang Z, Wang X, Liu F, et al., 2021, Adaptative Balanced Distribution for Domain Adaptation with Strong Alignment. *IEEE Access*, 2021(9): 100665–100676.
- [17] Ou W, Chen B, Dai X, et al., 2023, A Survey on Bid Optimization in Real-Time Bidding Display Advertising. *ACM Transactions on Knowledge Discovery from Data*, 18(3): 1–31.
- [18] PankiRaj J, Yassine A, Choudhury S, 2019, An Auction Mechanism for Profit Maximization of Peer-to-Peer Energy Trading in Smart Grids. *Procedia Computer Science*, 2019(151): 361–368.
- [19] Malik S, Thakur S, Duffy M, et al., 2023, Comparative Double Auction Approach for Peer-to-Peer Energy Trading on Multiple Microgrids. *Smart Grids and Sustainable Energy*, 8(4): 21.
- [20] Dramitinos M, Stamoulis G, Courcoubetis C, 2007, An Auction Mechanism for Allocating the Bandwidth of Networks to their Users. *Computer Networks*, 51(18): 4979–4996.
- [21] Yin M, Malkhi D, Reiter M, et al., 2018, HotStuff: BFT Consensus in the Lens of Blockchain, arXiv, <https://doi.org/10.48550/arXiv.1803.05069>

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A Small-Sample Bearing Fault Diagnosis Method Based on Multi-Image Fusion and Multi-Scale Dynamic Residual Dual Attention Mechanism

Ao Xu, Mu Li*

School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411100, China

*Corresponding author: Mu Li, mli@hnust.edu.cn

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In recent years, fault diagnosis methods based on convolutional neural networks (CNNs) have garnered significant attention in the field of rotating bearing fault diagnosis. Addressing the challenge of extremely limited fault signal samples, this paper proposes a small-sample bearing fault diagnosis method based on multi-image fusion and a dual-attention mechanism incorporating multi-scale dynamic residuals. This method first converts the fault signal into a two-dimensional image through continuous wavelet transform and Gram angle field (GASF/GADF), thereby transforming the fault diagnosis problem into an image feature learning problem. The model extracts basic features through the initial convolutional layer and sequentially learns deep features via multi-scale dynamic residual blocks and dual attention mechanisms. Among these, the multi-scale architecture captures features across different receptive fields through parallel convolutional branches, while the dual attention mechanism performs feature recalibration in both the channel and spatial dimensions. Experimental results demonstrate that the proposed method achieves an accuracy rate of 97.47% in bearing fault diagnosis tasks, representing a significant improvement over traditional CNN models. This validates the model's effectiveness and superiority in complex fault diagnosis scenarios.

Keywords: Fault diagnosis; Convolutional neural networks; Continuous wavelet transform; Gram angle field; Channel attention mechanism; Spatial attention mechanism; Rolling bearing fault; Multiscale residual blocks

Online publication: February 12, 2026

1. Introduction

As the core carrier of modern industrial production, mechanical equipment directly impacts a company's production efficiency and operational costs. Its health status has become crucial for ensuring the efficient operation of industrial systems. Rolling bearings, indispensable mechanical components within such equipment, support rotating shafts and their attached parts. Serving as the core elements of precision mechanical transmission systems, rolling bearings comprise a coordinated working system consisting of four major components: the inner ring, outer

ring, rolling elements, and cage ^[1,2]. The inner ring is rigidly connected to the rotating shaft via an interference fit. Its precision-machined outer raceway forms a dynamic contact surface with the rolling elements, transmitting shaft system loads to the rolling elements. The outer ring provides support and constraint through the bearing housing. Its inner raceway forms a relative motion path with the inner ring, with load distribution optimized through geometric parameter adjustments. However, rolling bearings also face risks of various potential defects such as wear and fracture. Once these issues manifest, they severely impair normal equipment operation, not only causing significant economic losses but also posing serious threats to personnel safety ^[3,4].

Signals used for rolling bearing fault diagnosis include acoustic emission signals, vibration signals, temperature signals, and current signals, among others ^[5,6]. Tao *et al.* converted one-dimensional raw vibration signals into two-dimensional time-frequency maps via short-time Fourier transform as input for a classification generative adversarial network (GAN). Leveraging the unsupervised clustering capabilities of this GAN, they achieved rolling bearing fault diagnosis ^[7]. Han *et al.* proposed a fault diagnosis model based on the fusion of multi-level wavelet packets and a dynamic ensemble convolutional neural network (DECNN). A multi-level wavelet coefficient matrix was constructed via wavelet packet transform to comprehensively represent non-stationary vibration signals ^[8]. Xia *et al.* combined vibration signals collected from multiple sensors of the same type at different locations into a two-dimensional matrix, which served as input for a 2DCNN model to identify faults in rolling bearings and gearboxes ^[9]. Jiao *et al.* proposed a complementary data-driven deep coupled dense convolutional network (CDCN) model for planetary gearbox fault identification ^[10].

2. Dataset selection and preprocessing

This study utilizes the bearing data set from Case Western Reserve University in the United States as experimental data ^[11]. The experimental setup is illustrated in **Figure 1**. The test bench comprises a 2 HP (2 horsepower) motor, a torque sensor/encoder, a dynamometer, and control electronics (not shown in the figure).

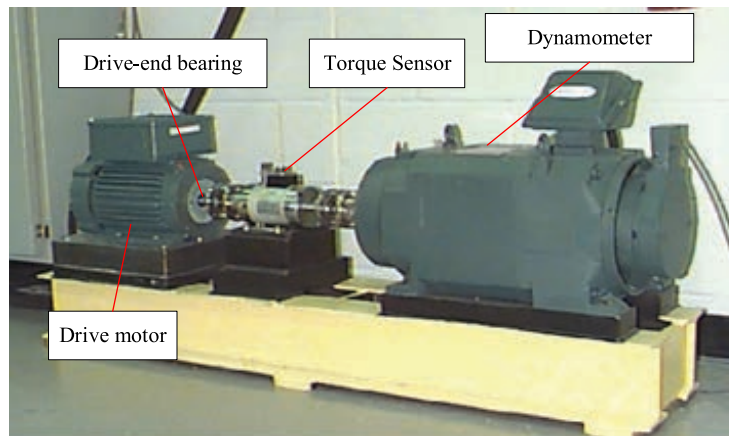


Figure 1. Experimental setup table.

3. Multi-image fusion

As shown in **Table 1** and **Figure 2**, in bearing fault diagnosis, faults cause signals to exhibit strong non-stationary characteristics and cross-scale energy distribution differences. CWT characterizes the temporal location of impact occurrence and its corresponding frequency band response through a combined time-frequency representation,

highlighting local transient phenomena and multi-scale textures.

Table 1. System parameters

Label	Fault location	Image type	Label	Fault location	Image type
0	Normal	CWT	6	IR007	CWT
1	Normal	GADF	7	IR007	GADF
2	Normal	GASF	8	IR007	GASF
3	B007	CWT	9	OR007 @6	CWT
4	B007	GADF	10	OR007 @6	GADF
5	B007	GASF	11	OR007 @6	GASF

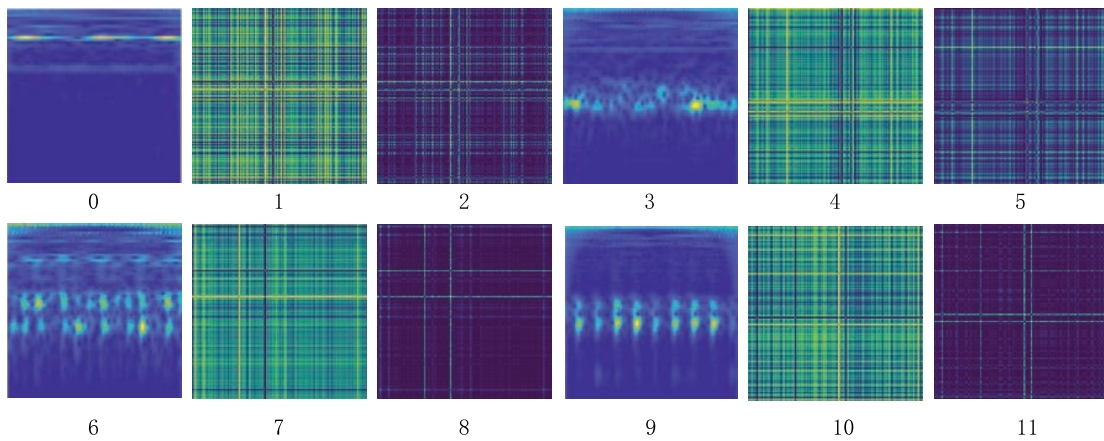


Figure 2. Multi-source images.

The wavelet energy map visually represents the energy distribution across different frequencies, where blue indicates low energy density, green and yellow denote moderate energy levels, and red signifies high energy content. The deeper the color, the greater the energy concentration. GASF depicts the consistency of shape/phase across two time points. If the amplitude variation trends of the sequence are similar at both moments, the corresponding GASF region will exhibit more consistent and continuous texture. GADF expression characterizes the direction and differences in changes at two time points, exhibiting greater sensitivity to mutations, impact edges, and upward/downward directions. Consequently, it tends to produce more pronounced texture contrasts near the impact location. This paper achieves bearing fault localization by transforming time-frequency domain signals into two-dimensional graphs via continuous wavelet transform and Gram angle field, which are then jointly fed as input to a 2DCNN.

4. Model construction

This paper proposes a small-sample bearing fault diagnosis method (MSDR-DAM) based on a dual attention mechanism that integrates multi-image fusion and multi-scale dynamic residuals. This model employs multi-image fusion, multi-scale analysis, dynamic residual learning, and a dual attention mechanism, wherein the dual attention mechanism comprises a channel attention mechanism and a spatial attention mechanism. This fault diagnosis process first transforms the raw fault signal into a time-frequency feature map via continuous wavelet transform

and Gram angular field transform as model input. It then extracts basic features through an initial convolutional layer, followed by deep feature learning through multi-scale dynamic residual blocks and a dual self-attention mechanism. where the multi-scale architecture captures features across distinct receptive fields through parallel convolutional branches, while the dual attention mechanism enables feature recalibration at both the channel and spatial dimensions. Next, a feature pyramid is constructed through two rounds of downsampling, progressively expanding the receptive field while compressing spatial dimensions. Finally, global features are aggregated using global average pooling, and a fully connected classifier outputs fault category probability. The entire training process employs dynamic learning rate adjustment and gradient clipping optimization strategies to ensure model convergence stability and optimal diagnostic accuracy. The overall model framework is illustrated in **Figure 3**.

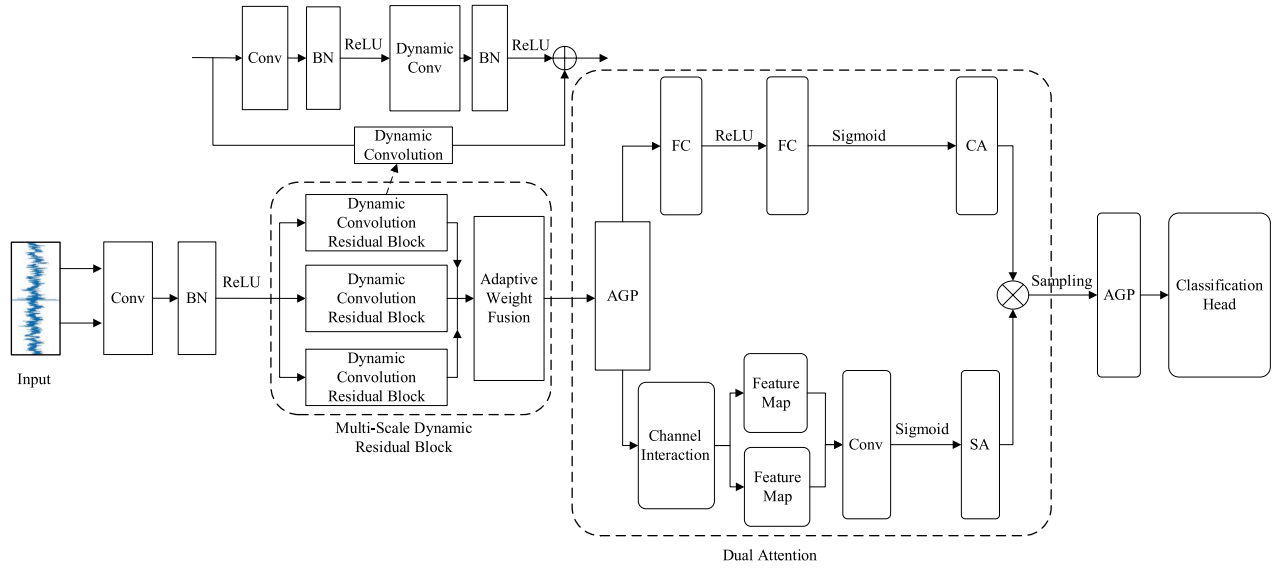


Figure 3. Overall framework of the model

4.1. Multi-scale dynamic residuals

The Multi-scale dynamic residual block is an innovative deep learning module whose core design philosophy lies in organically integrating parallel multi-scale feature extraction with a dynamic weight fusion mechanism. It processes input features through multiple parallel convolutional branches with distinct receptive fields and introduces an adaptive attention mechanism to dynamically adjust the contribution of each branch. Architecturally, the module comprises three parallel convolutional branches employing 1×1 , 3×3 , and 5×5 kernel sizes, respectively. The 1×1 convolution focuses on inter-channel information exchange and feature reorganization, the 3×3 convolution captures local spatial patterns, while the 5×5 convolution acquires a broader receptive field to capture global contextual information. The core innovation lies in its dynamic weight generation mechanism, mathematically expressed as follows:

$$W = \text{Softmax}(W_2 \cdot \text{GAP}(X))$$

where W_2 is the weight matrix for the 1×1 convolution and GAP denotes global average pooling.

The implementation first acquires global spatial information from input features via adaptive average pooling, then compresses the channel dimension to the number of branches using 1×1 convolution to generate

preliminary weights. Finally, the Softmax function ensures the sum of weights across branches equals 1, achieving a probabilistic distribution. During forward propagation, the feature fusion strategy can be expressed as follows:

$$Y = \sum w_i \cdot F_i(X),$$

where w_i is the dynamic weight of the i -th branch and $F_i(\cdot)$ denotes the transformation function of the i -th branch.

In the code implementation, the attention weight tensor is first generated. Then, the outputs of each branch are weighted by their respective weights, and finally, the weighted outputs of all branches are summed. The residual connection design adheres to the identity mapping principle, mathematically expressed as follows:

$$Z = \text{ReLU}(Y + G(X)),$$

where $G(\cdot)$ is the shortcut connection function.

When input and output channel dimensions match, it acts as an identity mapping; otherwise, it adjusts dimensions via 1×1 convolution. This design ensures both smooth gradient flow and the integrity of feature information.

4.2. Dual attention mechanism

This model employs a dual attention mechanism comprising two parallel branches: channel attention and spatial attention. The channel attention branch first compresses the spatial dimension through a global average pooling layer to extract channel-level statistical features. Subsequently, it utilizes a bottleneck structure composed of two fully connected layers to facilitate information exchange between channels. It generates channel attention weights via a sigmoid activation function, enabling adaptive recalibration across different feature channels. Simultaneously, the spatial attention branch generates two spatial feature maps by computing the mean and maximum values along the channel dimension. These maps are concatenated along the channel dimension and passed through a 7×7 convolutional layer to capture large-scale spatial dependencies. The sigmoid function then produces the spatial attention weight map. Ultimately, the channel attention output is multiplied position-wise with the spatial weights, enabling simultaneous fine-tuning of feature maps across both channel and spatial dimensions. This allows the model to adaptively focus on the feature channels and critical regions most relevant to fault diagnosis, significantly enhancing the discriminative power of feature representations.

4.3. Fault identification process

Task identification based on MSDR-DAM specifically involves the following three steps:

- (1) Step 1: Collect bearing signals as raw diagnostic data, construct a dataset encompassing normal operation and various fault conditions, and employ continuous wavelet transform and Gram angular field to generate two-dimensional maps for image fusion;
- (2) Step 2: Input the preprocessed data into the MSDR-DAM model for training. Once the model achieves stable accuracy that meets the criteria, validate its precision using the test set data and save the model's optimal hyperparameters and weights;
- (3) Step 3: Output model diagnostic logs and reports.

5. Experimental verification and analysis

5.1. Experimental platform

To validate the diagnostic effectiveness of the proposed model in this paper, the experimental platform employs an Intel Core i5-14600KF processor, an NVIDIA GeForce RTX 5060 Ti graphics card, and 32 GB of memory. The software environment is based on the CUDA 12.8 acceleration library, with the deep learning framework utilizing the PyTorch programming language on Python 3.9.

5.2. The experimental results

To validate the feasibility of the feature extraction method and diagnostic model under small sample conditions, this experiment utilizes the data from Section 2 for bearing fault diagnosis testing. By inputting data into the MSDR-DAM model, this paper employs a confusion matrix to quantitatively analyze diagnostic outcomes. The horizontal axis represents the diagnostic model's predicted results, while the vertical axis denotes the actual diagnostic results. Elements on the main diagonal indicate the number of samples where predictions match actual outcomes, whereas off-diagonal elements represent misclassified samples. The test diagnostic confusion matrix is shown in **Figure 4**. Samples under different fault categories can be correctly classified, with few misclassified samples, high classification accuracy, and strong generalization capability. Throughout the process, no complex coordinate transformations, expert prior knowledge, or manual debugging experience are required to successfully diagnose the location of the fault. This demonstrates that the method proposed in this paper has certain feasibility for small-sample bearing fault diagnosis.

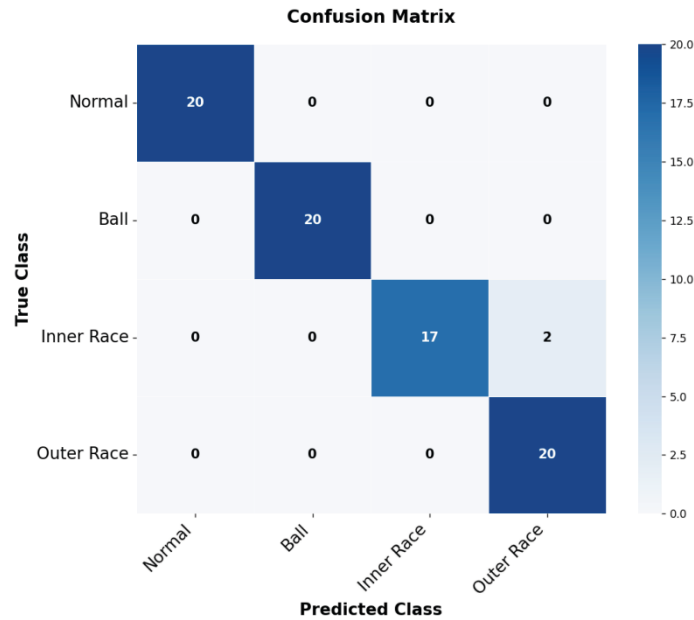


Figure 4. Confusion matrix.

As shown in **Table 2**, the proposed model achieves an accuracy of 98.43% in IGBT localization, demonstrating exceptionally high diagnostic precision. Furthermore, the high values of precision, recall, and F1 score confirm the model's capability in identifying samples.

Table 2. Comparative experiment

Diagnosis method	Accuracy	Precision	Recall	F1 score
1D-CNN	91.03%	92.50%	90.56%	90.69%
2D-CNN	95.00%	95.49%	95.00%	94.89%
Proposed Method	97.47%	97.73%	97.37%	97.42%

Table 2 demonstrates that the proposed model exhibits significant advantages across multiple key performance metrics. Compared to the 1D-CNN model, our approach enhances diagnostic accuracy by converting fault signals into two-dimensional data. Compared to the 2D-CNN model, this paper introduces multi-image fusion and an attention mechanism. This mechanism guides the network to more effectively capture discriminative features related to faults by simultaneously computing channel attention and spatial attention, thereby enhancing the diagnostic performance and robustness of the convolutional network. This model fusion approach enables the MSDR-DAM diagnostic model to excel in accuracy, efficiency, stability, and resource consumption for small-sample fault diagnosis in bearings.

6. Conclusion

This paper proposes a small-sample bearing fault diagnosis method based on multi-image fusion and a dual-attention mechanism for multi-scale dynamic residuals. The multi-scale dynamic residual module adaptively fuses fault features with varying receptive fields to capture multi-scale patterns ranging from transient impacts to steady-state modulations. The dual-attention mechanism for both channels and spatial dimensions simultaneously calibrates feature responses along these dimensions, precisely focusing on fault-sensitive regions. To validate performance, the dataset from the Bearing Data Center at Case Western Reserve University was employed for experiments. Results demonstrate that the proposed method achieves high-accuracy diagnosis even under minimal sample conditions, with bearing fault localization accuracy reaching 97.47%, significantly outperforming traditional models such as 1D-CNN and 2D-CNN.

Funding

The National Natural Science Foundation of China (Project No.: 61404049, 62273141); Natural Science Foundation of Hunan Province (Project No.: 2020JJ6031)

Disclosure statement

The author declares no conflict of interest.

References

- [1] Zhang T, Liu S, Zhang S, 2021, Review on Fault Diagnosis on the Rolling Bearing. Journal of Physics: Conference Series. IOP Publishing, 1820(1): 012107.
- [2] Zhang S, Ye F, Wang B, et al., 2019, Semi-Supervised Learning of Bearing Anomaly Detection via Deep Variational

Autoencoders. arXiv, <https://doi.org/10.48550/arXiv.1912.01096>

- [3] Zhang C, Guo J, Zhen D, et al., 2020, Rolling Element Bearing Fault Diagnosis based on the Wavelet Packet Transform and Time-Delay Correlation Demodulation Analysis. *Advances in Asset Management and Condition Monitoring: COMADEM 2019*, 1195–1203.
- [4] Xue B, Yi W, Jing F, et al., 2021, Complex ISAR Target Recognition using Deep Adaptive Learning. *Engineering Applications of Artificial Intelligence*, 2021(97): 1–9.
- [5] Wong P, Zhong J, Yang Z, et al., 2016, A New Framework for Intelligent Simultaneous-Fault Diagnosis of Rotating Machinery Using Pairwise-Coupled Sparse Bayesian Extreme Learning Committee Machine. *Proceedings of The Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 231(6): 1146–1161.
- [6] Safizadeh M, Yari B, 2017, Pump Cavitation Detection Through Fusion of Support Vector Machine Classifier Data Associated with Vibration and Motor Current Signature. *Insight*, 2017(59): 669–673.
- [7] Tao H, Wang P, Chen Y, et al., 2020, An Unsupervised Fault Diagnosis Method for Rolling Bearing Using STFT and Generative Neural Networks. *Journal of the Franklin Institute*, 357(11): 7286–7307.
- [8] Han Y, Tang B, Deng L, 2018, Multi-Level Wavelet Packet Fusion in Dynamic Ensemble Convolutional Neural Network for Fault Diagnosis. *Measurement*, 2018(127): 246–255.
- [9] Xia M, Li T, Xu L, et al., 2017, Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks. *IEEE/ASME Transactions on Mechatronics*, 2017(99): 1.
- [10] Jiao J, Zhao M, Lin J, et al., 2019, Deep Coupled Dense Convolutional Network with Complementary Data for Intelligent Fault Diagnosis. *IEEE Transactions on Industrial Electronics*, 2019: 1.
- [11] Smith W, Randall R, 2015, Rolling Element Bearing Diagnostics using the Case Western Reserve University Data: A Benchmark Study. *Mechanical Systems and Signal Processing*, 2015(64): 100–131.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Overview of the Integration of Large Language Models, Knowledge Graphs, and GraphRAG, along with Research on Their Industrial Applications

Xian Ye*

School of Mine Safety, North China Institute of Science and Technology, Yanjiao 101601, Hebei, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In recent years, Large Language Models (LLMs) have rapidly advanced in language understanding, reasoning, and generation, and are increasingly adopted as the “brain” of industrial intelligent systems. Nevertheless, in high-risk and strongly regulated domains they still exhibit hallucination, weak domain grounding, limited interpretability, and privacy as well as security constraints. Knowledge graphs (KGs) encode domain entities, relations, rules, and events explicitly, providing controllable semantics and an explainable reasoning substrate. Retrieval-augmented generation (RAG) injects external evidence into LLM prompting, while GraphRAG further introduces graph indexing and community-level retrieval to preserve global structure and support multi-hop reasoning. This review summarizes the evolution of LLMs, KG modeling and extraction, GraphRAG mechanisms, and a general fusion framework. Typical industrial applications are surveyed, and a coal mine flood emergency plan generation and evaluation approach is discussed to illustrate the practical value of graph-grounded large models. KG-enhanced retrieval also supports provenance tracking, allowing industrial users to audit the evidence behind model outputs.

Keywords: Large language model; Knowledge graph; GraphRAG; Retrieval-augmented generation; Coal mine flood; Emergency management

Online publication: February 13, 2026

1. Introduction

Large model-driven general AI has shifted industrial systems from pipeline-style NLP modules to unified model-centric architectures. LLMs offer strong zero-/few-shot transfer, natural human-computer interaction, and the ability to synthesize multi-source information, making them suitable as universal interfaces for production, governance, and safety services. However, industrial tasks emphasize factual correctness, timeliness, traceability, and compliance. Pure LLMs cannot reliably store long-tail or rapidly changing professional knowledge, and their black-box generation may lead to unsafe suggestions in critical decision processes. Existing work therefore

combines LLMs with external knowledge. RAG improves factuality by retrieving evidence from document stores, but vector-only retrieval may fragment long documents and ignore structural relations, limiting global understanding and multi-hop reasoning. KGs provide structured semantics and explicit causal/procedural connections, yet traditional KG-based QA often depends on brittle symbolic pipelines and heavy manual schema design. GraphRAG integrates graph structure into retrieval and reasoning, enabling global-aware evidence selection and explainable paths, and thus becomes a promising foundation for trustworthy industrial LLM systems. This paper reviews key techniques and proposes an LLM + KG + GraphRAG fusion route for industry scenarios, with a focus on coal mine flood emergency management.

2. Development and industrialization trends of large language model technology

2.1. Evolution of general large models

Since the transformer architecture, pretraining on massive corpora plus instruction tuning and alignment (e.g., RLHF or preference optimization) has produced foundation models with strong reasoning and generation capabilities. Research trajectories include scaling laws, long-context modeling (positional encoding variants, sparse attention, retrieval-memory hybrids), mixture-of-experts for efficiency, distillation/quantization and speculative decoding for deployment, and tool/function calling for structured actions. Inference optimization and hardware acceleration have reduced serving cost, while multimodal models extend LLMs to images, audio, and sensor-like signals, supporting richer industrial inputs such as inspection photos, charts, or monitoring reports. Despite this progress, general models remain weak on domain terminology, localized standards, and task-specific workflows. Their performance degrades under distribution shift (new devices, new policies, unseen hazards), and safety alignment for professional scenarios is still insufficient. Therefore, domain customization and reliable knowledge grounding are indispensable for industrial use.

2.2. Industry large models and task customization practice

Industrial adaptation usually follows two intersecting paths: domain fine-tuning and knowledge-grounded augmentation. Fine-tuning uses curated professional corpora and parameter-efficient methods (LoRA, adapters) to strengthen domain vocabulary, response style, and constraint compliance; continual learning is used to follow evolving regulations and incident patterns. Knowledge augmentation connects LLMs to enterprise databases and document stores through RAG, KGs, and agent workflows so that answers are evidence-based and controllable. In industry and smart-city contexts, private data are heterogeneous and sensitive; deployment thus stresses on-premise or secure-cloud serving, access control, and data-quality governance. For mine safety, models are customized to understand emergency plans, accident reports, geological texts, and monitoring indicators, and to generate disposal suggestions aligned with standards. LLMs also assist KG construction by extracting entities/relations from long technical documents, reducing manual labeling cost and enabling faster knowledge updates. For library, information, and scientific services, LLMs act as semantic organizers and report writers on top of curated KGs, improving literature retrieval and thematic analysis.

3. Knowledge graph and structured knowledge extraction technology

3.1. Modeling elements of industry knowledge graph

An industry KG encodes domain entities, attributes, relations, and events under a shared ontology. Entities

may represent equipment, hazards, resources, organizations, policies, or cases. Relations include hierarchical, causal, temporal, spatial, procedural, and constraint links. Industrial KGs emphasize multi-source fusion (plans, logs, sensors), spatiotemporal evolution, uncertainty/confidence, and versioned updates, making them suitable for decision support, risk assessment, and post-event auditing. Schema design usually aligns with standards or regulations, and quality assurance relies on constraint rules, consistency checking, and iterative expert validation.

3.2. Knowledge extraction based on pre-trained models and LLMs

KG construction requires NER, relation/event extraction, entity linking, and coreference resolution. Traditional supervised IE provides stable baselines but is costly to label. Pretrained encoders and seq2seq extractors reduce data needs, while LLM prompting enables few-shot or open IE, structured triple/JSON outputs, and automatic schema induction. In practice, hybrid pipelines are common, where statistical or rule extractors provide high-precision seeds; LLMs expand coverage by self-asking and self-correcting; normalization and deduplication fuse results; and human auditing remains the final safety gate. LLMs are also used for knowledge completion and conflict detection, supporting continuous KG evolution.

4. Technical route of 4RAG and GraphRAG

4.1. Vector RAG framework and its limitations

Vector RAG embeds text chunks, retrieves top-k similar units, and feeds them into LLM prompts. It enhances factuality for single-hop questions but struggles with complex industrial corpora as follows:

- (1) Fixed chunking breaks long-range coherence, and top-k evidence may miss the global narrative;
- (2) Embeddings emphasize topical similarity but may overlook explicit causal or procedural relations;
- (3) Retrieved snippets can be redundant or inconsistent, so the model may still hallucinate when integrating evidence.

These limits are evident in tasks requiring cross-document causal tracing, such as policy coordination, equipment fault localization, or emergency decision making.

4.2. Core idea and implementation process of GraphRAG

GraphRAG constructs a graph over text units and/or KG nodes to preserve structure during retrieval. Offline, it does as follows:

- (1) Segments documents into base units;
- (2) Extracts entities and relations for linking;
- (3) Builds graph edges using KG relations and/or semantic similarity;
- (4) Applies community detection to obtain topic-level subgraphs;
- (5) Summarizes each community to form a compact global representation.

Online, a query retrieves relevant community summaries to locate the correct knowledge region, and then expands into local nodes and paths for fine-grained evidence. By combining global summaries with local facts, GraphRAG improves multi-hop retrieval stability, reduces redundancy, and provides interpretable evidence chains.

4.3. Integration mode of GraphRAG and knowledge graph

Three integration patterns are observed as listed:

- (1) Document-graph GraphRAG builds graphs mainly from unstructured corpora for domains without mature

KGs; it is easy to deploy but less controllable;

- (2) KG-enhanced GraphRAG uses a domain KG as the core graph and links documents to KG nodes, enabling rule-consistent retrieval and reasoning;
- (3) Hybrid GraphRAG merges document graphs and KGs into a heterogeneous graph, combining semantic proximity search with symbolic multi-hop traversal.

The choice depends on KG availability, update frequency, interpretability requirements, and engineering cost.

5. General fusion framework of LLM + KG + GraphRAG

5.1. Offline stage: Multi-source knowledge modeling and graph index construction

The offline stage ingests domain plans, regulations, cases, technical manuals, databases, and logs. Documents are cleaned and segmented into base text units; entities, relations, and events are extracted to build or update the KG. Text units are embedded and linked to KG nodes, producing a heterogeneous retrieval graph. Community detection and summarization create a multi-level index supporting global overview plus local evidence access. Graph and vector indexes are stored for fast online retrieval and are periodically refreshed as new data arrives.

5.2. Online stage: Query comprehension, graph retrieval, and generative reasoning

Given a user query, the system performs intent and entity parsing, retrieves relevant community summaries or KG subgraphs, and expands along graph paths to gather supporting text. Evidence is re-ranked and organized into structured context (facts, relations, timelines, procedures). The LLM then generates answers under evidence constraints, optionally using self-verification, rule checks on KG paths, and tool calls (database lookup, calculation, simulation). Outputs can include cited evidence and reasoning paths for expert auditing and downstream execution.

5.3. Coal mine flood emergency plan generation and evaluation

For coal mine flood disasters, knowledge sources include emergency plans, laws, historical accidents, hydrology/geology reports, and monitoring indicators. A safety schema models water-inrush sources, precursor signals, affected equipment, response teams, and disposal actions with causal and temporal links. GraphRAG retrieves scenario-matched subgraphs and similar accident communities, providing grounded context on hazard evolution and proven measures. The LLM produces structured, stepwise emergency plans (monitoring–alarm–confirmation–control–evacuation–rescue–recovery). Plan evaluation can leverage KG rule compliance, evidence coverage, and similarity to validated historical cases.

6. Overview of industry applications

6.1. Analysis of government data governance and policy coordination

Zhu and Qin built a policy knowledge graph and adopted GraphRAG for government data governance and policy coordination^[1]. Their method converts policy texts into a unified graph space, uses community division to cluster policy topics, and performs path-based retrieval to trace cross-policy dependencies. GraphRAG locates relevant clauses quickly and provides multi-document evidence chains, while the LLM assists semantic interpretation and policy-effect analysis, improving policy synergy and execution efficiency.

6.2. Industrial private data and smart city knowledge services

Hu *et al.* proposed an ALBERT-XL pipeline for constructing KGs from industrial private data, enabling knowledge extraction without exposing raw sensitive logs ^[2]. Based on industrial KGs, Li *et al.* implemented a GraphRAG-based enterprise private knowledge base for construction projects, enhancing cross-document retrieval, fault troubleshooting, and experience reuse ^[3]. In smart-city governance, similar KG + GraphRAG services unify heterogeneous data for infrastructure management, incident response, and citizen Q&A, and LLMs translate retrieved evidence into actionable and understandable recommendations.

6.3. Library and information services and domain knowledge services

Xie *et al.* introduced TTKE-LLM, using LLM prompting plus engineering constraints to extract tourism entities and relations and to accelerate KG construction ^[4]. Ma *et al.* combined LLMs with GraphRAG to generate graph-guided abstracts for scientific literature, turning scattered papers into structured knowledge and improving literature navigation ^[5]. Compared with keyword-based intelligence analysis, KG + GraphRAG enables semantic clustering, relationship tracing, and explainable answers to complex queries, while LLMs synthesize readable intelligence reports.

6.4. Agricultural pest and disease control and expert system

Wu *et al.* developed a rice pest-and-disease expert system by coupling agricultural KGs with LLM reasoning ^[4]. Symptoms, varieties, pesticides, and environmental factors are encoded as KG nodes/relations, and GraphRAG retrieves multi-hop evidence for diagnosis. The LLM serves as a hypothesis generator and interactive expert, revising conclusions with retrieved evidence to improve interpretability and practical usability.

6.5. Power system, autonomous driving and other scenarios

In power systems, Liu *et al.* applied LLMs to electric-vehicle charging and swapping load forecasting ^[5]. With KG/RAG context on equipment states, user behavior, and grid constraints, their approach yields more robust and interpretable forecasts. In autonomous driving, Song *et al.* summarized large-model decision/planning progress, and Wu *et al.* further integrated vehicle KGs with LLMs to support multi-scenario decision making and safer reasoning ^[6,7]. Ai *et al.* introduced a GraphRAG-based assistant for spacecraft fault localization, showing that graph-guided retrieval can improve cross-document fault reasoning ^[8].

6.6. Mine accidents and safety management

For mining, Zhang *et al.* constructed a mine-accident KG using LLM-assisted extraction, capturing environment, causal factors, losses, and response measures, which supports case retrieval and risk pattern mining ^[9]. Xu *et al.* designed an LLM-based coal-mine safety assistant; combined with KG/GraphRAG retrieval, it assists hazard identification, case recall, and on-site decision suggestions with traceable evidence ^[9]. These results indicate that graph-grounded LLM systems can improve the safety training effectiveness and emergency disposal efficiency.

7. Key challenges and development trends

7.1. Challenges at the data and graph level

Industrial corpora are noisy and heterogeneous, and KG construction faces schema inconsistency, entity ambiguity, sparse labeling, and high expert cost. Continuous updates, temporal reasoning, and uncertainty modeling remain

difficult, and extraction errors may propagate into retrieval and generation. Future systems need automatic quality assessment, conflict resolution, and efficient human-in-the-loop editing to maintain graph reliability.

7.2. Challenges at the model and system levels

GraphRAG introduces extra latency, storage, and engineering complexity. Retrieval quality depends on graph construction, community division, and linking accuracy, while LLMs may still hallucinate when evidence is weak or contradictory. Efficient deployment therefore requires lightweight heterogeneous graph indexing, caching, adaptive retrieval depth, and model compression, together with strict privacy/access control for sensitive data.

7.3. Evaluation, compliance and future directions

Benchmarks for graph-grounded generation are scarce. Evaluation should measure factuality, evidence faithfulness, reasoning-path correctness, robustness, and expert trust, ideally via scenario-based tests. Compliance concerns cover data security, copyright, and domain safety regulations. Future research will likely focus on automated schema learning, spatiotemporal and multimodal GraphRAG, causal/functional graphs for decision support, joint graph-LLM training, and agentic multi-step planning-retrieval-generation with human oversight.

8. Conclusion

LLMs provide powerful language interfaces, KGs supply explicit and explainable domain knowledge, and GraphRAG enables global-aware multi-hop retrieval and path-based reasoning. Their fusion improves grounding, interpretability, and robustness for industrial intelligence. In coal mine flood emergency management, a KG-enhanced GraphRAG framework can generate and evaluate structured emergency plans aligned with regulations and verified cases, offering a feasible direction for intelligent emergency disposal. Overall, the LLM + KG + GraphRAG paradigm offers a balanced path between neural flexibility and symbolic controllability for future safety-critical industrial AI.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Zhu X, Qin D, 2025, From Evolution to Coordination: Exploring Policy Effectiveness-Enhancement Paths for Government Data Governance under the GraphRAG Framework. *Information and Documentation Services*, 46(5): 89–101.
- [2] Hu D, Wu X, Liu X, et al., 2025, Knowledge Graph Construction Method for Industrial Private Data Based on ALBERT-XL. *Smart Cities*, 11(7): 10–14.
- [3] Li G, Wu Z, Huang J, 2025, Research and Implementation of an Enterprise-Level Private Knowledge Base in the Construction Domain Based on GraphRAG. *Computer Knowledge and Technology*, 21(20): 23–25.
- [4] Wu H, Chen M, Guo Q, 2025, Rice Pest and Disease Expert System Based on Large Language Models and Knowledge Graphs. *Transactions of the Chinese Society of Agricultural Engineering*, 41(22): 244-255.
- [5] Song Z, Chen J, Jiang L, et al., 2025, Research Progress Review of Large-Model-Based Autonomous Driving

Decision-Making and Planning. *Automotive Technology*, 2025(10): 21–31.

- [6] Wu S, Jiang H, Huang K, et al., 2025, Vehicle Decision-Making System Based on Knowledge Graphs and Large Language Models. *Journal of Beijing Institute of Technology*, 6–12.
- [7] Ai S, He Y, Zhang W, et al., 2025, A GraphRAG-Based Method for Assisted Spacecraft Fault Localization. *Spacecraft Engineering*, 34(4): 84–90.
- [8] Zhang P, Sheng L, Wang W, et al., 2025, Construction of a Mine Accident Knowledge Graph Based on Large Language Models. *Industry and Mine Automation*, 51(2): 76–83.
- [9] Xu J, Gong Q, Xiang Y, et al., 2024, Design and Application of a Coal Mine Safety Assistant based on LLM. *Inner Mongolia Coal Economy*, 2024(1): 133–135.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Collaborative Quality Management in Industrial Engineering from a Supply Chain Perspective: AI-Driven Enterprise Quality Optimization

Jiakai Zhong*

School of Mining Engineering, China University of Mining and Technology, Xuzhou 221116, Jiangsu, China

*Corresponding author: Jiakai Zhong, 1670469597@qq.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Amidst the intensifying digital economy and global competition, supply chain quality management is evolving from traditional linear models toward networked systems characterized by data-driven and intelligent collaboration. This paper constructs an AI-driven “Supply Chain Quality Collaborative Management” framework through system optimization and artificial intelligence analytical capabilities from a supply chain perspective. The study first analyzes core challenges in supply chain quality collaboration across three dimensions: data fragmentation, standard discrepancies, and mechanism asymmetry. It highlights that traditional static and reactive quality controls struggle to adapt to complex, dynamic supply chain ecosystems. Subsequently, through systematic literature review and theoretical synthesis, the paper elucidates AI’s role in multi-source quality data fusion, semantic alignment, standardized governance, and intelligent incentives. It proposes collaborative optimization pathways based on deep learning, blockchain, and reinforcement learning. Through case studies in the automotive and pharmaceutical industries, the research validates the feasibility of AI in predictive maintenance and cross-linkage collaborative decision-making, demonstrating AI’s ability to significantly enhance the systemic resilience and decision-response capabilities of quality management. This paper innovatively integrates industrial engineering process optimization with cross-organizational governance mechanisms for supply chain quality management, providing a new theoretical framework and practical pathway for intelligent manufacturing and sustainable supply chain development.

Keywords: Supply chain quality management; Industrial engineering; Artificial intelligence; Quality collaborative optimization; Data-driven decision-making

Online publication: February 12, 2026

1. Introduction

Amidst of economic globalization and diversifying market demands, enterprises face dual pressures to reduce costs and enhance competitiveness. Consequently, supply chain quality management has become a core issue

determining corporate competitiveness and risk resilience. Against the backdrop of accelerating global industrial transformation toward digitalization, networking, and intelligence, traditional linear supply chains are evolving into mesh ecosystems characterized by “data-driven-platform-collaborative” features. The focus of corporate competition has shifted from individual product performance to a comprehensive contest of overall supply chain real-time responsiveness and collaborative efficiency. However, within complex supply chain networks, quality issues at any link can propagate through upstream and downstream channels, amplifying their impact (e.g., the bullwhip effect or ripple effect), ultimately affecting overall output quality and customer satisfaction. Traditional quality management models, predominantly static, linear, and post-event inspection-based, struggle to adapt to dynamic external environments due to information silos and the absence of real-time sharing and cross-organizational collaboration mechanisms. Industrial Engineering (IE), emphasizing system optimization, process standardization, and statistical process control, exhibits inherent synergy with supply chain quality management. Both disciplines prioritize data-driven decision-making, closed-loop control systems, and cross-organizational collaboration. This convergence enables holistic quality robustness and overall optimization through localized improvements. Consequently, establishing an integrated, preventive, and platform-based supply chain quality collaboration, management system has become the critical pathway to mitigate quality transmission risks and overcome traditional management bottlenecks.

Supply chain quality management still faces numerous practical challenges. Quality data sources are highly dispersed and heterogeneous, with inconsistent standards, formats, and interfaces across different segments, leading to information asymmetry, traceability difficulties, and delayed decision-making. Cross-enterprise quality standards vary, and differing implementation and inspection methods impact collaboration efficiency and consistency in quality assessment. Furthermore, in dynamic and uncertain environments, supply chain collaboration mechanisms lack resilience, struggling to respond to demand fluctuations or even supply chain disruptions triggered by geopolitical tensions or public health events. Against this backdrop, this study focuses on the intersection of industrial engineering and supply chain management. It introduces artificial intelligence (AI) technology to construct a cross-enterprise quality collaboration analysis framework. Leveraging AI enables deep mining of quality data, predictive optimization, and intelligent decision-making. This approach deeply integrates the logic control with the concept of cross-organizational collaborative governance in SCM. It not only of IE, emphasizes process optimization and quality, fills a research gap in AI-enabled supply chain quality collaboration theory, but also provides a new research paradigm for quality management in complex supply chain scenarios.

2. Theoretical foundations and literature review

2.1. Supply chain quality management theory

Supply chain quality management (SCQM), serving as the core link connecting internal quality control and external collaboration, has seen its theoretical framework evolve alongside advancements in industrial structures and management paradigms. Early traditional quality management systems, such as total quality management (TQM), emphasized principles of full employee participation, process orientation, and continuous improvement, playing a vital role in enhancing internal quality standards. As supply chain integration deepens, TQM principles have gradually expanded to the supply chain level, forming a multi-stakeholder collaborative governance quality management system. Six Sigma, centered on the DMAIC model, achieves high-precision quality control at critical supply chain nodes (such as raw material screening, production assembly, and logistics) by reducing process

variation and optimizing key processes. The introduction of collaborative management theory further advances the systematization and networking of supply chain quality governance. Its core lies in achieving coordinated optimization and dynamic responsiveness among upstream and downstream enterprises through unified standards, data sharing, and real-time monitoring mechanisms. With the integrated development of industrial internet, big data, and AI technologies, supply chain quality management is progressively evolving toward intelligent and data-driven approaches. This promotes supply chain transparency and resilience enhancement, establishing it as a critical support system for driving high-quality and sustainable enterprise development.

2.2. Quality optimization methods in industrial engineering

Quality optimization in industrial engineering can be summarized under two core pillars: “shop floor monitoring” and “upstream design.” At the shop floor level, statistical process control (SPC) serves as the core methodology. By employing Shewhart and EWMA control charts to distinguish common from special cause variation, stable monitoring of the manufacturing process is achieved for special cause. Within the Industry 4.0 framework, SPC integrates with IoT sensors and MES systems to enable real-time data collection, automated charting, and closed-loop early warning ^[1]. At the design level, quality function deployment (QFD) and failure mode and effects analysis (FMEA) are widely employed for requirement translation and risk pre-control. However, methods have certain limitations: QFD is prone to mapping deviations under mass customization and rapidly fluctuating requirements, while demanding high levels of interdisciplinary collaboration and data resource support. Meanwhile, the risk priority number (used in traditional FMEA RPN) often distorts rankings due to the multiplication of ordinal metrics and heavily relies on expert judgment. To address these challenges, current research is evolving toward fuzzy/multi-criteria coupling methods, time-updated “dynamic FMEA”, and digital twin-driven online risk assessment to enhance responsiveness and timeliness ^[2,3].

2.3. Current applications of AI in supply chain quality management

With the rapid advancement of AI technologies, the research paradigm in supply chain quality management is progressively shifting from experience-driven approaches toward data-driven and intelligent decision-making orientations. AI provides new technical support for quality control, risk prediction, and supply chain collaborative optimization through methods such as machine learning, deep learning, and intelligent reasoning. Machine learning models, leveraging their automated feature extraction and pattern recognition capabilities, can effectively predict product defect rates and optimize process control. For instance, random forest (RF) and support vector machines (SVM) can accurately predict defect occurrence probabilities and generate actionable insights for supply chain optimization ^[4]. Long short-term memory (LSTM) networks are suitable for time series modeling, enabling demand fluctuation forecasting and early supply chain risk warnings; while Bayesian networks provide intelligent reasoning frameworks for root cause analysis and optimal strategy generation in complex systems.

The convergence of AI and supply chain quality collaboration has emerged as a frontier topic of shared interest across academia and industry. AI-driven data fusion and sharing mechanisms provide tamper-proof transaction records, enhancing supply chain transparency and traceability. This boosts overall supply chain resilience and enables real-time decision-making and optimization ^[5]. Blockchain technology plays a pivotal role in ensuring data trustworthiness and traceability, delivering high-quality, transparent, and flexible training data for AI models ^[6]. Natural language processing technology can parse quality documents within and across organizations, enabling semantic-level standard mapping and specification unification; reinforcement learning

supports dynamic decision-making in multi-agent environments, driving supply chains toward adaptability and continuous improvement in complex settings. However, current research and practice face limitations, often focusing on localized AI applications or single nodes while lacking a systematic supply chain quality collaboration framework. Data security and privacy protection mechanisms remain underdeveloped, with cross-enterprise data sharing encountering legal and trust barriers.

3. Challenges in supply chain quality collaborative management and AI-empowered pathways

3.1. Core challenges

Supply chain quality collaboration management in practice faces multidimensional challenges, primarily concentrated across three dimensions: data, standards, and mechanisms. At the data level, numerous participants with widespread distribution result in fragmented quality data characterized by inconsistent formats and incompatible standards. Concurrently, corporate concerns regarding data security and privacy protection limit the depth and breadth of information sharing. At the standards level, differences create barriers to information exchange, weakening collaborative effectiveness in quality management norms, testing methods, and defect criteria among enterprises. To achieve efficient coordination, it is imperative to establish unified quality management systems and data exchange standards. At the mechanism level, unequal distribution of benefits among supply chain members often leaves some suppliers lacking intrinsic motivation for quality improvement. Combined with imperfect collaborative decision-making and constraint mechanisms, this hinders the formation of sustained, effective momentum for quality collaboration.

3.2. AI-empowered solutions

AI technology offers a systematic solution to the complex challenges of collaborative quality management in supply chains. By integrating multi-source data, ensuring data security, unifying quality standards, and optimizing collaborative mechanisms, AI can significantly enhance supply chain efficiency, sustainability, and systemic resilience^[7]. Deep learning and natural language processing (NLP) technologies enable the fusion and semantic alignment of multi-source, heterogeneous quality data, supporting unified modeling of quality information across different stages and formats. This improves data utilization efficiency and analytical accuracy^[8,9]. Blockchain and smart contract technologies ensure data security at the mechanism level during circulation, traceability, and immutability, establishing a trusted foundation for cross-organizational collaboration^[10]. At the standards level, AI aligns and automates the semantic interpretation of quality standards, inspection metrics, and defect classifications, driving the formation of standardized quality management systems^[11]. Reinforcement learning enables intelligent incentives for quality improvement and resource allocation through real-time data-driven dynamic decision-making and adaptive optimization mechanisms in multi-agent collaborative environments, enhancing system responsiveness and stability^[12,13]. Research indicates that AI-empowered supply chain collaborative optimization not only reduces defect rates and operational risks but also significantly boosts overall performance and long-term sustainability.

4. Industry case analysis

4.1. Automotive industry case

In automotive manufacturing, AI applications play critical roles in process quality control, planning and inventory

optimization, and equipment maintenance. Time-series machine learning predicts geometric deviations in structural component hole positions during work-in-process stages, achieving approximately 15% higher prediction accuracy than baseline models while reducing tolerance violations ^[14]. Demand forecasting/inventory planning systems reduce manufacturing inventory costs by 10–20% and suppress supply fluctuations by about 12% ^[15]. Predictive maintenance combining AHP and PFMEA for failure mode identification with random forest modeling achieved nearly 80% accuracy in predicting failure cycles on actual production lines, enabling proactive interventions for unplanned downtime ^[16].

4.2. Pharmaceutical industry case study

Within pharmaceutical supply chains, AI is playing a role in temperature-controlled quality assurance and proactive recall risk assessment. Supervised learning based on route, time slot, and temperature sensor data can identify cold chain boundary risks before shipment. A Brazilian cross-regional transport study with 7,078 samples demonstrated that a multi-layer perceptron achieved 93.7% correct classification on imbalanced data with a Kappa statistic of 0.82. This directly supports scheduling and route adjustment decisions, establishing a “preemptive handling” mechanism for temperature control risks ^[17]. Recall risk scoring tailored to product and process complexity employs methods like the Minimum Absolute Shrinkage and Selection Operator to screen key descriptors (e.g., route of administration, dosage form, release pattern, half-life). With an accuracy rate of approximately 71%, this approach enables early identification of quality vulnerabilities throughout the product lifecycle and provides quantitative thresholds for change management ^[18].

5. Challenges and future research directions

5.1. Implementation challenges

Despite offering innovative opportunities such as predictive analytics, real-time monitoring, and automated decision-making for collaborative supply chain quality management, the implementation of AI still faces multiple obstacles, primarily centered around data sharing, technological transparency, and collaborative incentives. Research indicates that enterprises generally exhibit low willingness to share sensitive data, which may hinder AI applications in cross-organizational quality collaboration, although governance frameworks can mitigate such issues to some extent. The “black box” nature of AI models erodes managerial trust, leading to concerns about decision-making transparency and limiting practical adoption ^[20]. At the mechanism level, cross-enterprise incentive designs remain immature, with power asymmetries and misaligned motivations making it difficult for suppliers to actively engage in quality improvement.

5.2. Future directions

Future research may expand along the following directions to systematically advance AI-enabled collaborative supply chain quality management. Regarding data collaboration and privacy protection, distributed machine learning techniques like federated learning hold significant application potential. This approach enables multiple parties without sharing raw data to jointly build AI models, effectively balancing data utilization with commercial privacy protection and providing a feasible technical pathway for cross-enterprise quality data analysis and collaborative modeling. At the process mapping and dynamic simulation level, the deep integration of digital twins and AI will construct dynamic virtual representations of quality states across all supply chain segments. Such systems can map quality behaviors in physical supply chains in real time, simulate disturbance impacts, and

predict potential risks, thereby providing high-fidelity, iterative simulation environments for quality early warning, strategy evaluation, and decision optimization. Regarding collaborative mechanisms and standardization, policy guidance and industry consensus should be strengthened to promote the development of cross-industry, cross-platform quality data sharing standards and collaborative decision-making processes. Unified data standards and collaborative frameworks will reduce system integration costs, enhance information exchange efficiency, and thereby boost the overall synergistic effectiveness of the supply chain ecosystem.

6. Conclusion

This study systematically investigates AI-enabled collaborative quality management within supply chains, proposing an integrated analytical framework that merges industrial engineering optimization principles with supply chain coordination mechanisms. It identifies and articulates feasible AI-driven pathways to overcome critical challenges, including cross-enterprise quality data silos, standardization gaps, and weak collaborative incentives. Through case studies in the automotive and pharmaceutical sectors, the research demonstrates AI's tangible benefits in enhancing operational efficiency and system reliability. Applications in quality monitoring, inventory optimization, equipment maintenance, and risk management highlight AI's capacity for precise prediction and proactive control. Notwithstanding these advances, the widespread implementation of AI in supply chain quality management continues to encounter obstacles such as data-sharing barriers, limited model interpretability, and underdeveloped incentive structures. Nevertheless, emerging technological frontiers, such as federated learning, dynamic FMEA, multi-modal data fusion, and policy-standard alignment, are steadily steering supply chain quality governance toward real-time responsiveness, adaptive decision-making, and trusted collaboration. Looking ahead, as AI technologies continue to evolve and cross-organizational governance models mature, AI-facilitated collaborative quality management is poised to achieve end-to-end optimization and strengthen systemic resilience across increasingly complex supply chain ecosystems. This progression will provide critical support for the transition toward high-quality and sustainable enterprise growth, establishing a new benchmark for intelligent supply chain operations in the digital era.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Dieguez T, Malheiro M, Leal N, et al., 2025, Systematic Literature Review on Manufacturing Execution Systems in the Era of Industry 4.0: A Bibliometric Analysis. *Innovations in Industrial Engineering IV*, 298–310.
- [2] Di Nardo M, Murino T, Osteria G, et al., 2022, A New Hybrid Dynamic FMECA with Decision-Making Methodology: A Case Study in an Agri-Food Company. *Applied System Innovation*, 5(3): 45.
- [3] Zio E, Miqueles L, 2024, Digital Twins in Safety Analysis, Risk Assessment and Emergency Management. *Reliability Engineering & System Safety*, 2024(246): 110040.
- [4] Jawad Z, Villányi B, 2025, Designing Predictive Analytics Frameworks for Supply Chain Quality Management: A Machine Learning Approach to Defect Rate Optimization. *Platforms*, 3(2): 6.
- [5] Kadam A, Vaidya T, Katragadda S, 2025, Digital Transformation of Supply Chain Quality Management: Integrating

AI, IoT, Blockchain, and Big Data. *Journal of Economics, Finance and Accounting Studies*, 7(3): 41–49.

- [6] Ada N, Ethirajan M, Kumar A, et al., 2021, Blockchain Technology for Enhancing Traceability and Efficiency in Automobile Supply Chain: A Case Study. *Sustainability*, 13(24), 13667.
- [7] Gulnaz S, Jia F, Chen L, et al., 2024, AI Adoption in Supply Chain Management: A Systematic Literature Review. *Journal of Manufacturing Technology Management*, 35(6): 1125–1150.
- [8] Jahin M, Shahriar A, Amin M, 2025, MCDFN: Supply Chain Demand Forecasting via an Explainable Multi-Channel Data Fusion Network Model. *Evolutionary Intelligence*, 18(3): 66.
- [9] Schöpfer H, Kersten W, 2021, Using Natural Language Processing for Supply Chain Mapping: A Systematic Review of Current Approaches. *Computational Linguistics and Intelligent Systems (CoLLInS)*, 71–86.
- [10] Hübschke M, Buss E, Holschbach E, et al., 2025, Blockchain in Supply Chain Management: A Comprehensive Review of Success Measurement Methods. *Management Review Quarterly*, 2025.
- [11] Dritsas E, Trigka M, 2025, Methodological and Technological Advancements in E-Learning. *Information*, 16(1): 56.
- [12] Rolf B, Jackson I, Müller M, et al., 2023, A Review on Reinforcement Learning Algorithms and Applications in Supply Chain Management. *International Journal of Production Research*, 61(20): 7151–7179.
- [13] Zhang Y, He L, Zheng J, 2025, A Deep Reinforcement Learning-Based Dynamic Replenishment Approach for Multi-Echelon Inventory Considering Cost Optimization. *Electronics*, 14(1): 66.
- [14] Msakni M, Risan A, Schütz P, 2023, Using Machine Learning Prediction Models for Quality Control: A Case Study from the Automotive Industry. *Computational Management Science*, 20(1): 14.
- [15] Walter S, 2023, AI Impacts on Supply Chain Performance: A Manufacturing Use Case Study. *Discover Artificial Intelligence*, 3(1): 18.
- [16] Ojeda J, Moraes J, Filho C, et al., 2025, Application of a Predictive Model to Reduce Unplanned Downtime in Automotive Industry Production Processes: A Sustainability Perspective. *Sustainability*, 17(9): 3926.
- [17] Mangini C, Da Silva Lima N, de Alencar Nääs I, 2020, Prediction of Cold Chain Transport Conditions Using Data Mining. *IFIP International Conference on Advances in Production Management Systems (APMS)*, 616–623.
- [18] Bhatt J, Morris K, Haware R, 2024, Development of a Predictive Statistical Model for Gaining Valuable Insights in Pharmaceutical Product Recalls. *AAPS PharmSciTech*, 25(8): 255.
- [19] Mangini C, Lima N, Nääs I, 2024, Risk Prediction Score for Thermal Mapping of Pharmaceutical Transport Routes in Brazil. *Logistics*, 8(3): 84.
- [20] Cooper M, 2025, Barriers to AI Adoption in Supply Chain Management: Perspectives from Industry Leaders, Preprints, DOI:10.20944/preprints202504.0581.v1

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mechanism-Based Fault Diagnosis of Reciprocating Natural Gas Compressor Valves

Yixuan Yang*

Xi'an Shiyou University, Xi'an 710065, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The gas valve is the most vulnerable component of a reciprocating natural gas compressor, and its state is closely related to the operating state of the compressor. However, valve faults cannot be clearly diagnosed through the frequency domain. To address this difficulty, this paper analyzes the valve vibration signals from the perspective of time-domain mechanisms. The main research work includes analyzing various vibration sources of the gas valve through on-site vibration data and key phase signals, and analyzing fault signals based on the impact of vibration sources. The analysis of on-site fault vibration signals has well confirmed the accuracy of vibration source classification and the reliability of time-domain analysis methods. The results show that fault diagnosis of compressor valves through time-domain signals is a reliable approach.

Keywords: Reciprocating natural gas compressor; Gas valve; Mechanism; Fault diagnosis

Online publication: February 27, 2026

1. Introduction

With the continuous growth of global energy demand, natural gas has become one of the main clean energy sources and is widely used in various industrial equipment and transportation vehicles ^[1]. As a crucial equipment in the production, transportation, and storage of natural gas, large reciprocating natural gas compressors (hereinafter referred to as compressors) bear key operational loads during operation ^[2]. However, during long-term high-load operation, the gas valves of compressors are prone to performance degradation or failure due to factors such as wear and tear of components, malfunctions, or human operational errors ^[3]. In severe cases, this may even lead to equipment downtime, seriously affecting production efficiency and safety. Although gas valves are continuously being improved and developed, they remain the most vulnerable components in compressors, with a service life of less than three months in harsh working environments ^[4-6]. Therefore, research on fault diagnosis and health monitoring technologies for compressor gas valves holds significant theoretical value and application prospects.

Currently, fault diagnosis methods based on vibration signals primarily rely on the time and frequency

domains. While spectral analysis is effective for identifying severe faults such as valve plate breakage, its effectiveness in analyzing milder faults like air valve spring failure is not sufficiently evident. Therefore, this paper performs envelope processing on the air valve vibration signals collected from the valve cover, and the extracted envelope signals contain information about the mechanism and faults of the air valve ^[7].

2. Collection objects and systems

Data acquisition is primarily conducted for a large reciprocating compressor of a natural gas company. The rated power of this compressor is 5816 kW, and its gas valve adopts a mesh valve structure. The acquisition points for the gas valve are shown in **Figure 1**. Explosion-proof magnetic suction piezoelectric acceleration sensors are used, and infrared key-direction signal sensors are also employed to obtain the time point when the piston moves to the outer dead center. The sampling rate of vibration signals in this paper is set at 12800 Hz.



Figure 1. Location map of vibration signal acquisition points for the air valve.

3. Analysis of vibration signal excitation source

3.1. Schematic diagram of air valve structure

The simplified structure of the gas valve is shown in **Figure 2**, consisting of a lift limiter, spring, valve plate, and valve seat ^[8]. When the pressure difference between the internal and external air is insufficient to overcome the spring force and the gravitational force of the valve plate, the valve plate is pressed against the valve seat, preventing normal airflow. However, as the piston moves, when the pressure difference reaches a certain value, the gas pushes the valve plate to the state shown in the figure, and the gas valve starts to allow gas to flow. Therefore, with the reciprocating motion of the piston, the gas valve opens and closes regularly, allowing the compressor to

operate normally. However, due to the continuous impact of the gas valve on the lift limiter and valve seat, the valve plate is easily damaged ^[9]. Therefore, fault diagnosis for the gas valve is particularly important.

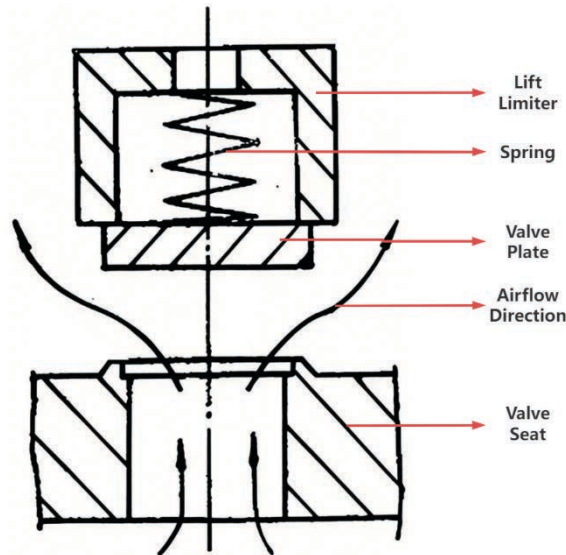


Figure 2. Simplified structural schematic diagram of the gas valve.

3.2. Analysis of vibration signal excitation source

Due to the numerous moving parts in the entire compressor unit, the vibration signal of the gas valve contains a lot of interference, and high-frequency vibrations make it difficult to intuitively discern the characteristics of the vibration time-domain signal. In this paper, the preprocessing of vibration signals employs the Hilbert envelope to strip away redundant information and highlight core features.

As shown in **Figure 3**, the vibration signals of all air valves on one side are plotted in a single graph to facilitate vibration source analysis. For a double-acting eight-valve compressor, the actions on both the left and right sides are completely symmetrical. Since the vibrations of the air valves are symmetrical in both positive and negative directions, the vibration data of each air valve is either positive or negative. The vibration data of intake valve 2 and exhaust valve 1 are taken as positive values, while the vibration data of intake valve 1 and exhaust valve 2 are taken as negative values. The reason for this is that during the movement of the piston from the outer dead center to the inner dead center, intake valve 2 and exhaust valve 1 undergo valve plate movement, so their vibration data are taken as positive values. Similarly, intake valve 1 and exhaust valve 2 are both taken as negative values. This way, for each stage of the air valve vibration signal, it is only necessary to focus on the shape of the y-axis side, facilitating the classification of vibration sources.

Phases 1 and 3 represent the movement of the piston from the outer dead center to the inner dead center. During this period, the intake valve 2 and exhaust valve 1 experience impact vibrations from the valve plates. It is evident that the signal intensity above the y-axis is high, while the signal below the y-axis is only passive vibration. Similarly, in phases 2 and 4, the piston moves from the inner dead center to the outer dead center, and the vibrations of the intake valve 1 and exhaust valve 2 are significantly stronger. Therefore, the impacts at points 1, 3, 5, and 7 in the figure are due to the collision vibrations between the intake valve plate and the lift limiter, while the impacts at points 2, 4, 6, and 8 are due to the collision vibrations between the exhaust valve plate and the lift limiter.

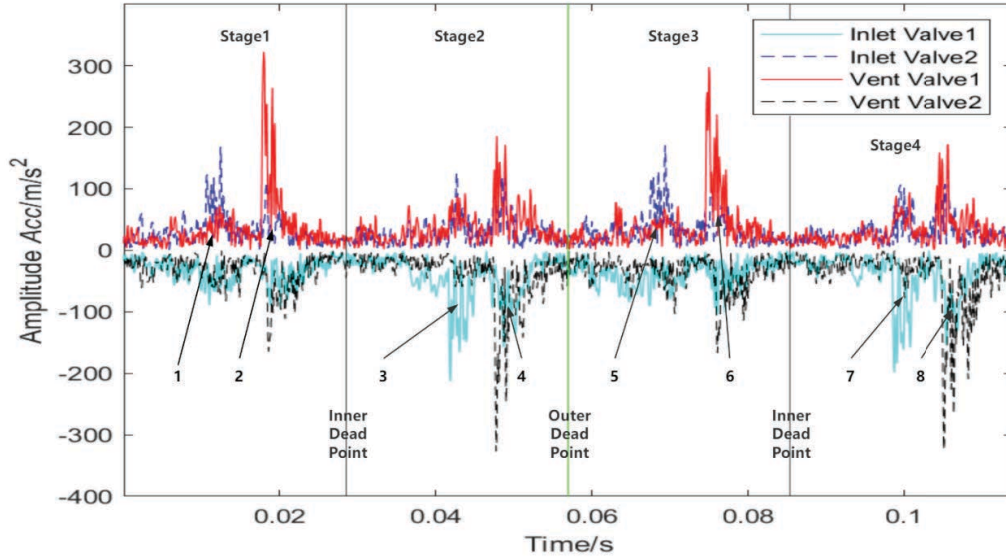


Figure 3. Envelope diagram of vibration signal of air valve on one side of the cylinder.

3.3. Analysis of valve plate movement

The movement of the valve plate primarily depends on the spring force and the pressure difference between the internal and external air pressures, while the rate of pressure change in the cylinder due to the gas inside mainly depends on the piston's movement speed^[10]. The formula is as follows:

$$\lambda = \frac{r}{l}$$

$$v = r\omega(\sin t\omega + \frac{\lambda}{2} \frac{\sin 2t\omega}{\sqrt{1 - \lambda^2 \sin^2 t\omega}})$$

Where r represents the crank radius, l denotes the length of the connecting rod, λ signifies the ratio of crank radius to connecting rod length, v stands for the piston's movement speed, ω indicates the angular velocity of the crankshaft rotation, and t represents time.

The piston speed, as shown in **Figure 4**, varies periodically, and the speed decreases as it moves closer to the stop point, resulting in a rate of change in cylinder pressure that first increases and then decreases from the stop point. Taking the intake valve 1 in **Figure 1** as an example, as the piston moves from the inner stop point to the outer stop point, the pressure difference between the inside and outside gradually overcomes the spring force, pushing the valve plate onto the lift limiter. As gas flows into the cylinder, due to the limited space of the intake chamber, the pressure inside the cylinder exhibits pulsating changes. However, the pressure difference still allows the valve plate to approach the lift limiter. However, as the cylinder pressure approaches the stop point, the rate of change slows down. Due to factors such as gas inertia, the valve plate may fall back onto the valve seat earlier, and its speed is much lower compared to when it impacts the lift limiter. Therefore, the impact is not evident in the on-site vibration waveform. However, due to factors such as spring stiffness and intake and exhaust pressure, the valve plate may bounce strongly when it impacts the lift limiter and valve seat, which may accelerate the damage to the valve plate and reduce its lifespan.

Although the operating principles of intake and exhaust valves are similar, there are still certain differences in their movements. The vibration of the intake valve plate hitting the lift limiter will precede the exhaust valve plate hitting the lift limiter, as shown in **Figure 3** where the impact at point 1 occurs earlier than the impact at point 2. This means that the piston displacement during the expansion process is less than that during the compression process. The theoretical proof is as follows.

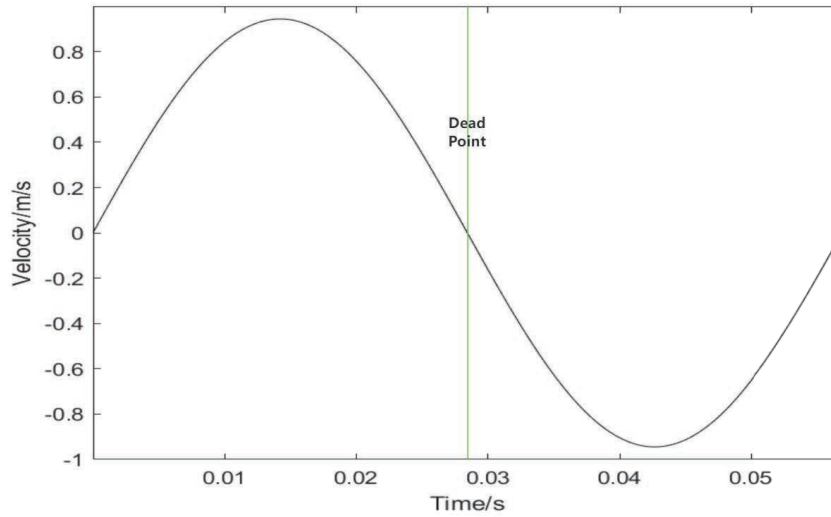


Figure 4. Schematic diagram of piston speed.

By simplifying a gas into an ideal gas, the gas satisfies:

$$pv^n = C$$

Where p represents the gas pressure, v denotes the gas volume, n stands for the polytropic index, and C signifies a constant. Generally, in engineering, it is assumed that the polytropic index is equal for both the expansion and compression processes.

Therefore, based on the data presented in **Figure 5**, we can deduce that:

$$p_1 v_4^n = p_2 v_3^n$$

$$p_1 v_2^n = p_2 v_1^n$$

That is:

$$\frac{v_3}{v_4} = \frac{v_1}{v_2}$$

$$1 - \frac{v_3}{v_4} = 1 - \frac{v_1}{v_2}$$

$$\frac{v_4 - v_3}{v_4} = \frac{v_2 - v_1}{v_2}$$

Because:

$$v_4 > v_2$$

So:

$$v_4 - v_3 > v_2 - v_1$$

Therefore, the piston displacement during the expansion process is smaller than that during the compression process, meaning the impact at point 1 in **Figure 3** will occur earlier than the impact at point 2.

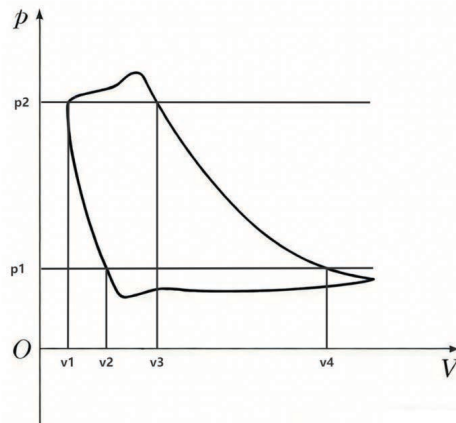


Figure 5. Indicator diagram of compressor.

4. Fault signal analysis based on mechanism

4.1. Fault signal manifestation

The gas valve is not only the most vulnerable component in a compressor, but it also exhibits various failure modes. Generally, there are four common types of failures: gas valve leakage, gas valve spring failure, valve plate fracture, and gas valve sticking. Due to the overlap in vibration characteristics among different failures, accurately inferring the type of failure is challenging. However, different failures can be roughly inferred based on the vibration characteristics of the time-domain vibration source.

When the gas valve leaks, during the stage when the valve should be fully closed and sealed, the waveform is not flat but exhibits continuous low-amplitude vibration, which is caused by gas leakage. Furthermore, due to the impact of leakage, the pressure change in the cylinder slows down, resulting in a decrease in the impact amplitude during the opening and closing of the gas valve.

When the valve spring fails, during the valve opening process, the valve plate will repeatedly impact the lift limiter, manifesting as dense, continuous multiple impact peaks; and due to insufficient spring force, the valve plate reseating speed slows down, resulting in a delayed collision impact with the valve seat.

When the valve plate fractures, the vibration of the gas valve not only exhibits the vibration pattern typical of gas valve leakage, but may also include chaotic impact spikes. The most severe situation is when valve plate fragments bounce chaotically inside the valve chamber. Valve plate fracture may also lead to valve plate sticking or seizure, resulting in the disappearance of impact.

When the air valve is stuck, the movement of the valve plate is obstructed, requiring a larger pressure difference between the inside and outside to move. This is manifested as a delay in the shock waveform on the time-domain diagram.

4.2. Case 1

This is the vibration waveform pattern formed by the fracture of the valve plate at the position of intake valve 1 in **Figure 1**. In **Figure 6**, there is only a weak impact vibration when the valve plate at position 1 is opened, indicating that when the piston moves from the inner dead center to the outer dead center, the poor sealing

caused by the fracture of the valve plate leads to air leakage, and the pressure change in the cylinder slows down. Therefore, the gas force phase experienced by the valve plate is much smaller than normal, resulting in a smaller impact. At position 2, an unreasonably high-amplitude impact is generated, indicating severe valve plate sticking, which is likely due to the presence of valve plate fragments. At position 3, where the intake valve should be closed, the vibration waveform exhibits low-amplitude continuous vibration, also indicating air leakage caused by the fracture of the valve plate. In summary, based on the vibration behavior of various vibration sources, it can be inferred that this waveform is generated by the fracture of the valve plate.

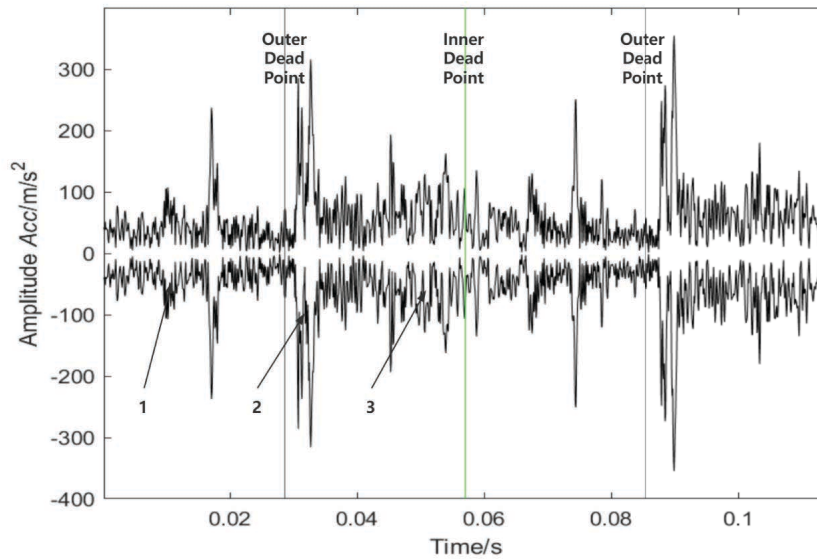


Figure 6. Vibration signal envelope diagram of Case 1.

4.3. Case 2

This is the vibration waveform formed by the fracture of the valve plate at the position of intake valve 2 in **Figure 1**. There is an unreasonably high-amplitude impact at position 1 in **Figure 7**, indicating a severe valve plate sticking phenomenon. It is likely that there are valve plate fragments, and the valve plate is forcibly pressed down onto the valve seat by the reverse airflow, resulting in a delayed strong impact. It is necessary to inspect and replace the valve plate in a timely manner.

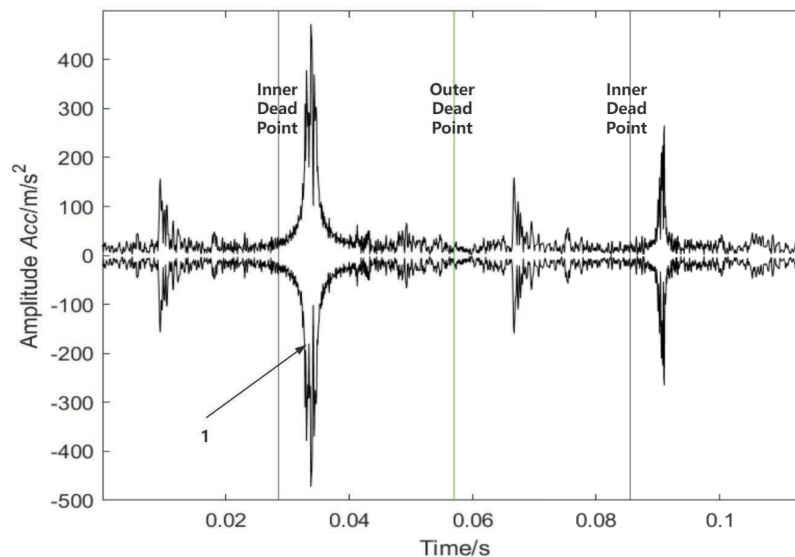


Figure 7. Vibration signal envelope diagram of Case 2.

5. Conclusion

In order to achieve reliable fault diagnosis for compressor valves, this paper proposes a mechanism-based diagnostic method that can determine the state of the valve through the vibration behavior of various vibration excitation sources. The effectiveness of the proposed method is verified through the inference of on-site fault signals. The results show that the method in this paper can achieve reliable fault diagnosis for valves. Subsequent research can explore how to use effective feature representation methods to classify various faults and utilize neural networks and other methods to achieve intelligent diagnosis of compressor valves.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Song R, Xu J, Chen B, 2024, Automatic Control Method for Vibration Faults of Reciprocating Natural Gas Compressors. *Journal of Machinery and Electronics*, 42(1): 71–75.
- [2] Yang C, Zhang D, Teng Y, 2021, Analysis of Monitoring and Fault Diagnosis Technology for Reciprocating Compressors in Shale Gas Fields. *China Equipment Engineering*, 2021(S1): 50–53.
- [3] Chang W, Mi S, Zhang Z, 2024, Analysis and Treatment of Common Faults in Marine Reciprocating Natural Gas Compressors. *Standardization and Quality of Machinery Industry*, 2024(6): 32–35.
- [4] Yu Y, 2005, *Technical Manual of Positive Displacement Compressors*, China Machine Press, Beijing, 328–330.
- [5] Ma H, Yu X, Lv Q, et al., 2022, Application of One-Dimensional Convolutional Neural Network in Fault Diagnosis of Reciprocating Compressor Valves. *Journal of Xi'an Jiaotong University*, 56(4): 101–108.
- [6] Hu F, Xu H, Tang A, 2013, Fault Diagnosis Analysis and Experimental Research on Marine Air Compressors. *Journal of Ship Engineering*, 35(5): 40–43.
- [7] Liu H, Lin J, Shen W, et al., 1996, Vibration Diagnosis Method for Reciprocating Compressor Valve Faults. *Compressor Technology*, 1996(1): 32–34.
- [8] Wu Y, 1989, *Mathematical Model and Application of Reciprocating Compressor*, Xi'an Jiaotong University Press, 60–66.
- [9] Wu G, 2008, *Research on Dynamic Characteristics and Fault Simulation of Reciprocating Compressor Valves*, thesis, China University of Petroleum.
- [10] Qu Z, Yang X, 2024, *Principles of Reciprocating Compressors*, Xi'an Jiaotong University Press.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Application of the SCSSA-VMD Denoising Method in Natural Gas Pipeline Leakage Detection

Tianxiang Xie, Dan Zhang

Xi'an Shiyou University, Xi'an, Shaanxi, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The decomposition performance of variational mode decomposition (VMD) on natural gas pipeline leakage pressure signals is highly sensitive to the subjective selection of its key parameters: the number of modes K and the penalty factor α . To address this issue, this paper proposes an enhanced sparrow search algorithm (SSA) that integrates sine/cosine searching and Cauchy mutation strategies, referred to as SCSSA, for optimizing the VMD parameter combination. Experimental results demonstrate that the SCSSA-optimized VMD method significantly outperforms denoising approaches based on the standard SSA and particle swarm optimization (PSO) in optimizing VMD parameters. Specifically, the proposed method achieves a higher signal-to-noise ratio (SNR) and a lower root mean square error (RMSE) in the denoised signal, effectively enhancing the denoising performance.

Keywords: Natural gas pipeline; Leakage detection; Sparrow search algorithm (SSA); Variational mode decomposition (VMD); Signal denoising

Online publication: February 12, 2026

1. Introduction

The detection of pipeline leakage signals serves as a crucial element in ensuring the safety and integrity of oil and gas transportation systems. These leakage signals generally display highly non-stationary and nonlinear behavior, making them inherently complex and difficult to analyze. Moreover, they are frequently contaminated by multiple forms of noise arising from diverse sources. Key contributors include internal fluid dynamic phenomena, such as turbulence induced by high-velocity flows and irregular pressure fluctuations, as well as external environmental disturbances, for instance, mechanical vibrations from pumping equipment or electromagnetic interference from nearby industrial activities. Additionally, sensor-related inaccuracies and instrumental limitations further contribute to the overall noise profile. The convergence and superposition of such noise components often mask the intrinsically weak and transient characteristics of leakage signals, which in turn substantially undermines the reliability of subsequent feature extraction processes and degrades the performance of leakage detection

algorithms. As a result, the development of efficient, robust, and adaptive denoising techniques is an essential prerequisite for enhancing the discernibility of leakage signatures and improving the accuracy of detection models when handling complex real-world pipeline data.

In this context, variational mode decomposition (VMD) has emerged as a prominent technique for processing such signals since its introduction ^[1]. VMD is a fully non-recursive, adaptive signal decomposition method. Its core principle involves constructing and solving a constrained variational optimization problem to adaptively decompose the original signal into a set of sub-signal components with specific center frequencies and limited bandwidth, known as intrinsic mode functions (IMFs). Compared to methods like empirical mode decomposition (EMD), VMD is grounded in a rigorous mathematical framework, which effectively suppresses mode mixing and offers excellent time-frequency localization properties. This makes it particularly suitable for extracting transient components of non-stationary signals from an intensely noisy background ^[2,3]. In the processing of pipeline leakage pressure signals, VMD can separate the weak pressure fluctuations induced by a leak from the complex noise background, providing a high-quality data foundation for subsequent fault diagnosis ^[4].

However, the performance of VMD is highly contingent upon the preset values of its key parameters, primarily the number of decomposition modes, K and the penalty factor (or bandwidth constraint parameter), α . An inappropriately small K value leads to insufficient decomposition, causing different signal components to alias within the same mode. Conversely, an excessively large K value results in redundant and spurious modes, leading to over-decomposition ^[5]. The parameter α controls the estimated bandwidth of each IMF, influencing the smoothness of the mode's center frequency and bandwidth. Suboptimal parameter configuration can significantly degrade the decomposition efficacy of VMD, thereby impacting the accuracy of denoising and feature extraction ^[6].

To address this limitation, researchers have proposed multiple strategies for VMD parameter optimization and hybrid denoising. For instance, Xiao *et al.* proposed an improved VMD method combined with a threshold algorithm for partial discharge signal denoising. By optimizing the parameters, they enhanced the decomposition specificity and effectively eliminated noise-dominant modes through subsequent thresholding ^[7]. Wu *et al.*, focusing on surface electromyography (sEMG) signal noise, introduced an improved sparrow search algorithm (ISSA) to adaptively optimize VMD parameters. This was combined with a second-generation wavelet threshold applied to the IMFs. This approach enhanced the parameter self-adaptation capability of VMD and achieved favorable denoising results ^[8]. These research ideas also provide valuable references for pipeline leakage signal processing. Furthermore, some scholars have employed fitness functions such as sample entropy and envelope entropy, combined with swarm intelligence optimization algorithms (e.g., particle swarm optimization, genetic algorithm) to automatically determine the optimal parameter combination for VMD, further improving its performance in specific applications ^[9,10].

In summary, when confronting the challenges posed by the non-stationary, nonlinear, and noise-contaminated nature of pipeline leakage signals, VMD demonstrates significant application potential due to its solid theoretical foundation and adaptive decomposition capability. Integrating VMD with other optimization algorithms and signal processing techniques (e.g., threshold denoising, wavelet transform) to construct intelligent, parameter-optimized VMD hybrid denoising models represents a crucial research direction for improving the accuracy and reliability of pipeline leakage detection.

Despite extensive research on pipeline signal denoising, existing methods still have limitations. This study aims to develop a more efficient and robust leakage signal denoising algorithm by leveraging an optimized VMD approach, thereby improving detection accuracy and reliability, and providing enhanced technical support for the

safe operation of natural gas pipelines.

2. Theoretical background and methodology

2.1. Principle of variational mode decomposition

VMD, introduced in 2014, is a fully non-recursive signal processing technique. Its core principle involves iteratively searching for the optimal solution of a variational model to adaptively decompose a complex signal into a discrete number of mode components a discrete set of band-limited (IMFs) with specific sparsity properties and central frequencies. The VMD method effectively overcomes endpoint effects and mode mixing problems common in other decomposition techniques. For a detailed mathematical formulation, readers are referred to the seminal work ^[1].

2.2. Enhanced sparrow search algorithm

The standard sparrow search algorithm (SSA) mimics the foraging behavior and anti-predation strategies of sparrows. While SSA benefits from a simple structure, few parameters, and rapid convergence, it often suffers from diminished population diversity in the later stages of optimization, leading to a high probability of converging to local optima ^[11].

To overcome these limitations, a multitude of enhanced variants of the salp swarm algorithm (SSA) have been put forward in recent years. These modifications generally concentrate on augmenting population diversity via mechanisms like adaptive parameter adjustment, hybrid operators combined with other meta-heuristic algorithms, or the incorporation of local search strategies. For example, certain studies have integrated chaotic mapping into the initialization stage to generate more uniformly distributed initial solutions. In contrast, others have incorporated mutation operators inspired by genetic algorithms to perturb stagnant populations and evade local optima. Moreover, adaptive weight strategies that dynamically balance exploration and exploitation capabilities during the optimization procedure have demonstrated potential in enhancing the algorithm's global search performance.

Notwithstanding these advancements, numerous existing improved SSA versions still encounter challenges when dealing with complex high-dimensional optimization problems or maintaining stable convergence rates across various types of objective functions. Consequently, further research is necessary to develop more robust and versatile SSA variants that can effectively address the diverse optimization requirements of real-world applications, such as engineering design, machine learning parameter tuning, and resource allocation problems.

To mitigate these limitations, this paper employs an enhanced SSA (SCSSA) that incorporates a sine/cosine search strategy and a Cauchy mutation operator, as proposed ^[12]. The SCSSA procedure is outlined as follows:

(1) Step 1: Initialization: Initialize the sparrow population size N , maximum iterations $Iter_{max}$, discoverer proportion P_D , vigilance proportion P_S , vigilance threshold R_2 , safety value ST . Randomly generate the initial positions of the sparrows as follows:

$$X_{i,j}(i = 1, 2, \dots, N; j = 1, 2, \dots, D, D \text{ is the problem dimension})$$

(2) Step 2: Fitness evaluation: Calculate the fitness value for each sparrow based on the objective function. Identify the best fitness f_g (best position X_{best}) and the worst fitness f_w (worst position X_{worst}) in the current population;

(3) Step 3: Discoverer update: The discoverers (the best $P_D \times N$ Nsparrows) update their positions using a strategy incorporating a nonlinear weight and a sine/cosine mechanism:

If $R_2 < ST$ (no threat from natural predators):

$$X_{ij}^{t+1} = \omega \cdot X_{ij}^t + r_1' \cdot \sin r_2 \cdot |r_3 \cdot X_{\text{best}} - X_{ij}^t| \quad (1)$$

where ω is a nonlinear decreasing weight factor $r_1' = 1 \times \left(1 - \left(\frac{t}{Iter_{\max}}\right)^\eta\right)^{\frac{1}{\eta}}$ ($\eta = 1.5$), $r_2, r_3 \in [0, 2\pi]$ are random numbers.

If $R_2 \geq ST$ (existence of predator threat), the r_2 in **Eq. (1)** is replaced by the r_2 .

(4) Step 4: Follower update: The remaining sparrows (followers) update their positions using a Cauchy mutation strategy to enhance global search ability and escape local optima:

$$X_{ij}^{t+1} = X_{\text{best}}^t + \text{Cauchy}(0,1) \otimes X_{\text{best}}^t \quad (2)$$

where $\text{Cauchy}(0,1)$ is a random number from the standard Cauchy distribution, and \otimes denotes element-wise multiplication.

(5) Step 5: Vigilantes update: For vigilantes (sparrows with fitness values at the tail of the population $P_S \times N$), position updates are performed under different scenarios:

If $f_i > f_g$ (sparrow is at the edge of the group):

$$X_{ij}^{t+1} = X_{\text{best}}^t + \beta \cdot |X_{ij}^t - X_{\text{best}}^t| \quad (3)$$

where β is a step size control coefficient following a normal distribution.

If $f_i = f_g$ (sparrow is in the center of the group):

$$X_{ij}^{t+1} = X_{ij}^t + k \cdot \frac{|X_{ij}^t - X_{\text{worst}}^t|}{(f_i - f_w) + \varepsilon} \quad (4)$$

where $k \in (0,1)$, $\varepsilon = 10^{-50}$, is the minimum value to avoid a denominator of 0.

(6) Step 6: Population update: Merge the updated positions of discoverers, followers, and vigilantes. Recalculate the fitness of all sparrows and select the best N sparrows to form the new population;

(7) Step 7: Termination check: If the maximum iteration count $Iter_{\max}$ is reached, output the optimal solution; otherwise, return to Step 2.

2.3. SCSSA-optimized VMD parameter selection

Selecting an appropriate fitness function is critical for guiding the SCSSA towards optimal VMD parameters (K , α). Pipeline leakage signals are typically nonlinear and non-stationary. Envelope entropy is highly sensitive to impulse components within such signals, making it a suitable candidate ^[13]. However, relying solely on envelope entropy can lead to IMFs with high purity but potential frequency aliasing. To address this, an orthogonal index is incorporated into the fitness function penalize mode mixing, effectively avoiding frequency overlapping problem and ensuring the rationality of decomposition structure.

The orthogonal index quantifies the degree of orthogonality between different IMFs, with higher values indicating less mode mixing. By combining envelope entropy and the orthogonal index, the fitness function can simultaneously minimize the complexity of the decomposed signals and maximize the independence between IMFs. This hybrid fitness function not only enhances the sensitivity to the impulse characteristics of pipeline leakage signals but also ensures the structural integrity of the decomposition results, laying a solid foundation for the subsequent accurate extraction of leakage features. In practical applications, the weights assigned to envelope entropy and the orthogonal index in the fitness function can be adjusted according to the specific characteristics

of the pipeline and the environmental noise level, allowing the SCSSA to adaptively search for the optimal VMD parameters under different working conditions.

3. Experimental setup and analysis

3.1. Experimental data and parameters

The experimental pipeline was constructed from seamless steel with an outer diameter of 108 mm, a wall thickness of 4.5 mm, and a total length of approximately 100 meters. The operating pressure ranged from 0.5 MPa to 2.0 MPa. A leakage orifice with a diameter of 1 mm was simulated ^[14]. The leakage pressure signals used in this study are derived from the publicly available dataset provided by. The parameter search spaces for VMD were set as $K \in [2, 10]$ and $\alpha \in [500, 5000]$. The SCSSA population size was set to 30, with a maximum of 50 iterations.

3.2. Signal decomposition and denoising process

The SCSSA (Slime Cossinidae Swarm Algorithm) was utilized to determine the optimal parameter combination for the VMD process, which was identified as $K = 8$ and $\alpha = 1496$. By applying VMD with these specifically optimized parameters to the noisy pipeline leakage signal, the method adaptively and effectively decomposed the original signal into eight distinct Band-Limited Intrinsic Mode Functions (BLIMFs). As clearly illustrated in **Figure 1**, each of these eight BLIMF components captures and represents information from different and specific frequency ranges present within the complex noisy signal, thereby facilitating a more detailed and structured analysis of the underlying data characteristics.

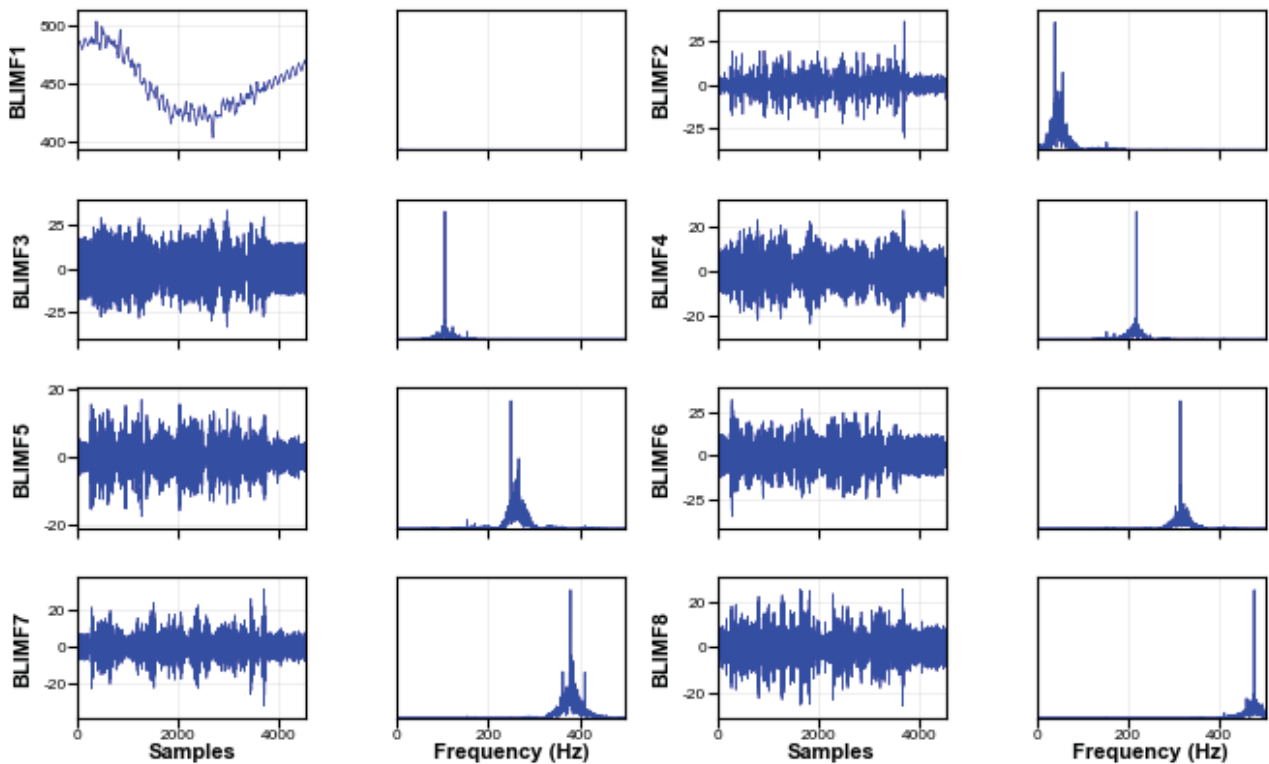


Figure 1. Time-frequency representations of BLIMF component.

Subsequent filtering of these BLIMFs was performed based on their correlation with the original signal, energy distribution, and frequency characteristics. BLIMFs strongly correlated with leakage features (e.g., representing pressure transient events) and possessing significant energy were retained. BLIMFs dominated by high-frequency noise were discarded. The denoised signal was reconstructed by summing the selected relevant BLIMFs. The comparison between the original signal and the denoised signal is shown in **Figure 2**. The visual comparison between the original signal and the denoised signal further confirms the effectiveness of the proposed method, showing a smoother waveform that retains the essential leakage-induced pressure variations.

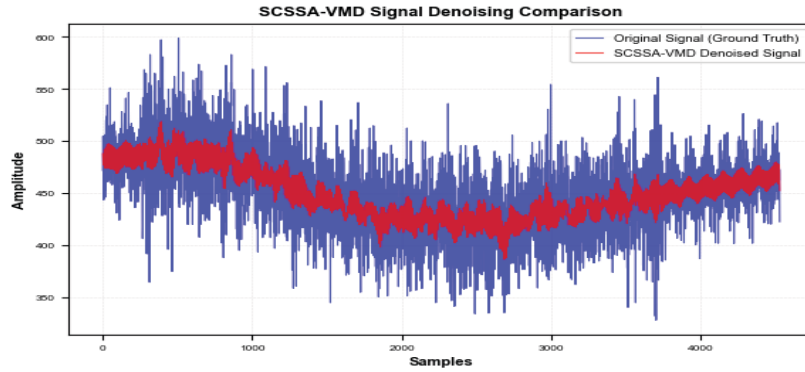


Figure 2. Comparison of original and denoised signals.

3.3. Results and comparative analysis

The denoising performance was quantitatively evaluated using signal-to-noise ratio (SNR) and root mean square error (RMSE). Higher SNR and lower RMSE values indicate superior denoising performance. The proposed SCSSA-VMD method was compared against two established optimization algorithms for VMD parameter tuning: the standard SSA-VMD and PSO-VMD.

As shown in **Table 1**, the SCSSA-VMD method achieved the highest SNR value of 24.490 dB and the lowest RMSE of 0.0996 among the three algorithms. In contrast, the SSA-VMD method obtained an SNR of 24.1588 dB and an RMSE of 0.1035, while the PSO-VMD method had the lowest SNR of 23.914 dB and the highest RMSE of 0.1065. These results clearly demonstrate that the proposed SCSSA-VMD method outperforms both the standard SSA-VMD and PSO-VMD in terms of denoising performance. The superior performance of SCSSA-VMD can be attributed to its enhanced ability to optimize the VMD parameters K and α . Specifically, the SCSSA-VMD method selected a K value of 8 and an α value of 1496, which are different from the values chosen by the other two algorithms. The higher K value might allow the method to better decompose the signal into more intrinsic mode functions, capturing more detailed information and thus improving the denoising effect. Meanwhile, the optimized α value could help in balancing the trade-off between the sparsity and smoothness of the decomposed components, leading to a more effective noise reduction. The quantitative comparison through SNR and RMSE metrics provides solid evidence that the SCSSA-VMD method is a more effective approach for VMD parameter tuning in denoising applications.

Table 1. Performance comparison of different denoising algorithms

Optimization algorithm	K	α	SNR (dB)	RMSE
SCSSA-VMD	8	1496	24.490	0.0996
SSA-VMD	3	3102	24.1588	0.1035
PSO-VMD	3	3652	23.914	0.1065

The SCSSA-VMD algorithm achieved the most optimal parameter combination, resulting in the highest SNR and the lowest RMSE. This indicates its superior ability to effectively separate noise from the underlying leakage signal while preserving critical diagnostic information.

4. Conclusion

This study successfully addressed the sensitivity of VMD to subjective parameter selection in pipeline leakage signal denoising by introducing an enhanced SCSSA for automatic parameter optimization. The SCSSA incorporates sine/cosine searching and Cauchy mutation to improve global exploration and avoid local optima. Experimental results demonstrate that the SCSSA-VMD method significantly outperforms traditional optimization approaches like SSA and PSO, achieving superior denoising performance quantified by higher SNR and lower RMSE. The proposed method effectively enhances the accuracy of pipeline leakage detection, contributing to improved safety and reliability in natural gas pipeline operations. Future work will focus on validating the method across a wider range of leakage scenarios and pipeline conditions.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Dragomiretskiy K, Zosso D, 2014, Variational Mode Decomposition. *IEEE Transactions on Signal Processing*, 62(3): 531–544.
- [2] Xue Y, Cao J, Yao Y, 2016, Application of the Variational-Mode Decomposition for Seismic Time-frequency Analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8): 3821–3831.
- [3] Zhou X, Wang X, Wang H, et al., 2023, Method for Denoising the Vibration Signal of Rotating Machinery through VMD and MODWPT. *Sensors (Basel, Switzerland)*, 23(15): 6904.
- [4] Xu T, Zeng Z, Huang X, et al., 2021, Pipeline Leak Detection Based on Variational Mode Decomposition and Support Vector Machine Using an Interior Spherical Detector. *Process Safety and Environmental Protection*, 2021(153): 167–177.
- [5] Jia G, Yang J, Liang H, 2025, A Combined Denoising Method of Adaptive VMD and Wavelet Threshold for Gear Health Monitoring. *Structural Durability & Health Monitoring*, 19(4): 1057–1072.
- [6] Li C, Liu Y, Liao Y, 2021, An Improved Parameter-Adaptive Variational Mode Decomposition Method and Its Application in Fault Diagnosis of Rolling Bearings. *Shock and Vibration*, 2021(1): 1–26.
- [7] Xiao S, Chen B, Shen D, et al., 2021, Improvement of VMD and Threshold Algorithms in Local Discharge

Denoising. *Journal of Electronic Measurement and Instrumentation*, 35(11): 206–214.

- [8] Wu T, Cai H, Liang J, et al., 2023, Noise Reduction of sEMG Signals Based on ISSA-VMD and Second-Generation Wavelet. *Electronic Measurement Technology*, 46(2): 93–100.
- [9] Zhou F, Yang X, Shen J, et al., 2020, Fault Diagnosis of Hydraulic Pumps Using PSO-VMD and Refined Composite Multiscale Fluctuation Dispersion Entropy. *Shock and Vibration*, 2020: 1–13.
- [10] Li J, Chen W, Han K, et al., 2020, Fault Diagnosis of Rolling Bearing Based on GA-VMD and Improved WOA-LSSVM. *IEEE ACCESS*, 2020(8): 166753–166767.
- [11] Xue J, Shen B, 2020, A Novel Swarm Intelligence Optimization Approach: Sparrow Search Algorithm. *Systems Science & Control Engineering*, 8(1): 22–34.
- [12] Li A, Quan L, Cui G, et al., 2022, Sparrow Search Algorithm Integrating Cosine and Sine with Cauchy Variation. *Computer Engineering and Applications*, 58(3): 91–99.
- [13] Zou N, Hu H, Bai Y, et al., 2021, A Joint Denoising Method Based on Optimized VMD and Wavelet Packet Denoising. *Mathematics in Practice and Theory*, 51(20): 135–142.
- [14] Yang T, 2024, Natural Gas Pipeline Leakage Data, *IEEE DataPort*, <https://dx.doi.org/10.21227/t7te-q548>

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A Framework for Advancing Intelligent Electrical Agricultural Machinery Technologies in Mountainous Regions

Hong Huang*

Chongqing Municipal Agricultural and Rural Affairs Committee, Chongqing 401121, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The complex terrain of the hilly and mountainous regions in southwestern China presents significant challenges to agricultural mechanization, resulting in a level that is markedly lower than the national average. Focusing on the key development needs for intelligent agricultural machinery in these areas. This paper systematically delineates four core technological domains: lightweight machine design, detachment and drag reduction in heavy clay soils, slope adaptability, and remote operation and maintenance. The study aims to provide technical insights for overcoming the bottlenecks in mechanizing hilly and mountainous agriculture, thereby contributing to rural revitalization and national agricultural development strategies.

Keywords: Electrical agricultural machinery; Mountainous regions; Design

Online publication: February 12, 2026

1. Introduction

Southwestern China, characterized by its predominantly hilly and mountainous terrain and abundant agricultural resources, serves as a strategic national hinterland and a vital production base for grains, oils, fruits, vegetables, tea, and sugar. This region is thus crucial for ensuring the security of agricultural product supply. Data indicates that cultivated land and crop-sown areas in the southwestern hills and mountains account for one-third of the national total, involving nearly 300 million agricultural workers. Against a backdrop of complex international dynamics, stabilizing agricultural production is of strategic importance for safeguarding national food security and advancing rural revitalization ^[1].

However, the region's complex terrain, steep slopes, and heavy clay soils pose formidable challenges. Large and medium-sized machinery common in plains struggle with field access, operation, and maneuverability. The limited availability of small machinery often necessitates manual labor, leading to high labor intensity and low efficiency, which in turn constrains agricultural mechanization and intelligence development ^[2]. While the national

comprehensive mechanization rate for crop cultivation, planting, and harvesting has reached 73%, it remains at only 55% in hilly and mountainous areas. Taking potato cultivation as an example, with a national planting area of approximately 4.67 million hectares, the southwestern mountainous regions, as the main production area, account for 40.6% of this total. Yet, their comprehensive mechanization rate is below 30%, with mechanized planting and harvesting rates under 10%. Furthermore, domestically produced tractors specialized for small plots and steep slopes are lacking. Large-scale rice transplanters and combine harvesters are prone to bogging down in wet fields, and there is a scarcity of lightweight machinery for transplanting and harvesting root crops (e.g., sweet potatoes) in heavy clay soil ^[3,4].

Consequently, there is an urgent need for targeted technological breakthroughs to address the critical bottleneck of having “no suitable machinery available, nor high-quality machinery accessible” for grain, oil, fruit, and vegetable production in these regions. Such advancements would promote modern agricultural development, foster rural economic prosperity and increased farmer income, and support national strategies for rural revitalization and building an agricultural powerhouse. This paper systematically elaborates on four key technology areas for intelligent agricultural machinery suited to the southwestern mountainous regions: lightweight design, detachment and drag reduction in clay soil, slope adaptability, and remote operation and maintenance. The goal is to guide future research, development, and application improvements.

2. Lightweighting technologies for mountainous intelligent agricultural machinery

Reducing the overall weight of machinery enhances its adaptability to small plots and steep slopes. This can be achieved through structural, material, and process lightweighting strategies ^[5].

2.1. Structural lightweighting technology

Structural lightweighting involves establishing optimization constraints through finite element-based continuum topology and discrete structural topology optimization algorithms. This process enables the identification of optimal lightweight structural solutions that meet performance requirements for body strength, modal characteristics, noise, vibration, and harshness (NVH), and crashworthiness, thereby achieving overall weight reduction ^[6].

2.2. Material lightweighting technology

Material lightweighting strategies include replacing ordinary steel with high-strength steel to reduce plate thickness while maintaining performance. For specific structural components, substituting with aluminum-magnesium alloys enables lightweight design that meets modal, stiffness, and reliability targets ^[7]. Research on optimizing dimensional parameters for plastics and composites can also expand the use of lightweight non-metallic materials, further reducing component weight.

2.3. Process lightweighting technology

Process lightweighting explores techniques like multi-material laser welding to reduce overall weight. Developing component manufacturing technologies based on hot stamping and part redesign effectively addresses issues with high-strength steel, such as limited deformation range, cracking susceptibility, and high spring back at room temperature ^[8]. Designing lightweight components (e.g., frames, control arms, steering rods) suitable for hydroforming technology is also highly effective for reducing individual part mass.

3. Detachment and drag reduction technology for heavy clay soils

Research on key technologies, including adhesion reduction and detachment for soil-engaging components, drag and energy consumption reduction, reliability enhancement, and structural adaptability design, is essential to improve machinery performance in the heavy clay soils prevalent in mountainous regions .

3.1. Adhesion reduction and detachment technology for soil-engaging components

Drawing on bionic principles, soil-engaging components can be designed with special structures and textures inspired by the unique surfaces of dung beetles, earthworms, and burrowing animals to minimize soil adhesion ^[11,12]. Surface modification techniques like carburizing and coating can further reduce adhesion. Additionally, incorporating detachment mechanisms such as mechanical scrapers or vibration assistance enables active adhesion reduction.

3.2. Drag and energy consumption reduction technology

Inspired by natural drag-reduction mechanisms (e.g., the skin structures of earthworms and pangolins), machinery surfaces can be designed with bionic characteristics to decrease contact area and friction coefficients with soil ^[13]. Designing appropriate vibration generation mechanisms to induce specific frequencies and amplitudes during operation can also reduce soil adhesion and friction forces on the equipment.

3.3. Reliability enhancement technology for soil-engaging components

Enhancing reliability involves developing impact- and wear-resistant materials and processing technologies based on a tribological analysis of the “environment-component-material” system ^[14]. Utilizing high-strength, wear-resistant materials ensures stable operation under complex loads. To meet the high toughness demands of mountainous environments, new materials can be developed using tough alloys and fracture toughness optimization, with heat treatment processes employed to refine microstructure and improve toughness and impact resistance ^[15].

4. Slope adaptability technology

Addressing the slope adaptability needs of both the machinery chassis and its implements through self-leveling technologies is crucial for improving operational safety and quality on inclines ^[16].

4.1. Implement slope adaptability technology

To mitigate issues like implement sinkage and poor ground contact, ground profiling mechanisms can be optimized through design. Simulation software aids in analyzing and optimizing key components to obtain optimal profiling parameters ^[16]. Research into depth detection methods for working components can reduce adjustment errors, enabling optimal parameter matching across different tillage conditions and thereby improving efficiency and quality.

4.2. Chassis self-leveling technology

To counter the poor stability and overturning risks of lightweight machinery in complex terrain, a three-point active leveling method based on position error can be employed. This approach addresses issues like leveling interruption due to limited actuator (hydraulic/pneumatic cylinder) stroke and helps manage the machine’s center

of gravity, enhancing anti-overturning capability and operational safety ^[17].

5. Remote operation and maintenance technology

The complex working conditions in southwestern mountains lead to diverse and frequent machinery faults. Studying fault diagnosis and prediction methods and developing remote O&M tools are vital for ensuring efficient operation.

5.1. Fault diagnosis and prediction technology

Sensor data combined with AI algorithms enable real-time monitoring and fault diagnosis, facilitating early detection of potential issues. Building on this, fault prediction models can forecast fault timing and severity, allowing for preventive measures to avoid major failures and downtime ^[18]. Fault simulation using digital twins and physical models can replicate scenarios to refine diagnostic and predictive algorithms, improving their accuracy and reliability.

5.2. Remote operation and management technology

IoT platforms can connect machinery to remote O&M centers for remote diagnosis, online updates, and maintenance guidance ^[19,20]. AR/VR technologies can provide remote assistance and virtual training, enhancing maintenance efficiency and safety. Mobile apps and cloud platforms enable real-time location tracking, status monitoring, and remote control, streamlining remote management.

5.3. Decision-making and optimization technology

AI and operations research algorithms can establish O&M decision models to automatically generate maintenance suggestions and repair schedules. Big data analytics and machine learning can optimize resource allocation and maintenance workflows, improve efficiency and reduce costs ^[21]. Multi-objective optimization techniques can balance factors like cost, reliability, availability, and safety to generate optimal O&M strategies.

6. Conclusion

This paper details key technologies for developing intelligent agricultural machinery in mountainous areas, encompassing lightweight design, detachment and drag reduction in clay soil, slope adaptability, and remote O&M. Lightweighting, achieved through integrated structural, material, and process optimization, enhances adaptability to small, sloped plots. For clay soils, combining bionic design, surface modification, and vibration techniques boosts operational efficiency. For slopes, chassis self-leveling and implement profiling optimization improve safety and quality stability. For O&M, a remote intelligent system based on IoT, digital twins, and AI enables fault prediction and optimized decision-making. These advancements provide a foundational framework for overcoming mechanization barriers and promoting sustainable agricultural development in challenging terrains.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Chen J, Chen X, Xie S, et al., 2025, Full-Coverage Path Planning for Multi-Machinery Anti-Subsidence in Partially Sinkage-Risky Fields Based on an Improved Seagull Optimization Algorithm with Simulated Annealing. *Computers and Electronics in Agriculture*, 2025(238): 110835.
- [2] Han X, Kim H, Jeon C, et al., 2021, Design and Field Testing of a Polygonal Paddy Infield Path Planner for Unmanned Tillage Operations. *Computers and Electronics in Agriculture*, 2021(191): 106567.
- [3] He L, Li Y, An X, et al., 2025, Real-Time Monitoring System for Evaluating the Operational Quality of Rice Transplanters. *Computers and Electronics in Agriculture*, 2025(234): 110204.
- [4] Issa I, Zhang Z, El-Kolaly W, et al., 2025, Innovated Design, Simulation and Evaluation of Potato Harvester Excavation and Separation Conveyors. *International Journal of Agricultural and Biological Engineering*, 18(2): 132–145.
- [5] Cai S, Luo Y, Li Q, 2021, State of the Art of Lightweight Technology in Agricultural Machinery and its Development Trend. *Journal of Mechanical Engineering*, 57(17): 35–52.
- [6] Fartash N, Selvi K, Ungureanu N, 2025, Exploring the Role of Advanced Composites and Biocomposites in Agricultural Machinery and Equipment: Insights into Design, Performance, and Sustainability. *Polymers*, 17(12): 1691.
- [7] Kong W, 2025, Researches on Lightweight Body Structure Design and Optimization of Electric Vehicles. *Applied and Computational Engineering*, 124(1): 183–189.
- [8] Soleimani A, 2024, The Soft Computing Approaches in Optimising Multi-Objective Mechanical Design of a Weeding Robot. *Smart Agricultural Technology*, 2024(9): 100674.
- [9] Zhang F, Luo Z, Zheng E, et al., 2025, Imitating Pangolin Scale Structure for Reducing Adhesion and Resistance of Rotary Tillage in Wet-Adhesive Soil. *Soil and Tillage Research*, 2025(245): 106306.
- [10] Xu Z, Qi H, Gao P, et al., 2024, Biomimetic Design of Soil-Engaging Components: A Review. *Biomimetics*, 9(6): 358.
- [11] Li J, Li X, Hu B, et al., 2023, Analysis of the Resistance Reduction Mechanism of Potato Bionic Digging Shovels in Clay and Heavy Soil Conditions. *Computers and Electronics in Agriculture*, 2023(214): 108315.
- [12] Li X, Wang Z, Gu T, et al., 2025, Design of Bionic Rotary Tiller Blades Based on Marmot Claws and Research on Anti-Entanglement and Drag Reduction Performance. *Smart Agricultural Technology*, 2025(12): 101586.
- [13] Zou Z, Li J, Wang X, et al., 2025, Wear Resistance Enhancement of Rotary Tillage Blades Through Structural Optimization and Surface Strengthening. *Materials*, 18(21): 5006.
- [14] Vrublevskyi O, Olejniczak K, Napiorkowski J, et al., 2025, Modeling Wear of Different Surface Layers of Agricultural Tools in Soil Abrasive Mass. *Measurement*, 2025(242): 116076.
- [15] Zeng J, Ma Y, Fang Z, et al., 2024, Process Optimization and Wear Performance of Plasma-Cladding Fe5 Coatings on Rotary Tillage Blades. *Applied Sciences*, 15(1): 77.
- [16] Khoroshko L, Kuznetsov P, 2020, Digitalization of Agricultural Machinery Rehabilitation. *Engineering Technologies and Systems*, 30(4): 711–722.
- [17] Wang B, Zhu J, Chai X, et al., 2025, Research Status and Development Trend of Key Technology of Agricultural Machinery Chassis in Hilly and Mountainous Areas. *Computers and Electronics in Agriculture*, 2025(226): 109447.
- [18] Xue G, Liu H, Wang G, et al., 2025, Design and Performance Testing of a Motorized Machine-Mounted Self-Leveling Platform for Hilly Orchards. *Agriculture*, 15(23): 2512.
- [19] Kumar V, Sharma K, Kedam N, et al., 2024, A Comprehensive Review on Smart and Sustainable Agriculture Using

IoT Technologies. Smart Agricultural Technology, 2024(8): 100487.

- [20] Wang P, Yue M, Yang L, et al., 2024, Design and Test of Intelligent Farm Machinery Operation Control Platform for Unmanned Farms. Agronomy, 14(4): 804.
- [21] Sun P, Liu X, 2025, Research on Remote Operation and Maintenance Based on Digital Twin Technology. Machines, 13(2): 151.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Intelligent Inspection and Closed-loop Management Innovation of Campus Fire Safety

Lantao Li*

Security Department, North China Electric Power University, Baoding 071003, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: As universities expand in scale and diversify in functions, traditional fire safety management faces challenges such as inefficient inspections and difficulty in tracing hazards. Technological innovation is key to enhancing efficiency, focusing on the core needs of campus fire safety. Intelligent inspection technology is applied to facility monitoring, hazard identification, and risk alerts, establishing a closed-loop system of monitoring, identification, disposal, and feedback. By leveraging the characteristics of IoT, AI, and big data technologies, the innovative solutions encompass technological integration, process optimization, and responsibility enhancement. This research aims to help universities overcome traditional bottlenecks, strengthen fire safety safeguards, and drive management transformation toward precision, intelligence, and efficiency.

Keywords: University campus; Fire safety; Intelligent inspection; Closed-loop management; Technological innovation

Online publication: February 12, 2026

1. Introduction

Higher education institutions, with their dense populations and concentrated teaching/research facilities, face critical fire safety challenges that directly impact faculty and student safety, property protection, and campus stability. Traditional manual inspections suffer from limited coverage, delayed hazard detection, and delayed corrective actions, making them inadequate for modern complex security management needs. With the rapid advancement of technologies like IoT, AI, and big data, intelligent inspection systems have emerged as a game-changer. These systems enable real-time facility monitoring, automatic hazard identification with precise location tracking, and closed-loop management to ensure full-process controllability of corrective actions. This represents an inevitable trend in fire safety management upgrades. This study focuses on relevant innovations to provide valuable references for optimizing fire safety management models in higher education institutions.

2. Current status and challenges of campus fire safety management in higher education institutions

2.1. The traditional inspection mode is inefficient and has prominent limitations

University campus buildings are dispersedly arranged, encompassing multifunctional zones such as teaching buildings, dormitories, laboratories, and libraries. The extensive yet scattered fire safety facilities require traditional inspection methods that rely on manual verification by staff, consuming substantial manpower and material resources while often leading to overlooked areas and human errors. The latent fire safety risks are difficult to detect through manual inspections in real time. Inspection results are typically recorded in paper ledgers or basic spreadsheets, with delayed information flow and hindered data sharing. This results in ineffective and non-targeted inspection work, failing to meet the demands of dynamic campus fire safety management ^[1].

2.2. The hidden danger rectification process is not closed-loop, resulting in disconnection of management

Currently, some higher education institutions face challenges in integrating hazard identification with corrective measures within their fire safety management systems. Inspections often lack a unified tracking mechanism, with ambiguous accountability and insufficient oversight during implementation, resulting in delayed feedback on remediation outcomes. Certain hazards persist due to delayed responses and incomplete rectifications, creating a vicious cycle of detection, neglect, and recurrence. The absence of systematic analysis of hazard data hinders comprehensive assessment of campus fire risks, prevents the formulation of targeted prevention strategies, and ultimately increases the likelihood of safety incidents.

2.3. Lack of precision in the maintenance and management of fire protection facilities

Higher education institutions maintain diverse fire safety systems, including automatic fire alarm systems, automatic fire suppression systems, emergency lighting, and evacuation guidance systems. Maintenance standards and cycles vary significantly across these facilities. Traditional maintenance approaches rely on fixed-frequency inspections and empirical judgments, lacking real-time monitoring of operational status, which often results in either excessive maintenance or neglect ^[2]. Outdated, malfunctioning, or damaged facilities are not promptly replaced, rendering them ineffective during emergencies and compromising campus fire safety. Incomplete maintenance records and inconsistent standards further hinder full lifecycle management, impeding subsequent maintenance efforts.

3. Technical support and application foundation of intelligent inspection for campus fire safety in universities

3.1. Real-time monitoring of fire protection facilities by using IoT technology

IoT technology equips fire protection systems with diverse sensors to collect operational data in real time. Fire hydrants and extinguishers are fitted with pressure sensors and position trackers, enabling real-time monitoring of pressure compliance and placement accuracy. The automatic fire alarm system incorporates temperature and smoke sensors, swiftly detecting anomalies and transmitting them to management terminals. Sensor data is wirelessly relayed to backend servers, enabling remote dynamic control of fire protection systems. This innovation eliminates the time and space constraints of traditional manual inspections, expands coverage, and enhances response efficiency, providing precise data support for fire safety management.

3.2. AI technology improves the accuracy of hidden danger identification

The image recognition and pattern analysis capabilities of artificial intelligence provide support for automatic identification of fire hazards in universities. Intelligent cameras deployed in critical campus areas utilize computer vision to automatically detect typical hazards such as improper use of fire and electricity, blocked fire exits, and obstructions to fire facilities ^[3]. By analyzing historical data through machine learning to build identification models, the system continuously improves detection accuracy and response efficiency. The intelligent system instantly delivers alerts to prompt administrators to take swift action, reducing human judgment errors and enabling early detection, timely warnings, and prompt resolution of hazards.

3.3. Big data technology helps to analyze risk situation

Big data technology integrates and analyzes campus fire safety data, including inspection records, hazard management archives, facility operational parameters, and real-time weather conditions. The analytical platform deconstructs this information through multidimensional analysis to identify high-risk zones, recurring hazard categories, and latent risk nodes. It generates risk assessment reports that enable management teams to develop targeted inspection and prevention strategies. This approach optimizes resource allocation, shifting from reactive responses to proactive prevention. The system provides data-driven support for fire facility maintenance planning and emergency response optimization, thereby enhancing the scientific rigor and predictive capabilities of campus fire safety management ^[4].

4. Intelligent inspection and closed-loop management system construction of campus fire safety in colleges and universities

4.1. Constructing a multi-level intelligent inspection system architecture

The campus fire safety intelligent inspection system adopts a four-layer architecture: perception layer, transmission layer, platform layer, and application layer. The perception layer utilizes sensor devices, smart cameras, and mobile inspection terminals to comprehensively capture data on fire facilities, environmental parameters, and personnel activities. The transmission layer ensures data flow through wireless communication networks and IoT gateways. The platform layer coordinates data storage, analysis, processing, and sharing while enabling centralized management and scheduling. The application layer provides diverse services, including task allocation and hazard reporting to meet campus fire safety management needs.

4.2. Establish a closed-loop management process for monitoring, identification, disposal, and feedback

Closed-loop management forms the core foundation for intelligent inspection. Smart equipment continuously monitors fire protection systems and environmental conditions, ensuring real-time data updates. AI automatically identifies potential hazards, with manual verification to guarantee accuracy. The rectification process clearly defines responsibilities, standards, and timelines, while the platform dynamically assigns tasks and tracks progress. Managers provide real-time supervision, and upon completion, executors upload progress reports. The system verifies, accepts, and archives the results, with data incorporated into big data analytics to support future optimizations, forming a closed-loop process.

4.3. Improving the responsibility and coordination mechanism of closed-loop management

The implementation of a closed-loop management system requires clear delineation of responsibilities and coordinated collaboration. It is essential to define the roles of the school's fire safety management department, secondary units, and individuals, with inspection, rectification, and maintenance responsibilities progressively assigned to ensure accountability at every level. A cross-departmental collaboration platform should be established to strengthen coordination between fire safety, logistics, student management, and academic departments, enabling information sharing, resource integration, and joint response. Hazard rectification should be led by the logistics department for facility maintenance, with student management departments assisting in regulating student conduct. The performance evaluation mechanism should incorporate fire safety management effectiveness into departmental and individual assessments, motivating proactive performance and ensuring the stable operation of the closed-loop system.

5. Innovative path of intelligent inspection and closed-loop management of campus fire safety in colleges and universities

5.1. Technology integration and innovation enhance the overall efficiency of the system

The deep integration of IoT, AI, and big data technologies drives the functional optimization of intelligent inspection systems. AI-powered image recognition and IoT sensor data mutually reinforce each other, expanding the scope of hazard identification and enhancing detection accuracy. Big data analytics thoroughly examine inspection and rectification data, providing scientific support for fire safety risk prediction and decision-making. Mobile internet technology enhances the functionality of inspection terminals, enabling real-time task acceptance and instant hazard reporting. This improves inspection convenience and efficiency while facilitating seamless coordination with smart fire protection equipment for remote control and intelligent resource allocation, ultimately elevating campus fire safety emergency response capabilities.

5.2. Innovation in management models to optimize resource allocation efficiency

Intelligent inspection and closed-loop management are revolutionizing fire safety management models in universities. By transcending traditional decentralized frameworks, they establish centralized intelligent control platforms to facilitate streamlined and refined management. Leveraging big data analytics, these systems optimize inspection routes and frequencies, delineate risk zones, conduct intensive checks in high-risk areas, and allocate resources scientifically in low-risk zones, thereby enhancing targeted and efficient inspections. The framework also establishes a full lifecycle management system for fire protection facilities, systematically documenting procurement, installation, and other process information. This promotes precise standardization of maintenance, reduces costs, and optimizes resource allocation.

5.3. Concept and cultural innovation to strengthen the safety awareness of all staff

The implementation of technology and closed-loop management requires a safety-oriented culture. Universities should enhance fire safety education by utilizing platforms such as official websites, WeChat accounts, and smart inspection systems to disseminate fire safety knowledge, share case studies of hazards, and showcase rectification outcomes. This strengthens faculty and students' awareness and self-protection capabilities. The intelligent platform introduces a hazard reporting channel to facilitate feedback on campus fire risks, fostering a collaborative

environment where everyone participates in governance. Integrating fire safety culture into the overall campus culture, activities like themed events, knowledge contests, and emergency drills reinforce responsibility. This approach transforms fire safety awareness and active participation into a natural habit, laying a solid ideological foundation for the smooth advancement of related initiatives.

6. Conclusion

The intelligent inspection and closed-loop management innovation in campus fire safety is pivotal to addressing complex challenges and advancing modern management standards. Empowered by technologies like IoT, AI, and big data, these systems overcome traditional limitations by enabling dynamic monitoring of fire facilities, precise hazard identification, and proactive risk alerts. The closed-loop system of monitoring, identification, response, and feedback ensures full-chain controllability in hazard rectification. To enhance management effectiveness, universities must align with their specific contexts, fostering collaborative innovation in technology, management, and safety culture while refining relevant mechanisms. With continuous technological advancements, campus fire safety management will become increasingly intelligent, refined, and efficient, creating a secure environment for faculty and students and providing robust support for the high-quality development of higher education institutions.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Feng Q, 2025, Research on Improving the Efficiency of University Campus Facilities Resource Allocation Based on Data-Driven. *Education Reform and Development*, 7(10): 270–277.
- [2] Zou B, Wang J, Li X, et al., 2026, Interface In-Situ Generated Two-Component Phosphates Derived from Black Phosphorus Nanohybrid Endow Outstanding Thermal Stability and Fire Safety to Polycarbonate Composites. *Polymer Degradation and Stability*, 2026(245): 111884.
- [3] Zhang Y, Huang X, Usmani A, 2026, The Challenge of Optimizing Building Renovations for Urban Sustainability and Fire Safety. *Nature Cities*, 2026(prepublish): 1–2.
- [4] Xiao Y, Chen D, Huang H, et al., 2025, Micro-Rebar Architecture Enable Fire-Safety and Tough PVA/AG Aerogels. *ACS Appl Mater Interfaces*, 18(1): 2748–2763.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Innovative Research on Construction Models and Operational Management of Electrical and Electronic Laboratories

Lei Wang*

Non-commissioned Officer Academy of PAP, Hangzhou 311400, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Electrical and electronic laboratories are crucial for developing engineering talent, yet they face challenges such as outdated hardware, rigid management, and faculty shortages. This paper proposes an integrated reform model featuring virtual-physical equipment upgrades, open and intelligent management platforms, a dual-qualified teaching team, and a full-process safety assurance system. It offers a practical framework for modernizing such laboratories and supporting the cultivation of high-quality innovative engineering professionals.

Keywords: Electrical and electronic laboratory; Construction model; Management model; Talent development; Safety management

Online publication: February 12, 2026

1. Introduction

Electrical and electronic engineering is a foundational technology that underpins the development of strategic emerging industries such as modern information technology, intelligent manufacturing, new energy, and aerospace. The proficiency of relevant professionals in the knowledge system of electrical and electronic engineering directly determines the quality of their training. In higher engineering education, theoretical instruction and practical training are as indispensable as the two wheels of a cart or the two wings of a bird. Electrical and electronic laboratories serve as the primary battleground for practical teaching. They function as both a “testing ground” for theoretical knowledge and a “training camp” for fundamental skills, while also acting as an “incubator” for cultivating engineering thinking, stimulating innovation capabilities, and shaping a scientific spirit.

As the wave of higher engineering education reform, epitomized by the “New Engineering” initiative, deepens, and as industry demands increasingly higher levels of engineering practice capabilities and innovative qualities from graduates, the traditional concepts and management models of electrical and electronic laboratories

face unprecedented challenges. Historically, these laboratories primarily focused on verification experiments, featuring single-function equipment, rigid management approaches, and insufficient resource accessibility. Such laboratory structures struggle to meet the new demands for cultivating students' abilities to solve complex engineering problems and integrate across disciplines. Consequently, systematically reconstructing the laboratory's hardware environment, innovatively optimizing its operational management mechanisms, building a high-caliber experimental teaching team, and establishing a robust safety defense system to comprehensively enhance the laboratory's overall effectiveness have become a critical and urgent issue in university teaching reform^[1,2].

Against this backdrop, this paper aims to transcend the limitations of conventional thinking. It seeks to comprehensively examine and explore the construction and management of electrical and electronic laboratories from a more holistic, systematic, and forward-looking perspective. By analyzing the current state, diagnosing issues, and proposing solutions, it seeks to explore a new path for laboratory construction and management that not only meets current teaching needs but also adapts to future technological developments. This approach combines scientific rigor, efficiency, and safety, providing a solid platform to cultivate high-caliber electrical and electronic professionals capable of addressing future challenges.

2. The core role and value reinvention of electrical and electronic laboratories in talent development

Against the backdrop of engineering education accreditation, the cultivation of students' engineering practice and innovation capabilities has been elevated to unprecedented importance. The role of electrical and electronic laboratories extends far beyond merely supporting theoretical instruction as they are undergoing a profound reevaluation of their core functions.

2.1. From theoretical validation to knowledge-building accelerator

Traditional laboratory instruction primarily aims to verify theoretical formulas and theorems, such as validating Kirchhoff's Voltage Law (KVL) and Current Law (KCL) through actual circuit connections. However, modern educational philosophy places greater emphasis on active knowledge construction. A well-designed laboratory environment should guide students in transitioning from "verifiers" to "explorers." For instance, after completing fundamental circuit characteristic tests, students can engage in extended tasks such as "analyzing the impact of parameter variations on system performance" or "investigating errors introduced by non-ideal components." When students observe phenomena in experiments that deviate slightly from theoretical predictions, they should be guided to analyze the source of these deviations, whether it stems from instrument accuracy, connection wire resistance, or environmental interference. This process itself represents a deeper level of understanding, critical evaluation, and internalization of theoretical knowledge, achieving a leap in knowledge construction from merely knowing "what" to comprehending "why"^[3].

2.2. From skill training to engineering capability development arena

Proficiency in using instruments such as oscilloscopes, signal generators, and multimeters, along with the ability to correctly solder and assemble circuits, constitutes the fundamental skills in the field of electrical and electronic engineering. However, modern engineering capabilities extend far beyond these basics, encompassing comprehensive competencies such as system design, project management, teamwork, and communication skills. Laboratories should serve as integrated training grounds for cultivating these competencies. By adopting the

Project-Based Learning (PBL) model, students work in teams to experience the complete engineering process: project initiation, solution design, simulation optimization, procurement and material preparation, hardware implementation, debugging and testing, report writing, and final presentation. For instance, a project like “Microcontroller-Based Intelligent Temperature and Humidity Monitoring System” not only hones students’ circuit design, programming, and debugging skills but also subtly cultivates essential soft competencies: project planning, cost control, team division of labor and collaboration, and technical documentation writing. These represent the core competencies indispensable for future engineers ^[4].

2.3. From imitative learning to innovation-sparking catalyst

Innovation stems from free exploration and persistent trial and error. Traditional “recipe-style” experiment manuals largely constrain students’ thinking. Establishing open innovation labs equipped with abundant components, open-source hardware (e.g., Arduino, Raspberry Pi), programmable logic devices (e.g., FPGA), and essential testing facilities encourages students to pursue self-selected topics and engage in free exploration ^[5]. Moreover, an “Innovation Fund” program that can support students’ unconventional ideas should be established, and competitions like electronic design contests or the Challenge Cup should be organized to foster learning and innovation through competition ^[6]. When students repeatedly test different solutions and overcome technical hurdles to solve real-world problems, their critical thinking, divergent thinking, and innovative problem-solving abilities are most effectively stimulated and honed.

2.4. From individual learning to collaborative community for teamwork and professional growth

Modern engineering projects are almost invariably the result of teamwork. The open environment of the laboratory naturally provides students with a space for collaborative exchange. Within project teams, members must jointly discuss technical solutions, allocate tasks reasonably, cross-check code and circuits, and collectively overcome debugging challenges. This process not only enhances technical expertise but also cultivates team spirit, accountability, and communication skills. Simultaneously, the laboratory’s rigorous 5S management (Sorting, Straightening, Sweeping, Sanitizing, Sustaining), strict safety protocols, and instrument usage standards subtly instill a meticulous, disciplined scientific approach and strong professional ethics. This lays a solid foundation for students’ seamless transition from “student” to “professional” ^[7].

3. Analysis of deep-seated issues in the construction and management of electrical and electronic laboratories

Although the importance of laboratories has become widely recognized, their construction and management still face numerous deep-seated contradictions that require urgent resolution in many universities, particularly regional institutions and newly established colleges.

3.1. The aging and unbalanced hardware infrastructure cannot support cutting-edge teaching

Current laboratory equipment faces numerous issues as follows:

- (1) Equipment upgrades lag behind technological advancements. Many labs still rely on traditional box-style workbenches with fixed core functions and unchanged experimental procedures for years. Advanced

equipment widely adopted in industry, such as virtual instruments, programmable system-on-chip devices, industrial robots, and IoT modules, remains scarce in academic settings. This gap between “what is taught in schools” and “what is used in enterprises” forces graduates into lengthy retraining cycles;

- (2) Equipment configurations suffer from structural imbalance. On one hand, low-end equipment for basic experiments is redundantly over-configured with low utilization rates. On the other hand, high-end and specialized equipment essential for cutting-edge research and comprehensive, innovative experiments is severely lacking. This contradiction of “too much of some and too little of others” limits the depth and breadth of experimental teaching content;
- (3) A widespread tendency to “prioritize hardware over software” prevails. Equipment procurement often focuses solely on hardware specifications while neglecting complementary software, simulation platforms, teaching resource libraries, and post-purchase upgrade services. Consequently, advanced hardware devices fail to achieve their full potential due to the absence of suitable “souls” to complement them^[8,9].

3.2. The closed and inefficient management model stifles the enthusiasm of both faculty and students

Laboratory management faces multiple challenges. The “teacher-centered” closed-door approach strictly limits operating hours to scheduled classes only. Experiment content, procedures, and methods are predetermined by instructors, leaving students to execute tasks passively. While this “nanny-style” management facilitates control, it severely stifles students’ initiative for independent learning and exploration. Simultaneously, low levels of laboratory informatization result in high management costs. Many labs still rely on paper logbooks for equipment borrowing and consumable collection, while experiment reservations require multi-level relay through class representatives or student committee members. This traps administrators in tedious administrative tasks, preventing effective analysis and optimization of equipment status, usage data, and experimental outcomes. Furthermore, the evaluation system is simplistic and lacks incentive-driven mechanisms. Student assessments often focus solely on whether lab reports are neatly formatted or data aligns with expectations, overlooking critical aspects like critical thinking, experimentation, collaboration, and innovative contributions during the process. Similarly, evaluations for lab instructors and administrators fail to adequately reflect the complexity and creativity inherent in their roles, ultimately undermining team stability and work enthusiasm^[10].

3.3. The insufficient quantity and capability of the teaching staff constrain the improvement of educational standards

University laboratory teaching staff face multiple challenges. In terms of team structure, there exists a “generation gap” issue. High-level laboratory instructors require both solid theoretical foundations and extensive engineering practice experience. However, universities often prioritize academic publications and research projects during talent recruitment, making it difficult to attract “dual-qualified” professionals with industry backgrounds. Among existing faculty, younger instructors lack experience, while older ones may have outdated knowledge structures, creating a generational disconnect. Career advancement pathways remain narrow. Within many university evaluation systems, faculty in experimental teaching series face disadvantages in promotion and resource allocation compared to theoretical course instructors and research-focused faculty. This discourages talented individuals from pursuing experimental teaching roles, resulting in poor team stability and high turnover. Furthermore, the training

system is inadequate, with insufficient systematic and cutting-edge training for laboratory technicians. This hinders their ability to keep pace with new technologies and master new equipment, thereby affecting the updating of experimental teaching content and the improvement of teaching effectiveness^[11,12].

3.4. The weak safety management awareness system poses significant latent risks

Laboratory safety management harbors significant hidden dangers that cannot be overlooked. There is a severe lack of safety culture and weak safety awareness. Faculty and students commonly hold the mistaken notion of “prioritizing operations over safety,” believing that low-voltage electrical and electronic experiments pose minimal risk. They lack sufficient understanding of potential hazards such as electric shock, short circuits, fires, device explosions (e.g., reverse connection of electrolytic capacitors), and high-frequency radiation. This leads to safety education and training remaining superficial and failing to take root. Conversely, there are significant shortcomings in policy enforcement and regulatory blind spots. Although formal laboratory safety codes exist, factors like tight class schedules and insufficient oversight personnel often lead to compromised implementation. This is particularly evident in student-led open labs and innovation projects, where unsafe behaviors lack effective real-time monitoring and early warning mechanisms. Furthermore, inadequate investment in safety facilities and poor maintenance are prominent issues. Laboratories commonly lack essential safety signage, emergency lighting, and first-aid equipment. Critical safety installations like residual current devices (RCDs) and fire extinguishers are not subject to mandatory periodic inspections, making their effectiveness difficult to guarantee. Should an accident occur, the consequences could be severe^[13].

4. Systematic improvement measures and implementation pathways for the construction and management of electrical and electronic laboratories

To address the aforementioned issues, systematic reform measures must be implemented, advancing in a coordinated manner across four dimensions (hardware, management, personnel, and safety) to achieve the comprehensive transformation and upgrading of laboratories.

4.1. Establish a modern hardware platform system that integrates virtual and physical elements and advances in a tiered, progressive manner

To comprehensively enhance laboratory construction standards, coordinated advancement across three levels is required as outlined:

- (1) At the foundational level, emphasis should be placed on updating and optimizing traditional laboratory equipment. Core, frequently used fundamental equipment should undergo systematic upgrades to ensure precision and reliability. Concurrently, older equipment with functional viability should have its lifespan extended through third-party technical services and regular calibration. Additionally, modular retrofits can be implemented for certain traditional lab benches, converting fixed experimental circuits into freely pluggable modules. This approach increases experimental flexibility, facilitates student-designed circuitry, and reduces long-term maintenance costs;
- (2) In the enhancement layer, vigorously develop a “virtual-physical integration” experimental model. On one hand, introduce high-performance virtual instruments by procuring PC-based virtual instrument systems and building integrated hardware-software experimental platforms with relevant software. This allows students to perform circuit simulation and program design on computers before conducting real-time data

acquisition and system control via hardware platforms, achieving seamless integration between simulation and reality. In addition, establish a cloud-based circuit simulation laboratory. Utilize specialized software to build an online simulation platform, enabling students to design circuits and perform simulation analyses anytime, anywhere via personal computers. This effectively alleviates the spatial and temporal constraints of physical laboratories, making it particularly suitable for experiment preparation, complex system analysis, and experimental instruction during special periods;

- (3) At the innovation level, an open, comprehensive innovation platform must be established. This includes setting up open innovation labs with dedicated spaces equipped with diverse tools like 3D printers and laser engravers, operating on a 24*7 reservation basis to support student course projects and graduation designs. Additionally, cutting-edge industrial-grade equipment such as PLCs should be introduced based on the university's disciplinary strengths, exposing students to real-world industrial technology ecosystems and broadening their technical horizons ^[14].

4.2. Implementing a modern management model characterized by “smart openness and data-driven decision-making”

To elevate laboratory construction and management to new heights, coordinated efforts must be made across three dimensions: system development, management mechanisms, and evaluation incentives. For instance:

- (1) An integrated laboratory information management system should be established, with core functional modules encompassing access control, equipment reservations, consumables requisitioning, fault reporting, data statistics, and security monitoring. Students can reserve lab benches and borrow instruments via a mobile app or web portal, enabling one-click service. Simultaneously, the system automatically records equipment utilization rates, failure rates, and other metrics. Through data analysis, administrators can accurately assess equipment configuration rationality, optimize consumables procurement plans, and provide data support for teaching reforms;
- (2) Fully implement an open management operation mechanism. While ensuring scheduled teaching activities, maximize laboratory access hours, including evenings and weekends, by permitting cross-disciplinary and cross-grade student entry. Beyond mandatory experiments, laboratories will offer diverse, abundant elective projects for student selection. Students are actively encouraged to propose independent experimental topics aligned with their interests and research directions. To support students in conducting independent innovation experiments, the laboratory will establish a “Project Mentor System.” This system will assign specialized mentors to each student's independent innovation project, providing comprehensive and detailed guidance throughout all stages: experimental design, implementation, and problem-solving;
- (3) We will reform the experimental teaching evaluation and incentive mechanisms by establishing a diversified assessment system. This system will incorporate pre-experiment preparation and operational compliance into overall grades, utilize information systems to track student participation and contributions, and introduce awards such as the “Experimental Teaching Achievement Award.” Achievements like guiding academic competitions will serve as the key criteria for professional title evaluations and performance assessments, thereby stimulating the intrinsic motivation of the teaching team ^[15].

4.3. Build a dual-qualified faculty team that combines full-time and part-time instructors with outstanding capabilities

To strengthen the development of the experimental teaching faculty, a dual-pronged approach combining internal cultivation and external recruitment is essential. At the internal cultivation and advancement level, establish a regular training mechanism to periodically organize faculty participation in training on new technologies, equipment, and teaching methods. Simultaneously, encourage and fund experimental instructors to undertake visits, exchanges, and job shadowing at high-level domestic and international universities or renowned enterprises. We must also promote mutual growth between teaching and research by encouraging laboratory instructors to engage in research projects, transforming research outcomes into comprehensive and research-oriented laboratory programs, and support them in summarizing experiences from laboratory management and teaching reforms, applying for teaching research projects, and publishing papers on pedagogical innovation. Regarding external recruitment and resource sharing, we should flexibly recruit industry experts by appointing senior corporate engineers and technical directors as industry mentors or adjunct professors. They should be invited to regularly deliver lectures on campus, supervise graduation projects, and collaborate on experimental development, thereby introducing cutting-edge engineering case studies and practical expertise into the academic environment. Concurrently, establish joint university-industry laboratories through collaborations with leading enterprises. Companies provide equipment, technical expertise, and training support, while the university contributes facilities and talent resources. This approach not only enhances hardware capabilities but also cultivates a permanent cohort of industry mentors

[16].

4.4. Establish a safety assurance system characterized by “full participation and end-to-end controllability”

To fortify laboratory safety defenses, a systematic approach must be implemented across three dimensions: institutional culture, technical controls, and emergency response as listed:

- (1) Prioritize institutional frameworks and cultural norms: On one hand, establish a robust safety accountability system by clearly defining responsibilities at all levels, including the college, laboratory directors, faculty advisors, administrative staff, and students, through signed safety agreements that enforce the principle of “whoever uses is responsible, whoever manages supervises.” Additionally, cultivate a safety-conscious laboratory culture. Conduct regular safety awareness campaigns through new student orientation, safety knowledge competitions, promotional posters, and WeChat official account posts. Implement safety knowledge assessments as a mandatory prerequisite for laboratory access, where individuals who fail the assessment are prohibited from entering;
- (2) Prioritize both technical and human safeguards: Strengthen safety infrastructure by equipping every lab with adequate firefighting equipment, emergency medical kits, and emergency power-off buttons. Physically isolate high-voltage experimental zones with warning signage and provide protective gear for high-frequency/high-voltage experiments. Implement intelligent safety monitoring by installing video surveillance and smoke detectors for real-time alerts. Equip critical devices with smart power management systems for automatic overcurrent, overvoltage, and leakage protection, integrated with management systems to log abnormal operations;
- (3) Emergency management and response must be strengthened. Detailed contingency plans for incidents such as electric shock, fire, and chemical spills should be developed and prominently displayed.

Additionally, at least one full-scale fire evacuation and first-aid drill involving all personnel should be conducted each semester to ensure faculty and students are familiar with escape routes and basic first-aid techniques, thereby effectively enhancing practical capabilities for handling emergencies ^[17].

5. Conclusion

The construction and management of electrical and electronic laboratories constitute a complex systemic endeavor involving philosophy, technology, institutional frameworks, and culture. Their modernization is by no means an overnight transformation. This paper systematically constructs a four-pronged reform framework, grounded in a “hybrid hardware platform,” driven by an “intelligent and open management model,” anchored by a “dual-qualified faculty,” and safeguarded by a “comprehensive safety system covering all personnel and processes.”

This framework aims to achieve as follows:

- (1) Meet multi-tiered teaching demands from fundamental verification to cutting-edge innovation through tiered hardware development;
- (2) Unlock laboratory resource potential and stimulate faculty-student initiative via information-driven and open management reforms;
- (3) Cultivate a dual-qualified faculty pool proficient in both theory and practice;
- (4) Fortify the lifeline of sustainable laboratory development through a safety system emphasizing both cultural values and technological safeguards.

Looking ahead, with the rapid advancement of technologies such as artificial intelligence, big data, 5G, and the Internet of Things, electrical and electronics laboratories will inevitably evolve toward greater intelligence, networking, and personalization. Only by adhering to the principles of “student-centeredness, competency-based education, and safety as the bottom line,” and through continuous exploration and bold implementation, can electrical and electronic laboratories truly become the cradle for cultivating outstanding engineering talent. This will contribute indispensable educational strength to China’s transformation from a major manufacturing nation to a manufacturing powerhouse.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Wan Z, Yao Y, 2021, An Exploration of Issues in Building Open Electrical Engineering Laboratories Under the New Engineering Paradigm. *Science and Technology & Innovation*, 2021(19): 49–50.
- [2] Chen L, Du Q, Qiu W, et al., 2018, Exploration on Reform and Practice of University Laboratory Safety Education under Background of “New Engineering”. *Experimental Technology and Management*, 35(8): 260–262.
- [3] Liu W, Zhang Y, Zhou P, 2025, Design of High Frequency Electronic Circuit Experiment Teaching Reform Scheme Based on Virtual and Reality Combination. *Laboratory Science*, 28(2): 119–123.
- [4] Qian Q, 2022, Application of Project Teaching Method in Electronic and Electrical Course. *Application of IC*, 39(3): 166–167.
- [5] Zhang L, Yang W, Li D, et al., 2016, Exploration on Cultivating Students’ Innovative Thinking Mode Based on

Opening Experiment. *Experimental Technology and Management*, 33(7): 29–32.

- [6] Qu Y, Zhao S, Zheng W, 2019, Exploring the Development of a Student Innovation Capacity Cultivation System Based on the “Challenge Cup” Competition: Taking the Food Science and Engineering Major as an Example. *Heilongjiang Education (Research and Evaluation of Higher Education)*, 2019(1): 55–56.
- [7] Xu Y, 2022, Application of 5S Concept in College Laboratory Management. *Jiangsu Science and Technology Information*, 39(28): 58–60.
- [8] Chen Z, Yang Z, 2021, Exploring the Effect of Resource Misallocation on the Basic Research Efficiency. *Science and Technology Management Research*, 41(9): 69–76.
- [9] Gong S, 2016, Research on Measuring and Testing Technology of Aging Test Platform for Electronic Components, thesis, Harbin Institute of Technology.
- [10] Wang X, Dong H, Zhang H, et al., 2025, Optimization and Implementation of Safety Management Strategies for Electrical and Electronic Laboratories. *Tech-Inspired*, 2025(27): 142–144.
- [11] Song K, 2024, Research on the Countermeasures of the Construction of “Double Qualified” Teachers in Higher Vocational Education. *Journal of Jiangsu Vocational Institute of Architectural Technology*, 24(4): 80–83.
- [12] Han H, Zhang Y, 2016, Opportunities and Challenges Facing University Laboratory Instructors in the New Era. *Science & Technology Vision*, 2016(23): 50–70.
- [13] Jiang X, Fei T, Yu B, et al., 2025, Analysis of Safety Management Mechanism in Electrical and Electronic Laboratory. *Electronic Technology*, 54(6): 246–247.
- [14] Guo B, 2011, The Research and Application of Electronic Information Experimental Platform Based on Virtual Instrument, thesis, Hubei University of Technology.
- [15] Zhao T, Wang M, Liu Y, et al., 2023, Design and Development of an Information Management System for Teaching Laboratories in Higher Education Institutions. *Computer Knowledge and Technology*, 19(8): 59–61.
- [16] Zhang T, 2023, Construction of Experimental Teacher Team under the Background of “New Engineering”. *Experimental Technology and Management*, 40(S1): 20–23.
- [17] Yun X, 2021, Research on the Optimization of the Safety Management Responsibility System of University Laboratory in China, China University of Geosciences, Beijing.

Publisher’s note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Design of Integrated Brushless Motor Drive and Control System for Robotic Arm Joints

Wei Liu, Tingyu Li

School of Information and Control Engineering, Jilin University of Chemical Technology, Jilin 132022, Jilin, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: To meet the requirements of high performance, low cost, and modularity in robotic arm joints, this study designs an integrated brushless motor drive and control system. The system selects the STM32G473 microcontroller as the control chip and adopts field-oriented control as the primary motor control algorithm. Meanwhile, the drive and control system design is completed from both hardware and software aspects. Finally, the study performs closed-loop experiments on the robotic arm joint. The experimental results demonstrate that the designed drive and control system for robotic arm joints exhibits a favorable dynamic response and steady-state performance, making it suitable for controlling desktop-level and lightweight robotic arms.

Keywords: Robotic arm joint; Brushless motor; Motor drive; FOC

Online publication: February 12, 2026

1. Introduction

Since the mid-20th century, with the rapid development of computer technology and automatic control, robotic arm technology has emerged and quickly become an important driving force for industrial automation and intelligence^[1]. Over several decades of development, robotic arm technology has gradually matured, with research extending beyond traditional industrial manufacturing into emerging domains such as healthcare, service, and education. These new application scenarios impose higher demands on the lightweight design and flexibility of robotic arms^[2]. In the control of a robotic arm, the joint motor, as the actuator driving the arm, has a direct impact on stability, dynamic performance, and execution accuracy^[3]. However, traditional robotic joints often employ a separate drive-control design, which generally suffers from high costs and low integration. These drawbacks become increasingly pronounced in miniaturized and modular application scenarios^[4]. Therefore, balancing performance, cost, and integration in joint drive and control systems has become a current research focus, and integrated drive and control design has emerged as a new development trend. In response to the above needs, this study designs an integrated drive and control system for robotic arm joints, conducting research from two aspects: hardware circuit design and motor control algorithm development. Experiments verify the system's performance advantages in brushless motor control.

2. System hardware design

In terms of hardware selection, factors such as performance, cost, and integration are considered. The system hardware block diagram is shown in **Figure 1**.

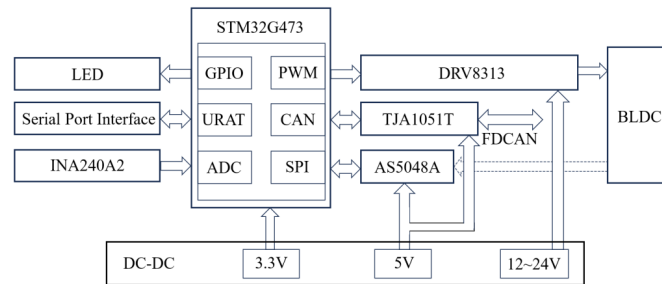


Figure 1. Hardware block diagram of the brushless motor drive and control system.

The main control chip selects the STM32G473 microcontroller as the computational core for the motor control algorithm. The drive section utilizes the DRV8313 three-phase brushless motor driver, and together with the AS5048A magnetic encoder and INA240A2 current-sensing chip, implements closed-loop motor control. Additionally, the drive and control board offers an FDCAN interface for data exchange with the host computer. The circuit design of each module is described in detail below.

2.1. Control module

The circuit design of the control module is shown in **Figure 2**, which comprises the MCU and its peripheral circuits. The motor drive control board selects the STM32G473 microcontroller as the control chip, featuring a single-precision FPU and a DSP instruction set, with a maximum frequency of 170 MHz. It is capable of handling the complex computations required by motor control algorithms. The MCU communicates with sensors, the driver module, and the host computer via SPI and FDCAN interfaces. Additionally, a UART serial port and an LED indicator are reserved to facilitate subsequent system development and debugging.

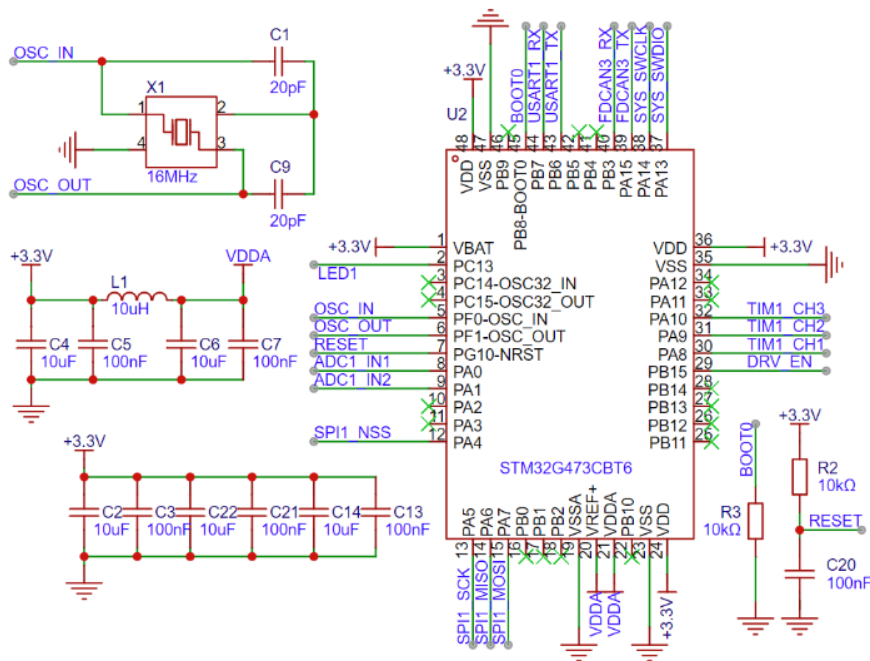


Figure 2. Control module circuit.

2.2. Motor drive module

The motor drive module employs the DRV8313 integrated driver chip, which internally integrates a complete pre-driver and three-phase half-bridge circuit. The input supply voltage ranges from 8V to 60V, and each half-bridge channel can deliver a peak current of up to 2.5A. The circuit of the drive module is shown in **Figure 3**. The STM32 controls the enable and conduction states of the three-phase drive through ENx and INx pins, respectively. The pre-driver amplifies these control signals to drive the power MOSFETs, which ultimately output to the three-phase windings via OUTx pins, achieving motor drive and control.

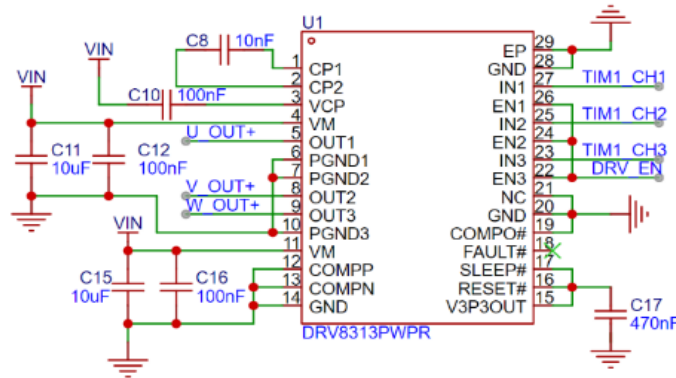


Figure 3. Motor drive module circuit.

2.3. Current sensing module

The circuit of the current sensing module is shown in **Figure 4**. The three-phase currents are measured using a combination of the dual-shunt resistor sampling method and the built-in current sensing function. The circuit directly connects the shunt resistors in series with the output terminals of motor phases U and V, amplifying the differential voltage across each resistor with a fixed gain of 50 V/V. The resulting output voltage signals are then fed into the ADC channels of the STM32.

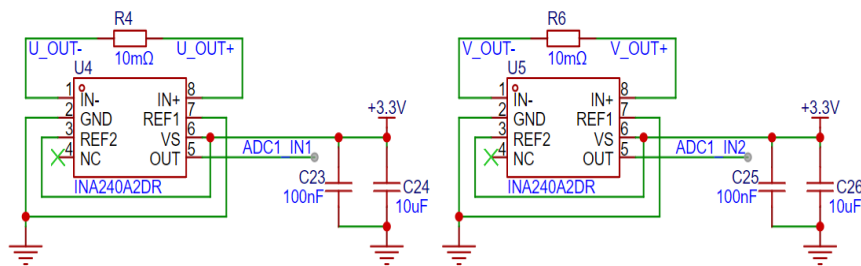


Figure 4. Current sensing module circuit.

2.4. Communication module

In this study, the robotic arm joint employs a distributed control architecture. Each joint controller exchanges high-frequency, bidirectional data with the host computer within a short time frame. Consequently, the communication scheme must ensure high speed, low latency, and strong reliability. To meet these requirements, the system uses the FDCAN protocol and employs the TJA1051T high-speed CAN transceiver, which supports data rates up to 5 Mbit/s to ensure reliable communication. The FDCAN communication module circuit is shown in **Figure 5**.

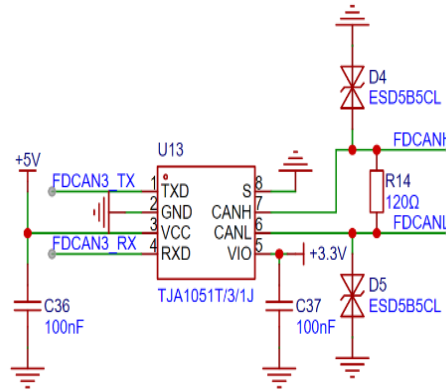


Figure 5. Communication module circuit.

2.5. Step-down module

The input voltage range of the motor drive and control board designed in this paper is 12V to 24V. To meet the power requirements of different modules, a multi-stage step-down scheme is adopted. As shown in **Figure 6**, the first-stage step-down circuit employs the synchronous step-down DC-DC converter SY8303AIC to convert 24 V to 5 V. The second-stage step-down then utilizes the linear LDO regulator RT9013-33GB to convert 5V to 3.3V.

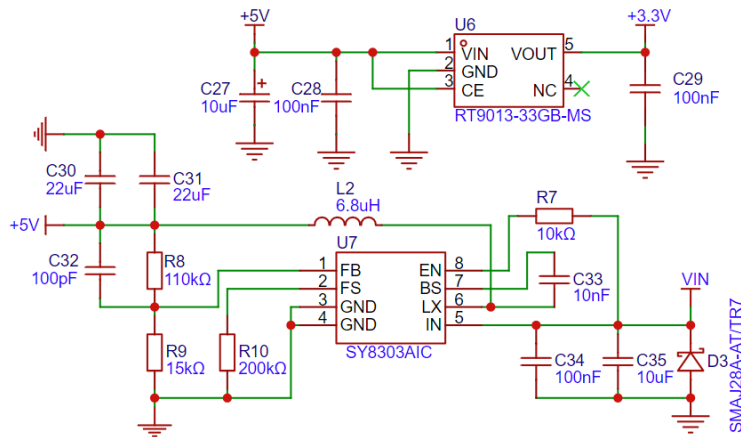


Figure 6. Step-down module circuit.

2.6. Magnetic encoder module

To improve the control accuracy and real-time performance of the robotic arm joints, the AS5048A magnetic rotary encoder is selected to acquire the motor angle information. This encoder features 14-bit high-resolution absolute angular measurement. The magnetic encoder module circuit is shown in **Figure 7**. Communication with the STM32 is achieved via the SPI interface, enabling high-speed data transmission and ensuring that joint angle information is fed back to the STM32 in real-time.

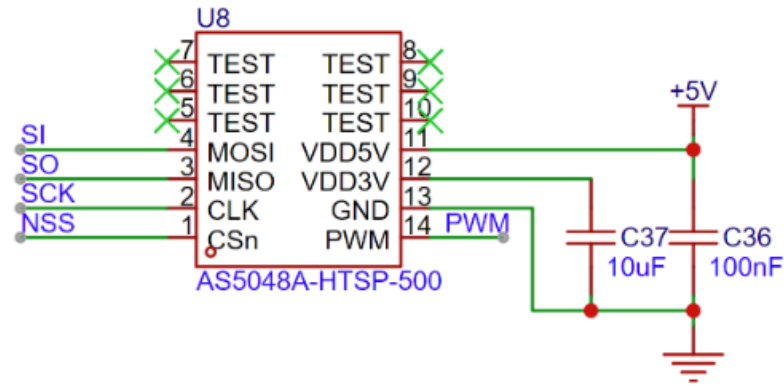


Figure 7. Magnetic encoder module circuit.

2.7. PCB design

The robotic arm joints employ an integrated design for the drive unit, which requires the integration of the reducer, motor, position sensor, and drive control board within the joint. Therefore, the designed PCB should be as compact as possible. Board size constraints make two-layer routing difficult, so the design adopts a four-layer PCB. The PCB design and physical layout are shown in **Figure 8** and **Figure 9**, with a board outline of 49 mm * 10 mm and all components placed on the top layer only. The top and bottom layers serve as signal routing layers, the GND layer forms a continuous ground plane, and the power layer carries only partial power traces.

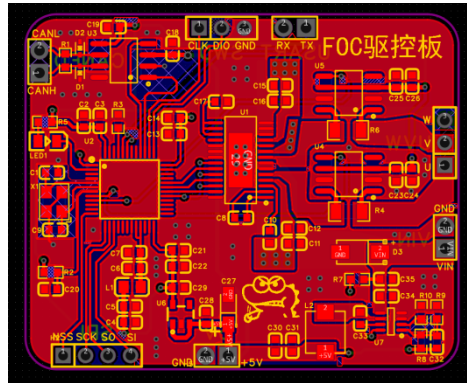


Figure 8. PCB layout of the board.

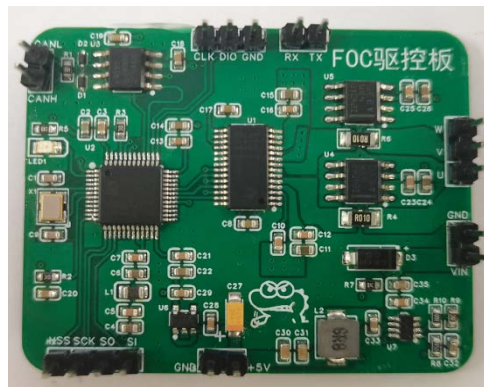


Figure 9. Physical board of the board.

3. Brushless motor FOC control

The Common control methods for BLDC motor include square wave control, sine control, and FOC control. In contrast, the first two methods suffer from limitations in terms of precision and stability. The FOC control can achieve high efficiency, smooth torque output, and high-precision position control over the full speed range, better meeting the stability and precision requirements of robotic arm joints. Therefore, this study adopts FOC as the core control algorithm for the BLDC motor.

FOC, also known as magnetic field-oriented control, is a type of vector control ^[5]. The basic framework of FOC control is shown in **Figure 10**. Its core idea is to decompose the stator current into a direct-axis component (I_d) controlling the field excitation and a quadrature-axis component (I_q) controlling the torque through a series of coordinate transformations, and regulating them independently to achieve high-precision control of the motor's magnetic field and torque. Meanwhile, three cascaded loops-position, speed, and current-are adopted, and the performance of each loop is optimized using PID control.

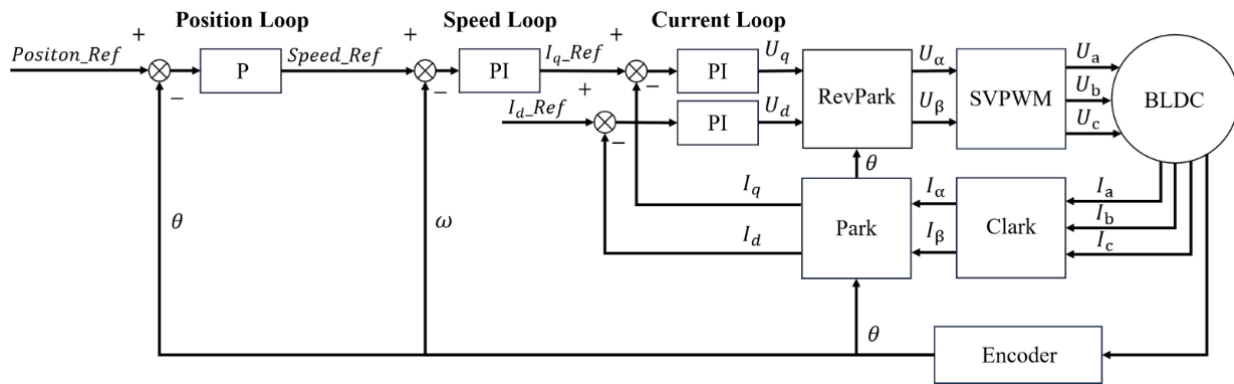


Figure 10. FOC three-loop control framework.

3.1. Clarke and Park transformation

The Clarke transformation is based on the principle of electromagnetic equivalence between the three-phase winding and two orthogonal windings ^[6]. It projects the current variables from the three-phase stationary reference frame onto a two-phase stationary orthogonal coordinate system. As shown in **Figure 11**, the current variables I_a , I_b , and I_c in the A-B-C coordinate system are projected on the α and β axes, to obtain the current components I_α and I_β in the two-phase stationary coordinate system.

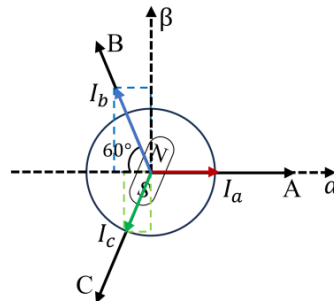


Figure 11. Clark transform.

Based on the coordinate relationships shown in **Figure 11**, the Clarke transformation can be expressed as follows:

$$\begin{cases} I_\alpha = I_a \\ I_\beta = \frac{1}{\sqrt{3}}(I_a + 2I_b) \end{cases} \quad (1)$$

After the Clarke transformation, the control variables are reduced from three to two, while the signal frequency remains unchanged before and after the transformation. However, the transformed variables I_α and I_β , remain sinusoidal and thus nonlinear.

Next, the Park transformation is applied to convert the two-phase stationary α - β coordinate system into the rotating d - q coordinate system, aligning with the rotor magnetic field. In this coordinate transformation, the d -axis is always aligned with the rotor flux direction, and the q -axis is orthogonal to the d -axis. This ensures that the current components maintain a fixed relationship with the magnetic field. As shown in **Figure 11**, the current components I_α and I_β in the α - β coordinate system are projected onto the d - q coordinate system.

Based on the coordinate relationships shown in **Figure 12**, the Park transformation can be expressed as follows:

$$\begin{cases} I_d = I_\alpha \cos \theta + I_\beta \sin \theta \\ I_q = -I_\alpha \sin \theta + I_\beta \cos \theta \end{cases} \quad (2)$$

Where θ is the rotor electrical angle. Through this transformation, the d -axis current I_d and q -axis current I_q in the rotating d - q coordinate system are obtained, thereby achieving decoupling and simplification of motor control.

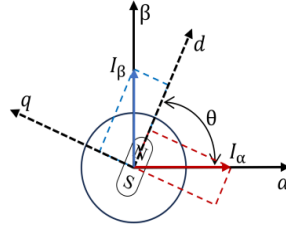


Figure 12. Park transform.

3.2. Space vector pulse width modulation

Through the inverse Park transformation, the AC reference voltage signals in the two-phase rotating d - q coordinate system, U_α and U_β , are obtained. These signals then need to be converted into three-phase voltages U_a , U_b , and U_c to drive the three-phase inverter circuit. Therefore, Space Vector Pulse Width Modulation (SVPWM) is employed to generate six high-frequency PWM signals. By controlling the on/off sequence of the six power MOSFETs, the desired three-phase sinusoidal AC voltages are produced [7].

In a three-phase inverter circuit, let the conduction of the upper switch in a half-bridge $Sx(x= a, b, c)$ correspond to state 1, and the conduction of the lower switch correspond to state 0. The upper and lower switches of the same half-bridge cannot conduct simultaneously. Based on this definition, the combination of the three-phase switch states can generate eight distinct basic voltage space vector states: U_{100} , U_{110} , U_{010} , U_{011} , U_{001} , U_{101} , U_{111} , U_{000} .

Taking one of the non-zero voltage vectors U_{100} for analysis, when the upper switch of phase A and the lower switches of phases B and C are conducting. Assuming equal phase resistances and a DC bus voltage of U_{dc} , the phase voltages U_{aN} , U_{bN} and U_{cN} relative to the neutral point N are as follows: $2/3U_{dc}$, $-1/3U_{dc}$ and $-1/3U_{dc}$. The synthesized expression of the three-phase symmetrical sinusoidal voltage vector is given in **Equation (3)**:

$$\vec{U}_{out} = \frac{2}{3} \left(U_{aN} + U_{bN} e^{j\frac{2\pi}{3}} + U_{cN} e^{j\frac{4\pi}{3}} \right) \quad (3)$$

After substituting U_{aN} , U_{bN} and U_{cN} into Equation (3), the resultant vector U_{100} can be obtained as:

$$U_{100} = \frac{2}{3} \left[\frac{2}{3} U_{dc} - \frac{1}{3} U_{dc} \left(\cos \frac{2\pi}{3} + j \sin \frac{2\pi}{3} \right) - \frac{1}{3} U_{dc} \left(\cos \frac{2\pi}{3} - j \sin \frac{2\pi}{3} \right) \right] = \frac{2}{3} U_{dc} \quad (4)$$

Similarly, the remaining non-zero voltage vectors can be calculated. In the α - β coordinate system, the six non-zero vectors have equal magnitudes of $2/3 U_{dc}$, and each adjacent non-zero vector is separated by 60° . The two zero vectors have zero magnitude and are located at the center. The endpoints of the six non-zero basic voltage vectors form a regular hexagon, dividing the plane into six sectors, as shown in **Figure 13**.

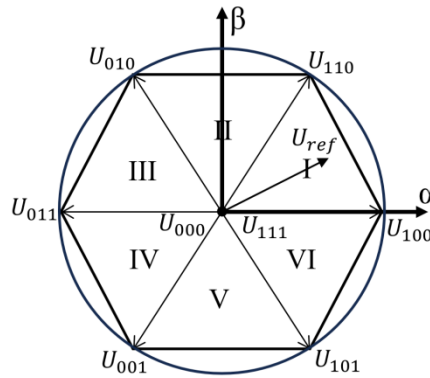


Figure 13. Voltage space vector diagram.

According to the principle of volt-second balance, any desired voltage vector U_{ref} in the α - β coordinate system, composed of U_α and U_β , can be synthesized by controlling the duration of the two adjacent non-zero basic vectors within the sector and the zero vectors. Let a PWM period be T_s , and the desired reference vector during this period be U_{ref} . Denote the two adjacent non-zero basic vectors of the sector as U_x and U_y , and the zero vector as U_0 . Then, according to the volt-second balance principle, the **Equation (5)** can be derived:

$$U_{ref} \cdot T_s = \vec{U}_x \cdot T_x + \vec{U}_y \cdot T_y + \vec{U}_0 \cdot T_0 \quad (5)$$

In summary, the SVPWM technique synthesizes a reference voltage vector U_{ref} from U_α and U_β obtained via the inverse Park transformation. The sector in which the reference vector is located is then determined. The dwell times of the two adjacent active vectors within the sector, along with the zero vectors, are appropriately allocated. Once the required dwell times of the adjacent active vectors and zero vectors are calculated, the PWM signals are generated. These PWM signals drive the three-phase inverter, ultimately achieving precise control of the brushless motor.

4. Results and discussion

To evaluate the performance of the BLDC motor drive and control system designed in this study, a position closed-loop test was conducted. This test also aimed to verify the FOC algorithm and assess the stability, accuracy, and real-time performance of the control system. The initial position was set to 0° . The motor was commanded to move to target positions of 90° , 180° , and 360° . Real-time data were recorded and analyzed using MATLAB.

As shown in **Figure 14** and **Figure 15**, when the motor moved from 0° to 90° , the system response time was 0.138s. The settling time was approximately 0.498s. After stabilization, the maximum error was about 0.25° . The

average error remained within 0.2° .

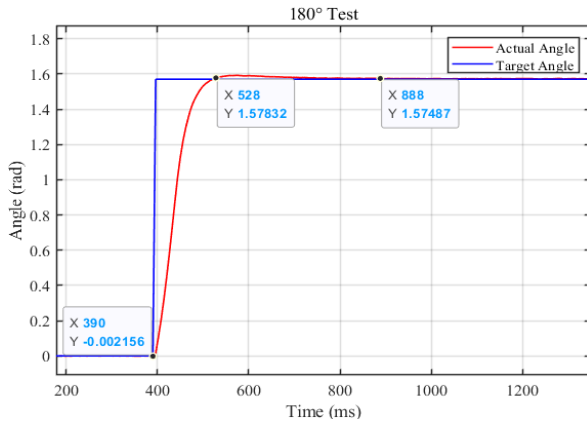


Figure 14. Response curve at 90° position.

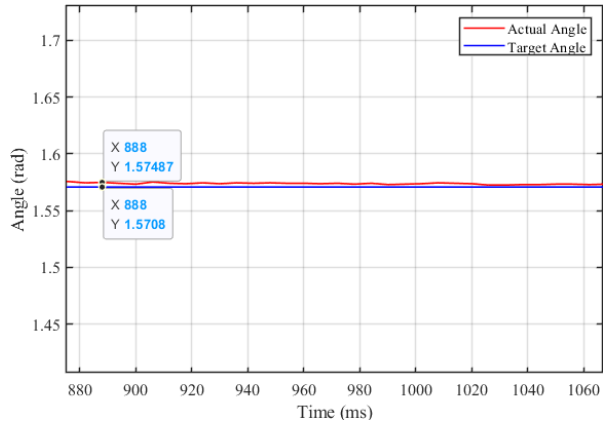


Figure 15. Position error curve at 90°

As shown in **Figure 16** and **Figure 17**, when the motor moved from 0° to 180° , the system response time is approximately 0.216s, the settling time is about 0.372s, the maximum steady-state error is around 0.18° , and the average error does not exceed 0.13° .

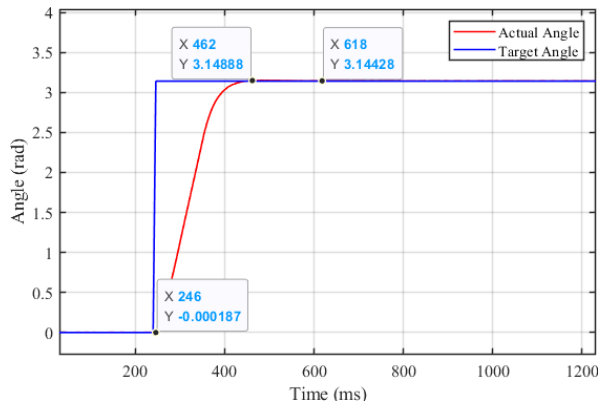


Figure 16. Response curve at 180° position.

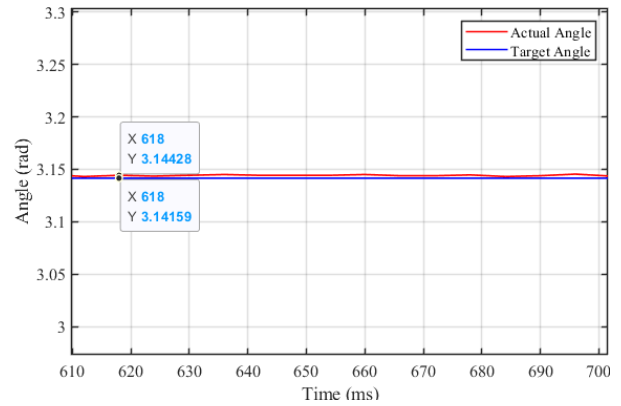


Figure 17. Position error curve at 180° .

As shown in **Figure 18** and **Figure 19**, when the motor moved from 0° to 360° , the system response time is approximately 0.378s. The maximum steady-state error is about 0.15° , and the average steady-state error is controlled within 0.12° .

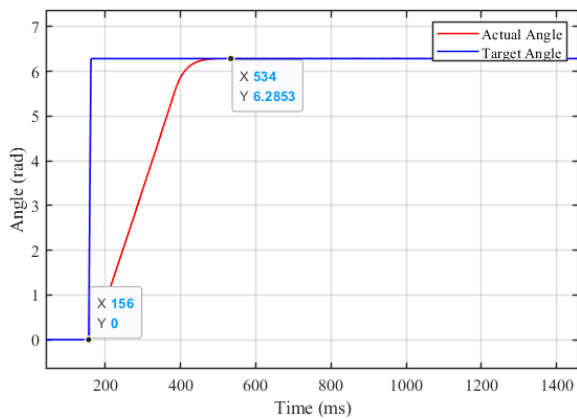


Figure 18. Response curve at 360° position.

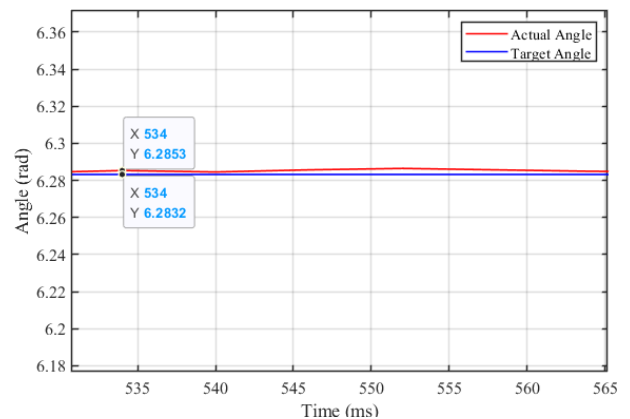


Figure 19. Position error curve at 360° .

The above experiments indicate that the system demonstrates superior dynamic response, with minimal steady-state position error. The maximum steady-state error does not exceed 0.25° , and the average steady-state error remains within 0.2° .

5. Conclusion

This study designs an integrated BLDC motor drive and control system for robotic arm joints, featuring high performance, high integration, and low cost. Experimental results show that the designed system exhibits excellent dynamic response and steady-state performance, meeting the requirements of desktop, lightweight robotic arms.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Ye S, Sun L, 2021, Design and Implementation of Intelligent Control Program for Six Axis Joint Robot. *Big Data Analytics for Cyber-Physical System in Smart City*, 180–186.
- [2] Sathuluri A, Sureshbabu A, Frank J, et al., 2023, Computational Systems Design of Low-Cost Lightweight Robots. *Robotics*, 12(4): 91.
- [3] Tingting W, 2023, Research on Optimized Control Method for Robotic Arm Joint Motors, thesis, Changchun University of Technology.
- [4] Le Q, 2024, Neural Network FOC Control of Robotic Arms Driven by Brushless Motors, thesis, Xi'an Shiyou University.
- [5] Li N, Li M, Zhang C, et al., 2024, Design of Brushless Motor Control System Based on FOC Algorithm. *International Conference on Mechatronic Engineering and Artificial Intelligence*, 130712T.1–130712T.7.
- [6] O'Rourke C, Qasim M, Overlin M, et al., 2019, A Geometric Interpretation of Reference Frames and Transformations: dq0, Clarke, and Park. *IEEE Transactions on Energy Conversion*, 34(4): 2070–2083.
- [7] Ma R, Guo X, Zhang Y, et al., 2023, PMSG Offshore Wind Power System Control Using SMC and ADRC with Fast SVPWM in Complicated Environment. *Electrical Engineering*, 105(5): 2751–2767.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research on Motion Simulation of Panda Manipulator Based on ROS2

Boru Wang*, Wei Liu

Jilin University of Chemical Technology, Ji'lin, Jilin, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Focusing on the research issues of path optimization and collision avoidance in robotic arm motion planning, this paper will construct a high-fidelity simulation environment using the ROS2 framework. By integrating MoveIt, Rviz visualization tool packages, and the Panda robotic arm URDF file, along with leveraging the DDS communication mechanism of ROS2, low-latency data interaction between the planning module and the simulation environment is achieved. The upper computer software is developed to conduct simulation studies on the path planning and trajectory interpolation principles of the robotic arm, thereby verifying the reliability of the ROS2 distributed architecture in robotic arm simulation.

Keywords: ROS2; Robotic arm; Path optimization; Trajectory interpolation

Online publication: February 12, 2026

1. Introduction

With the development of science and technology and the improvement of people's living standards, robotic arms are being increasingly widely applied in various fields, and users have also set higher performance requirements for them ^[1]. To meet the ever-more complex application demands, it is necessary to conduct various simulations during the development of the robotic arm body to save R&D costs and enhance R&D efficiency. Currently, the mainstream simulation platforms for developing robotic arms include Matlab, TeamBots, CARMEN, ROS, and so on. Among them, ROS2 is more widely utilized by universities, research institutions, and enterprises. ROS2 is the second-generation robot operating system specifically designed for modern robotic systems. It employs a DDS-based (Data Distribution Service) peer-to-peer (P2P) communication mechanism, supports task responses at the millisecond level, and has achieved significant improvements in real-time performance, security, cross-platform compatibility, and distributed architecture, making it an ideal choice for robot development ranging from consumer-grade to industrial-grade applications ^[2-4].

2. ROS system-related tools

2.1. Configuration of the robotic arm

Before controlling the robotic arm, it is necessary to use the MoveIt Setup Assistant to configure the robotic arm's description file in order to generate the required YAML files and various launch files for controlling its motion. This article utilizes the built-in Panda robotic arm URDF file provided by ROS. The file describes the parameters of the Panda robotic arm, such as the fact that it has seven joints, the installation positions of the motors on the links, and the motion angle and speed limitations of the joints, among other details. The steps for importing the URDF file into the MoveIt Setup Assistant to generate the configuration package are illustrated in **Figure 1**.

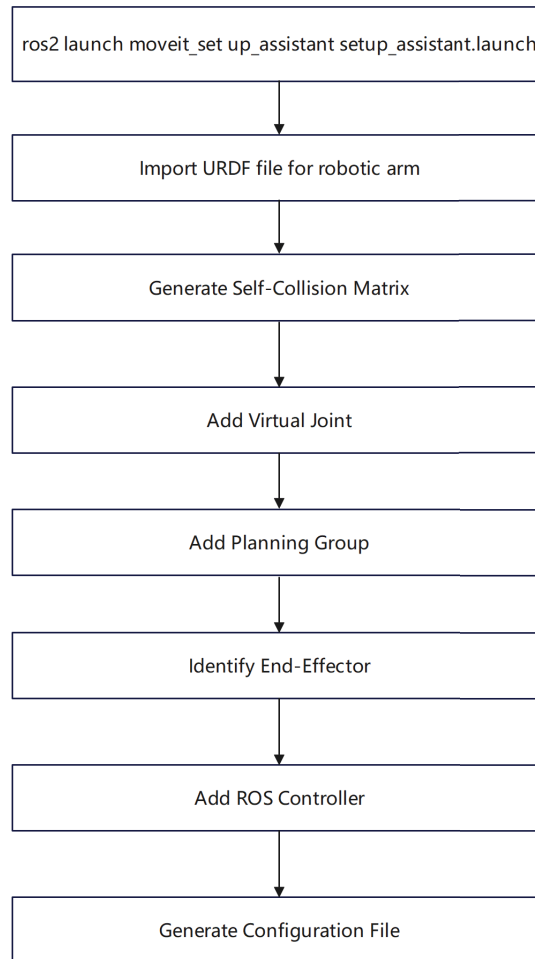


Figure 1. Configuration process for the panda robotic arm.

During the configuration process, the KDL algorithm library can be selected to solve the forward and inverse kinematics of the robotic arm. The generated configuration package includes a config folder and a launch folder. The config folder contains a series of YAML files that describe the initial position of the robotic arm, joint limits, kinematics algorithm type, as well as MoveIt controllers and ROS controllers ^[5]. The descriptive information in these files is closely related to the motion state of the robotic arm.

2.2. MoveIt system framework

MoveIt plays a highly significant role in the motion control process of robotic arms. Its functional block diagram

is illustrated in **Figure 2**. The `move_group` node retrieves the URDF and SRDF files of the robotic arm from the parameter server^[6]. It interacts with the robot controller via the `JointTrajectoryAction` interface. Meanwhile, users can configure MoveIt through the interfaces provided by MoveIt.

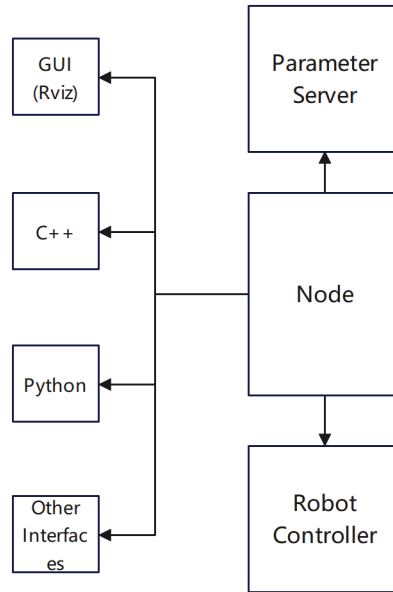


Figure 2. Block diagram of the MoveIt system.

2.3. Rviz visualization

Rviz is a 3D visualization tool capable of real-time displaying the joint states and motion trajectories of a robotic arm. Additionally, it allows for the control of the robotic arm through its built-in plugins^[7]. By running a launch file and observing the three-dimensional state of the robotic arm in Rviz, it can be seen that after running the demo file, the state of the robotic arm in Rviz is as shown in **Figure 3**, where all joint angles of the robotic arm are 0, and the Roll-Pitch-Yaw (RPY) angle values are 3.14, 0, and 0, respectively.

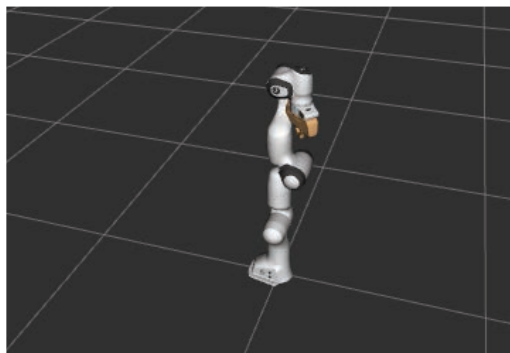


Figure 3. Panda robotic arm.

3. Route planning algorithm

3.1. Principles of the RRT algorithm and the RRT* algorithm

The RRT algorithm, whose full name is Rapidly-exploring Random Tree algorithm, is a sampling-based path planning algorithm^[8]. Its basic process is illustrated in **Figure 4**. It incrementally constructs a tree that extends

from the starting point to the target region, expanding one node at a time in an incremental manner. It gradually explores the space until the target point or target region is found. The experimental simulation is shown in **Figure 5**.

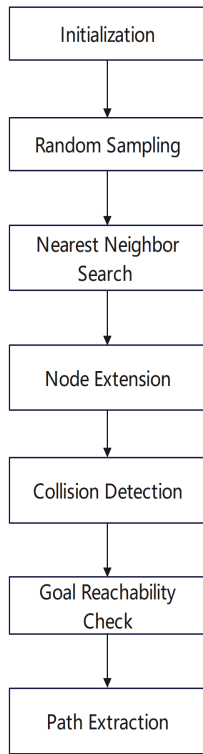


Figure 4. Flowchart of the RRT algorithm.

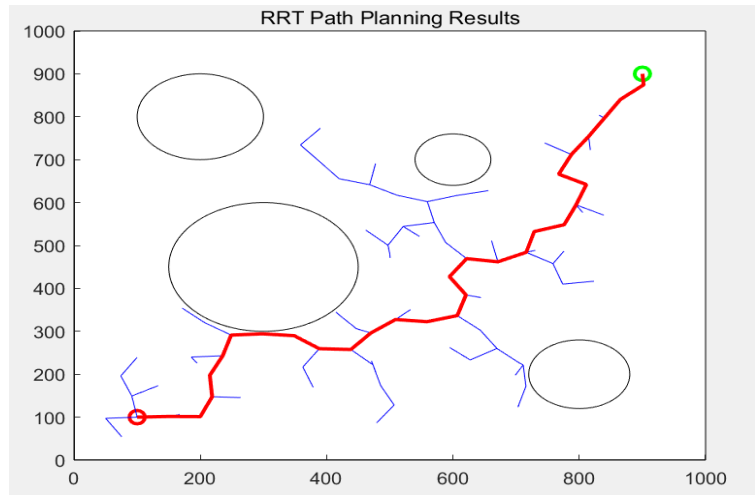


Figure 5. Simulation results of the RRT algorithm.

The RRT* algorithm is an improved version of the RRT algorithm, with its primary enhancement being the introduction of a path optimization mechanism. This mechanism ensures that a feasible path is found while further optimizing the path quality, such as by identifying shorter or more optimal paths ^[9]. It achieves this by reselecting parent nodes and rewiring the random tree, thereby reducing path costs. This is the main distinction between the RRT* algorithm and the RRT algorithm. The experimental simulation is illustrated in **Figure 6**.

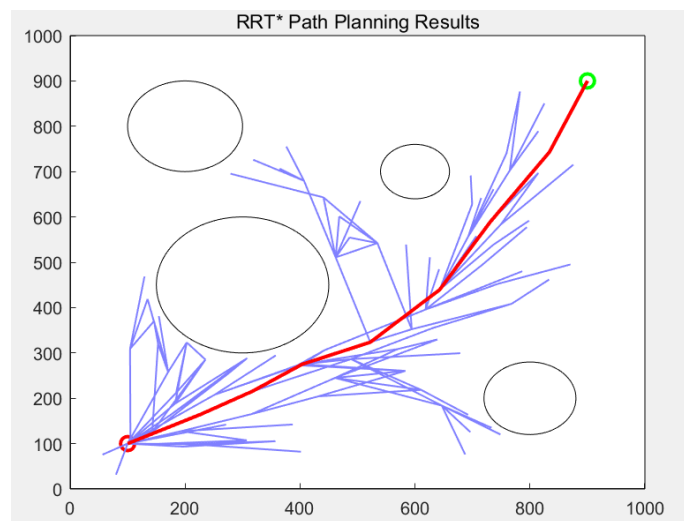


Figure 6. Simulation results of the RRT* algorithm.

3.2. Comparative testing of RRT and RRT* algorithms

On the upper computer, select the RRT and RRT* algorithms respectively, with the target point set to (0.5, 0.3, 0.4) for both. Click the “Add obstacles” button to introduce obstacles, where the length, width, and height of the obstacles are set to 0.5, 0.05, and 0.5, respectively, as shown in **Figure 7**. Then, set the upper limit for the planning time, click execute to conduct repeated experiments, and record the experimental data obtained, as depicted in **Figure 8**.

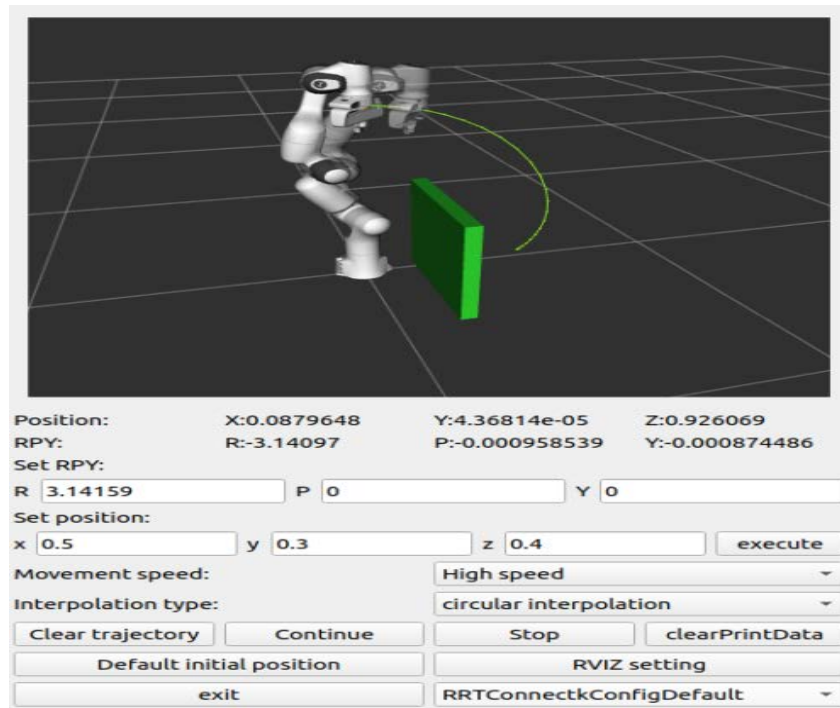


Figure 7. Upper computer interface.

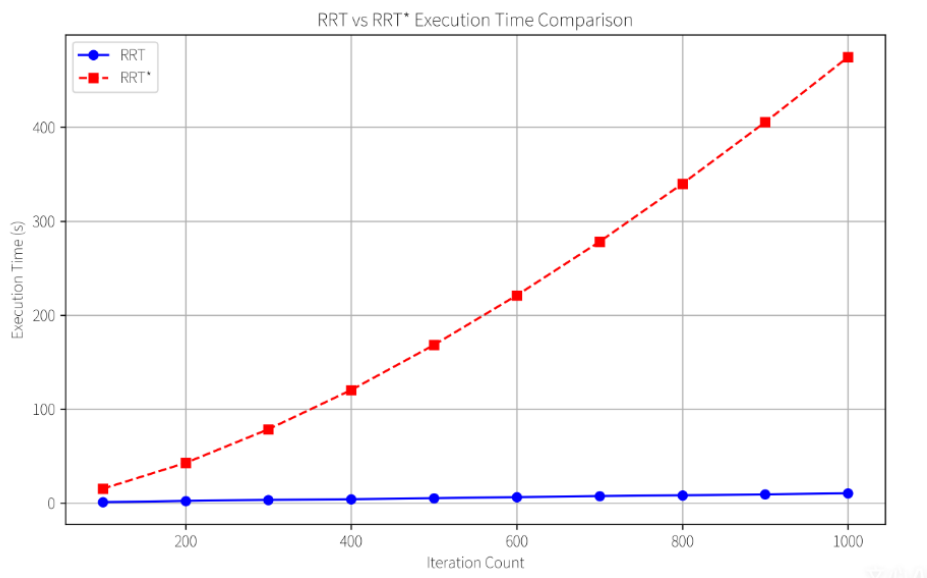


Figure 8. Comparative diagram of algorithm execution.

In the experiments, the RRT algorithm can always find a path from the starting point to the destination as long as it is given a sufficiently long planning time. However, due to its random sampling nature, the path it generates is often not optimal and exhibits low search efficiency. When the upper limit of the planning time is reduced, the RRT algorithm has a relatively high probability of failing to plan a path. In contrast, the RRT* algorithm demonstrates a significant improvement in search efficiency during the experiments, with a notably shorter planning time and relatively stable fluctuations in repeated planning durations.

4. Interpolation trajectory planning

When using the Configuration Assistant to generate a robotic arm configuration package, the RRT* algorithm from the OMPL motion planning library is selected as the path planning algorithm. After setting the destination coordinates, the planner will generate a path. However, this path only represents a route from the starting point to the destination and does not include the motion parameters for each joint of the robotic arm. Therefore, after the path is planned, it is necessary to perform interpolation on the path. Once the coordinates of each interpolation point are calculated, these coordinate data are transmitted to the inverse kinematics solver module to determine the velocity, acceleration, and other motion parameters of each joint at every point. Thus, calculating a series of suitable interpolation points is crucial for the smooth operation of the robotic arm^[10]. After setting the destination coordinates and RPY (Roll, Pitch, Yaw) angles in the upper computer software and clicking the execute button, the robotic arm will plan a path. To achieve smooth motion of the robotic arm's end-effector along the planned path, linear interpolation or circular interpolation is typically required.

4.1. Principles of linear interpolation

Given the starting point coordinates as (x_1, y_1, z_1) and the destination point coordinates as (x_2, y_2, z_2) , the straight-line distance between the two points in space is calculated as follows:

$$L = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (1)$$

The number of interpolation points is:

$$N = \begin{cases} \frac{L}{d}, \frac{L}{d} (Integer) \\ \text{int} \left(\frac{L}{d} \right), \frac{L}{d} (Non - integer) \end{cases} \quad (2)$$

Where t_n represents the timing step length, v is the end-effector's motion velocity, and t_n denotes the total timing interpolation duration. The average incremental displacement along the x, y, and z axes for each interpolation point, relative to the previous one, is given by:

$$\Delta x = (x_2 - x_1)/N \quad (3)$$

$$\Delta y = (y_2 - y_1)/N \quad (4)$$

$$\Delta z = (z_2 - z_1)/N \quad (5)$$

Therefore, the x, y, and z coordinate values of the interpolation point at any arbitrary (i+1) moment can be calculated as follows:

$$x_{i+1} = x_1 + i\Delta x \quad (6)$$

$$y_{i+1} = y_1 + i\Delta y \quad (7)$$

$$z_{i+1} = z_1 + i\Delta z \quad (8)$$

After obtaining the coordinate values of the interpolation points at any given moment, MoveIt will use the KDL solver to perform inverse kinematics calculations, deriving the angle for each joint. Subsequently, the robotic arm will move according to these calculated values. On the upper computer software, when Cartesian linear interpolation is selected and the joint point coordinates (0.5, 0.5, 0.5) are input, with RPY (Roll, Pitch, Yaw) set to 3.14, 0, and 0 respectively, and the execute button is clicked, the robotic arm will move to the target point. After reaching the target, if new coordinates (0.5, 0.1, 0.5) are input while maintaining the same orientation, the robotic arm will move along the planned linear trajectory. The motion pattern of the robotic arm is illustrated in **Figure 9**, with a total of 67 interpolation points and a movement duration of 6.7 seconds. As can be seen from the figure, the trajectory of the robotic arm consists of two straight lines, fulfilling the requirements of linear interpolation.

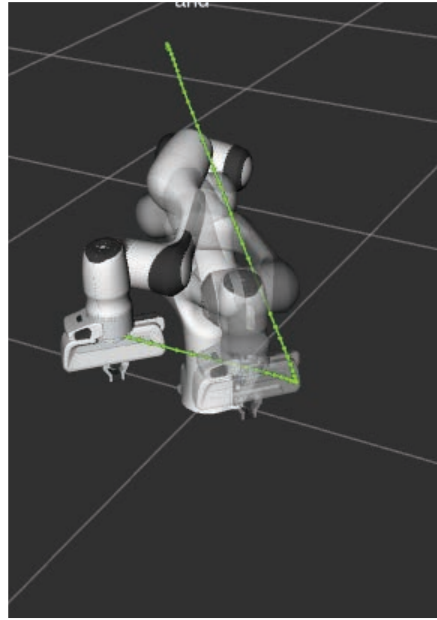


Figure 9. Linear interpolation of the robotic arm.

4.2. Circular interpolation

4.2.1. Principles of planar circular interpolation

A planar circular arc refers to an arc whose plane coincides with one of the three fundamental planes (e.g., the XOY plane). In this plane, given three non-collinear points A, B, and C, the circular arc formed by these points is illustrated in **Figure 10**.

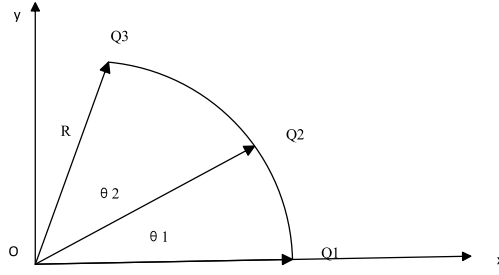


Figure 10. Planar circle defined by points Q1, Q2, and Q3.

R represents the radius of the planar circle, θ_1 and θ_2 are the central angles subtended by the arcs from Q1 to Q2 and from Q2 to Q3, respectively. Therefore, the total central angle is $\theta = \theta_1 + \theta_2$. Moreover, according to the chord length formula $L = 2R^* \sin \frac{\theta}{2}$, we have:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} = 2R^* \sin \frac{\theta_1}{2} \quad (9)$$

$$\sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2} = 2R^* \sin \frac{\theta_2}{2} \quad (10)$$

$$\theta_1 = 2\arcsin\left(\frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{2R}\right) \quad (11)$$

$$\theta_2 = 2\arcsin\left(\frac{\sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}}{2R}\right) \quad (12)$$

Given the angular displacement $\Delta\theta = t_s v / R$ within a fixed time t_s and the joint angular velocity v , the total number of interpolation steps can be calculated accordingly.

$$N = \text{int}(\theta / \Delta\theta + 1) \quad (13)$$

$\partial_i = i\Delta\theta$ represent the central angle from the starting point to the interpolation point at time instant i . Therefore, the coordinates of the interpolation point at time instant $i+1$ can be derived as follows:

$$y_{i+1} = R \sin(\partial_i + \Delta\theta) \quad (14)$$

$$x_{i+1} = R \cos(\partial_i + \Delta\theta) \quad (15)$$

4.2.2. Principles of spatial circular arc interpolation

Spatial circular arc interpolation is based on planar circular arc interpolation and can typically be carried out in three steps as follows:

- (1) Transform the three-dimensional spatial problem into a two-dimensional planar problem by identifying the plane determined by three non-collinear points, which serves as the plane where the planar circular arc lies;
- (2) Utilize the planar circular arc interpolation algorithm to calculate the coordinate values of each interpolation point within this plane;
- (3) Compute the transformation matrix between the coordinate system established by the planar circular arc

and the base coordinate system, and then convert the coordinates of each interpolation point into their corresponding values in the base coordinate system.

As shown in **Figure 11**, a plane and a circular arc with a known radius R can be determined by three non-collinear points P_1 , P_2 , and P_3 . Through establishing a coordinate system named $O_R X_R Y_R Z_R$ on this plane, making the origin of the coordinate system coincide with the center of the circle determined by points P_1 , P_2 , and P_3 , where Z_R is the outward normal to the $O_R X_R Y_R$ plane where the planar circular arc lies, planar circular arc interpolation theory can be applied to the circular arc defined by points P_1 , P_2 , and P_3 in the plane.

If the rotation angle between the coordinate axis Z_R of $O_R X_R Y_R Z_R$ and the base frame is ∂ , the angle between X_R and the X -axis is θ , Given that O_R in the original coordinate system are (x_0, y_0, z_0) , the rotation matrix for rotating the $O_R X_R Y_R Z_R$ by an angle about the Z -axis is:

$$R_{Z_R}(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (16)$$

The rotation matrix for rotation by ∂ is:

$$R_{X_R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \partial & -\sin \partial \\ 0 & \sin \partial & \cos \partial \end{bmatrix} \quad (17)$$

So, the overall rotation matrix is:

$$R = R_{Z_R} R_{X_R} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \partial & -\sin \partial \\ 0 & \sin \partial & \cos \partial \end{bmatrix} \quad (18)$$

Transformation matrix is as shown:

$$T = \begin{bmatrix} \cos \theta & -\sin \theta \cos \partial & \sin \theta \sin \partial & x_0 \\ \sin \theta & \cos \theta \cos \partial & -\cos \theta \sin \partial & y_0 \\ 0 & \sin \partial & \cos \partial & z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (19)$$

Therefore, the relationship for converting the coordinate values of the interpolation point P into the base coordinate values Q is as follows:

$$Q = TP \quad (20)$$

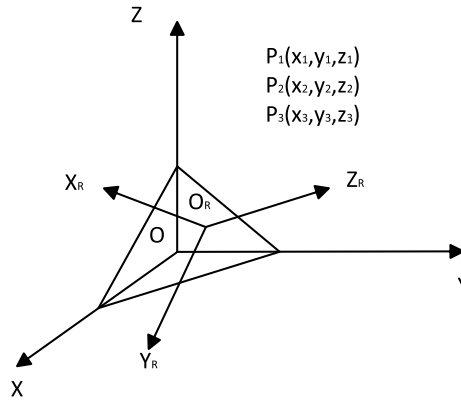


Figure 11. Base coordinate system and planar circular arc coordinate system.

On the upper-computer software, circular arc interpolation is selected. After inputting the coordinate values (0.5, 0.5, 0.5) and clicking the execute button, the robotic arm reaches the target point. Subsequently, another set of coordinates (0.5, 0.1, 0.5) is input, and the execute button is clicked again. The execution results, as shown in **Figure 12**, indicate that the number of interpolation points is 155, and the total time taken is 15.5 seconds. The interpolated trajectory consists of two arcs, meeting the requirements for circular arc interpolation.

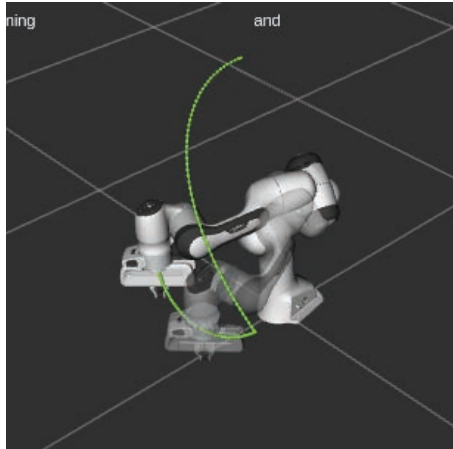


Figure 12. Arc trajectory interpolation.

Through simulation, it was found that when comparing the number of interpolation points and the time required for linear interpolation and circular arc interpolation under the same starting and ending coordinates, the number of interpolation points and the time for linear interpolation were significantly lower. However, during the experiments, it was also discovered that linear interpolation sometimes failed to reach the ending coordinates. Therefore, before connecting to the actual robotic arm, simulation can be utilized to observe in advance whether linear interpolation can reach the target point. If linear interpolation can reach the ending point, it can be given priority to reduce the number of interpolation points and calculation time. In other cases, circular arc interpolation should be used. During the operation of the robotic arm, the rqt tool can be opened, and the Visualization/Plot plugin under Plugins can be selected to observe the angular data of each joint of the robotic arm, as shown in **Figure 13**. It can be seen that after interpolation calculations are performed by MoveIt's internal algorithm, the angles of each joint of the robotic arm change smoothly.

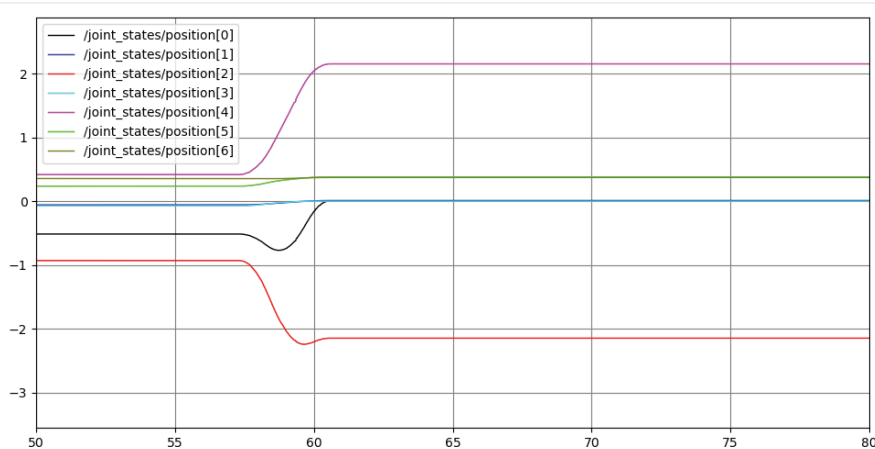


Figure 13. Joint angle curves.

5. Conclusion

This paper takes the Panda robotic arm as the research subject, displaying its status and controlling its motion through upper-computer software. By introducing obstacles into the planned scenarios, it compares the random sampling efficiency of the RRT algorithm and the RRT* algorithm in path planning for the robotic arm. Furthermore, it elaborates on the principles of linear and circular arc interpolation following path planning, as well as the importance of conducting simulations before connecting to the actual robotic arm. Through simulations, a smooth operation of each joint of the robotic arm was observed, laying a theoretical foundation for future research on robotic arms based on ROS2.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Cao H, Zhang X, Zhang Z, et al., 2022, Research Status and Trend Analysis of ROS-Based Real-Time Control Systems for Robotic Arms. *Computer Measurement & Control*, 30(3): 1–7.
- [2] Hu J, 2022, Obstacle Avoidance Research for Six-Axis Industrial Robots Based on ROS. *Science & Technology Information*, 20(14): 10–12.
- [3] Qu L, Gao K, Xing Y, et al., 2022, Research on Motion Planning for Robotic Arms based on ROS. *Machine Tool & Hydraulics*, 50(22): 43–47.
- [4] Zhao P, Hong R, Fang C, 2022, Design of a UR5 Robot Control System based on ROS. *Journal of Nanjing Tech University (Natural Science Edition)*, 44(2): 161–168.
- [5] Wang B, 2023, Research on Motion Planning for Six-Axis Robotic Arms based on ROS. *Computer Programming Skills & Maintenance*, 2023(3): 140–142.
- [6] Cheng L, Hua J, Hu J, 2022, Research on Trajectory Planning and Simulation for 6-DOF Robotic Arms based on ROS Platform. *Manufacturing Automation*, 44(1): 38–41.
- [7] Hui D, Chen G, Ma J, et al., 2023, Modeling and Motion Planning for Robotic Arms based on ROS. *Journal of Nanjing Institute of Technology (Natural Science Edition)*, 21(1): 52–58.
- [8] Jiang H, Chao Y, Zhou J, et al., 2023, Improved RRT Algorithm for Robotic Arm Path Planning. *Mechanical Design and Manufacturing*, 2023(12): 288–292.
- [9] Zhang B, 2022, Research on Motion Control of a 6-DOF Robotic Arm based on ROS, thesis, Jilin Institute of Chemical Technology.
- [10] Zhang S, 2021, Trajectory Planning and Grasping for Robotic Arms based on ROS System, thesis, Dalian University of Technology.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

Technology of Radar Detection

Rui Hu, Lei Pan*, Jing Gao, Haojie Hu, Fang He

Rocket Force University of Engineering, Xi'an 710025, Shaanxi, China

*Corresponding author: Lei Pan, pl_528@163.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Radar detection technology utilizes radio waves for target detection, localization, and identification. It involves emitting electromagnetic waves and receiving the reflected echoes from targets, then analyzing the echo characteristics to obtain target information. This paper focuses on the fundamental principles, advantages, and disadvantages of radar detection technology. It emphasizes synthetic aperture radar (SAR), passive radar detection technology seeker, and millimeter-wave active homing guidance target identification technology, as well as the characteristics and development status of other detection methods such as phased array radar, showcasing the multi-directional development trend of radar detection technology.

Keywords: Radar detection technology; Synthetic aperture radar (SAR); Passive radar detection technology; Millimeter-wave active homing guidance target identification technology; Development status

Online publication: February 12, 2026

1. Introduction

Radar is defined as radio detection and ranging, which measures the position of targets in the air, on the ground, and on the water, and is also called radio positioning. Radar uses a directional antenna to emit radio waves into the air. When the radio waves encounter a target, they are reflected back and received by the radar. The distance data of the target is obtained by measuring the time experienced by the radio waves propagating in the air, and the angle data of the target is determined according to the antenna beam pointing^[1]. The prominent advantages of radar are its large coverage, long operating range, and strong ability to penetrate haze. The disadvantages are that active detection easily exposes itself, making it susceptible to detection and interference by the enemy, low resolution, and vulnerability to anti-radiation missile attacks.

2. Development analysis of synthetic aperture radar (SAR)

Synthetic aperture radar (SAR) can realistically display the shape, size, motion state, and attitude of targets, breaking through the limitations of the original radar, which could only obtain four-dimensional information of

the target's range, azimuth, pitch, and velocity. Airborne SAR is an interdisciplinary field of SAR and precision-guided research. Theoretically, SAR has the ability to perform high-resolution imaging of any area at any time, and it has a long operating range. Its operation is not limited by climatic conditions and solar illumination, and it can penetrate vegetation and surface layers to detect hidden targets ^[2]. Therefore, the active radar seeker using SAR technology can effectively improve its all-weather and all-time detection capability and accuracy. The application directions of airborne SAR technology are as follows in **Figure 1**.

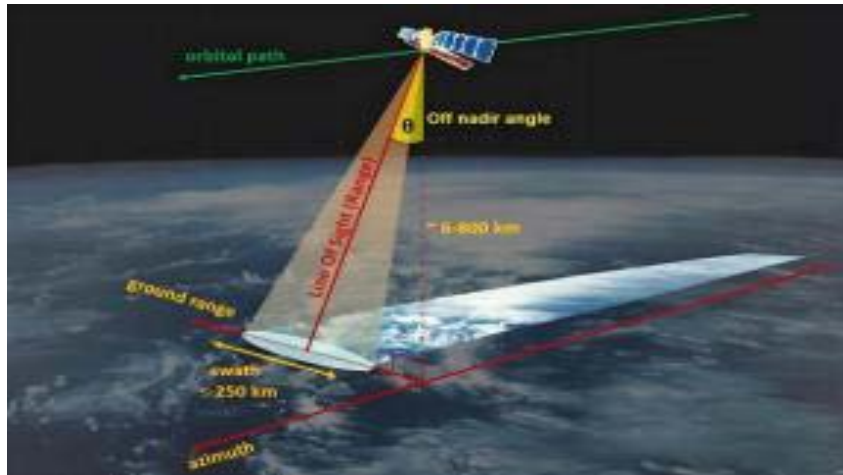


Figure 1. Synthetic aperture radar (SAR).

2.1. Correcting inertial navigation errors

Inertial navigation system (INS) is a self-contained navigation method that has no optical or electrical connection with the outside world. It has good concealment and is not restricted by meteorological conditions. Therefore, it has become a major guidance method widely used in medium- and long-range missiles. Its inherent defects are that the navigation error accumulates with time, and long-term operation will bring large accumulated errors. Using missile-borne SAR for scene matching to correct INS errors is an effective measure to improve guidance accuracy. Real-time images containing typical terrain features are first obtained by the missile-borne SAR and matched with the reference map pre-stored in the electronic map database to obtain the position information of several points in the scene. Then, the missile body position is calculated according to the relative position relationship between the missile and the scene, and the position error of the INS is further corrected. The compensated error can reach the P-code accuracy of GPS. If the data rate is high enough, the velocity error of the INS can also be corrected.

2.2. Striking time-sensitive targets

Improving the capability to engage time-sensitive targets (TSTs) such as mobile missile launchers and mobile air defense guidance radars is an important development direction for precision-guided missiles. For such targets, the method of determining their location and then attacking cannot be adopted. The area where the target may be located must be searched in the mid-course phase of the autonomous guidance, and after the target is acquired and confirmed, the missile transitions to homing guidance for precision strike. The main problems faced are the influence of strong ground clutter on target detection and the accurate identification of the target. Target strikes in **Figure 2** demonstrate.



Figure 2. Time-sensitive target strike.

The adoption of SAR technology improves the resolution in the azimuth direction. The reduction of azimuth resolution units will greatly improve the signal-to-clutter ratio (SCR), thereby increasing the detection probability and high-resolution imaging of the target can effectively improve the recognition probability. For instance, conduct a large-scope imaging in scanning working patterns such as Doppler Beam Sharpening (DBS), and take advantage of the obtained low-resolution image to detect targets ^[3]. If a target is detected, perform forward squint high-resolution imaging on the destination area to identify whether it is a target of interests, and after confirming, switch to single pulse tracking for exploitation.

2.3. Attack point selection

When there are multiple targets in the range of the terminal guidance radar, the performance of the target recognition and tracking system will be greatly affected if the target cannot be identified effectively and the real target cannot be selected correctly, as the selection of attack point is a difficult technical problem for the seeker. The missile-borne SAR is used to image the formation target group (**Figure 3**). According to the extracted target characteristics and the law of Tactical Formation mode, the selection ability of guidance radar to valuable targets can be effectively improved. The size, shape and other features are extracted from the high-resolution SAR image of the ship target. Combined with the relative position information of the strong scattering points such as the bridge, the key parts of the ship target can be selected.

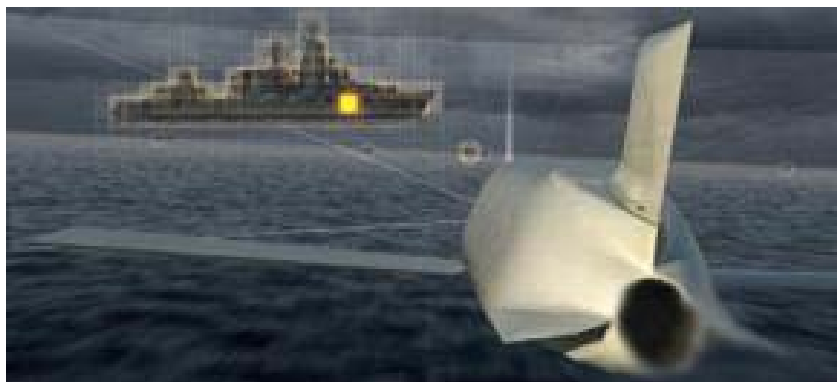


Figure 3. The anti-ship missile chooses the ship's vital position to strike.

2.4. Damage assessment

According to the target image transmitted by the missile through the data link, planners can evaluate the damage of

the target, judge the damage effect of the missile on the target, and decide whether to plan the subsequent missile to continue attacking the target, or turn to attack other targets. This instantaneous damage assessment capability reduces the number of missiles needed for a given attack mission. Based on the high-resolution missile-borne SAR image, the geometric shape features and internal structure features of the man-made target are extracted. If the degree of feature change before and after striking reflects the degree of target destruction. The grade evaluation method can be used to evaluate whether the man-made target is destroyed, the degree of destruction and the destroyed part.

3. Development analysis of passive radar detection technology

The passive radar seeker is the key component of the anti-radiation missile (ARM), known as the “Eye” of the ARM, whose main function is to complete the sorting, interception and tracking of the radiation source, its technical performance will directly affect the operational performance of ARM. The passive radar seeker should have a wide frequency band to cover most of the enemy’s radars, and the seeker should have high sensitivity to ensure that the missile can attack the target from both the main lobe and the side lobe of the radar. The seeker must possess high-angle measurement accuracy and angular resolution capability to enable target search, acquisition, tracking, as well as anti-jamming functionalities. The working principle of mono-pulse receiving technology seeker is that the target radiation signal received by the seeker antenna is sent to the signal processing system by the receiver, and the signal is sorted and identified.

Once it is determined as the target signal, the target is intercepted immediately, the angle error signal is extracted, and the error signal is sent to the Torquer of the gimbal after processing, so that the antenna can always point to the target and realize angle tracking. After the angle tracking of the Seeker, the command device in the signal processing system is connected to the autopilot, and the flight control system continuously corrects the route to the target according to the predetermined guidance law.

The early 1960s was the initial stage of the development of passive radar seeker. Due to technological constraints, first-generation ARM seekers featured narrow frequency bands, could only track the main lobe of target radars, and exhibited poor anti-jamming performance. Second-generation ARM seekers employed broader frequency bands with improved receiver sensitivity, enabling engagement of multiple target types. However, their high cost resulted in suboptimal cost-effectiveness ratios, incompatibility with small aircraft, and relatively low launch rates, rendering them inadequate for increasingly complex battlefield electromagnetic environments. Building upon previous generations, third-generation ARM seekers incorporate advanced microwave and signal processing technologies to expand frequency coverage, enhance receiver sensitivity, and strengthen signal processing capabilities^[4]. Through composite guidance, they achieve improved autonomous homing and anti-jamming performance, supporting multiple operational modes including self-defense, random, pre-programmed, and known/unknown range engagements.

The representative models of the third generation of direct attack on ARM are the AGM-88(HARM) of the United States (**Figure 4**), the X-31 of Russia, the ARF of France, the ALARM of Britain, the ARMIGER of Germany, etc., another type of patrol ARM is represented by Israel’s “Star-1”, South Africa’s “Lark”, the United States AGM- 136, Germany’s Dar and so on.



Figure 4. Agm-88 anti-radiation missile.

4. Research and analysis on millimeter wave active homing guidance target identification technology

ATR technology is an important indicator of the intelligentization level of precision-guided weapons. Millimeter wave technology is a supporting technology for the development of radar weapon systems towards intelligentization. In the millimeter wave band, the improvement of detection accuracy and the miniaturization of missiles enable millimeter wave precision-guided weapons to strike critical parts of targets. From the perspective of information processing, the difference between millimeter wave radar and microwave radar lies in the different amount of target information provided by the two. In the millimeter wave band, due to the improvement of resolution, the amount of information provided increases by orders of magnitude, making intelligent information processing functions such as target discrimination, target identification, interference discrimination, and target vulnerable area identification more realistic radar functions.

Considerable achievements have been made internationally in the research of millimeter-wave active homing guidance target identification technology and its related fields. However, there are still many challenging problems that need to be solved urgently in its development. These problems can be summarized as follows.

4.1. The identification algorithm is required to have the adaptability to arbitrary attitude angles of the target

Taking missile ground attack as an example, due to the complexity of terrain, the pose of ground vehicle targets relative to the radar seeker is random and cannot be predetermined through prior tracking. Therefore, it is necessary to design algorithms capable of omnidirectional matching and recognition^[5]. Currently, the widely adopted approach primarily relies on template matching. While this method can address target feature matching under different pose angles, it requires a substantial number of templates. Reducing feature dimensionality and compressing the template library serve as effective strategies. Furthermore, variations in target features also lead to recognition difficulties, which impose additional modeling burdens for template-based recognition methods and consequently result in decreased recognition rates.

4.2. The recognition algorithm is required to have the adaptability to the clutter background

Given the complex ground background conditions characterized by high false alarm rates and numerous false targets, it is essential to develop flexible and practical algorithms to separate targets from the background and discriminate against background-induced false targets^[6]. As noted by Indian scholar Mhamodi *et al.* in their study

on millimeter-wave seeker signal processing, when integration time is constrained by real-time requirements, the approach for enhancing weak signal detection in strong ground clutter environments should adopt relaxed false alarm criteria. The high false alarms generated during detection can subsequently be eliminated through target recognition algorithms, which simultaneously increases the difficulty of target identification and underscores its critical importance.

4.3. The target recognition algorithm is required to be realizable

Owing to the fact that the recognition algorithm is loaded on the seeker information processor, the processor is required to have fast processing speed, small data storage space and small hardware structure, therefore, modern advanced digital signal processing technology and parallel processing technology must be adopted.

4.4. The collection, analysis, modeling and simulation of target and background data require a large amount of work and a long development cycle

Due to the diversity of the background environment and the variability of the target attitude, a large target and background echo database is needed to train, test and improve various algorithms of target recognition and evaluate and select the best, many experiments are needed to finalize the algorithm and processor structure.

5. Development of other radar detection methods

5.1. Laser synthetic aperture radar (SAL)

Laser synthetic aperture radar (SAL) combines optical coherence detection technology and synthetic aperture imaging technology, and its weak signal detection ability can reach the order of photons, and the imaging resolution can break through the diffraction aperture limit of the telescope, high-resolution images can be obtained regardless of the detection range. At present, airborne SAL detection imaging experiments have been completed at home and abroad.

5.2. Phased array radar

In phased array radar seekers, the active phased array antenna replaces the traditional radar seeker's antenna, mechanical seeker platform, and transmitter. Fundamental changes have occurred in aspects such as antenna configuration, scanning method, platform stability, and structural dimensions, providing the means for the seeker to achieve new performance capabilities. The phased array radar seeker boasts multiple technical advantages, including high power density, rapid electrical scanning for tracking, multi-target information extraction, space-time adaptive processing (STAP), adaptive anti-jamming capabilities, compact size, and high reliability^[7]. It represents the development direction for precision-guided radar seekers and has become a subject of common interest worldwide.

The phased array radar seeker is basically in the active test and development stage, and only a few countries are successful in the research, such as the United States, Russia, Britain, Germany, Japan, etc., in this regard, the United States has been at the forefront of the world.

5.3. Terahertz radar

Ultra-narrow antenna beam can be easily realized in terahertz band. Under the same antenna size, the beam of terahertz wave is much narrower than that of millimeter wave and microwave, which can obtain higher antenna

gain and angle tracking accuracy. Compared with microwave radar and infrared detector, the application of terahertz technology to radar will bring many advantages, such as high range resolution, strong penetration, low interception rate and strong anti-jamming ability, it is one of the main application directions of terahertz technology. Throughout the development process and current progress level of terahertz radar system, it is still in the stage of demonstration and demonstration of experimental system.

5.4. Quantum radar

In missile weapon operation, quantum radar can effectively counter active deception jamming and achieve high-resolution anti-stealth detection, which will have a subversive impact on anti-stealth air defense operation. In 2012, the University of Rochester successfully demonstrated the application of quantum information technology to enable radar to effectively detect stealthy targets with deception capabilities^[8]. In 2015, the University of York in the United Kingdom developed a dual-cavity converter using a nano-oscillator to realize microwave and optical wave coupling, which can be used as a core device for future quantum radaQuantum Radar (**Figure 5**). In 2017, the 14th Institute of China Electric Power Science and Technology released public information that its quantum radar prototype could already detect targets hundreds of kilometers away. If finalized, it will represent China's leading position in the world in this technology. In the future, quantum radar technology will be applied to various platforms such as missile-borne and airborne.



Figure 5. Quantum radar.

6. Conclusion

Radar detection technology has become a pivotal detection means in military, civilian, and scientific research fields, evolving from traditional radio positioning into a multidimensional, high-precision integrated system. Among these advancements, SAR significantly enhances target recognition and guidance accuracy through high-resolution imaging, strengthening all-weather operational capabilities. Passive radar, centered on ARM, utilizes wide-frequency band and high-sensitivity seekers to effectively counter complex electromagnetic environments. Millimeter-wave active homing guidance technology addresses challenges such as target posture adaptability and background interference via high resolution and intelligent algorithms, driving the development of precision guidance toward greater intelligence. Emerging technologies like phased array radar, terahertz radar, and quantum radar further expand capabilities in anti-stealth, anti-jamming, and extreme environment detection. Currently, radar technology still faces challenges in real-time processing, miniaturized integration, and environmental adaptability. Looking ahead, the integration of artificial intelligence and quantum sensing technologies will steer radar systems toward higher precision, stronger anti-jamming capabilities, and multifunctional integration, providing more robust support for both defense and civilian applications.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Skolnik M, 2019, Radar Handbook, 4th ed, McGraw-Hill, New York.
- [2] Currie N, Brown M, 2020, Synthetic Aperture Radar: Systems and Signal Processing, Wiley.
- [3] Martone M, Rizzoli P, Villano M, 2024, Advanced Bistatic and Multistatic SAR Imaging: Recent Results from the TanDEM-X and TanDEM-L Missions. *IEEE Transactions on Geoscience and Remote Sensing*, 2024(62): 1–17.
- [4] Chen Z, Li H, Wang J, 2024, Metasurface-Based Radar Cross-Section Reduction and Beam Steering for Stealth Applications. *IEEE Antennas and Wireless Propagation Letters*, 2024(23): 456–460.
- [5] Zhang Y, Liu B, Sun X, 2024, A Multi-Scale Attention Network for SAR Target Recognition with Limited Samples. *Remote Sensing*, 16(5): 887.
- [6] Mhamodi A, 2018, Robust Target Detection in Millimeter-Wave Radar Under Clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 54(5): 2141–2153.
- [7] Zhang G, 2022, Phased Array Radar Technology, Publishing House of Electronics Industry, Beijing.
- [8] Shi X, Zhang W, Liu Y, 2024, Progress in Quantum Radar: Towards Practical Quantum Illumination with Microwave Photonics. *IEEE Journal of Selected Topics in Quantum Electronics*, 30(2): 3800309.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

An Evolutionary Game-Based Dynamic Signal Control Framework for Oversaturated Urban Networks

Weibin Zhao¹, Xinhai Xia^{2*}

¹School of Future Transportation, Guangzhou Maritime University, Guangzhou, China

²Guangdong-Hong Kong-Macao Greater Bay Area Shipping Research Institute, Guangzhou Maritime University, Guangzhou, China

**Corresponding author: Xinhai Xia, xiaxinhai@126.com*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Urban road networks frequently operate in an oversaturated state during peak hours, where traditional traffic signal control strategies, predominantly grounded in the assumption of fully rational user behavior, fail to capture the bounded rationality inherent in drivers' route choice decisions under congestion. To address this gap, this paper proposed a novel integrated framework that couples evolutionary game theory (EGT) with dynamic signal control, leveraging the Macroscopic Fundamental Diagram (MFD) for real-time feedback between network-wide traffic states and individual decision-making. Specifically, we model drivers within a control zone as a population choosing between two bounded-rational strategies: "waiting straight" versus "detouring". A replicator dynamics model governs the evolution of strategy adoption, with payoffs dynamically modulated by the MFD to reflect congestion-dependent travel costs. This behavioral layer is embedded within a receding horizon control (RHC) architecture that optimizes green splits and cycle lengths in real time to minimize total zone-wide delay, solved via Particle Swarm Optimization (PSO). Extensive simulations were conducted on a 6×6 grid network in SUMO under high-demand conditions (network saturation, approx. 0.92). Results demonstrate that the proposed method reduces average vehicle delay by 18.7% (from 142.8 s to 116.8 s), decreases queue spillback occurrences by 32.4%, and achieves convergence to an evolutionarily stable state (ESS) within 25 minutes, outperforming fixed-time, adaptive MAXBAND, and multi-agent deep reinforcement learning (MADDPG) baselines. This work establishes a closed-loop paradigm for behavior-aware, state-responsive traffic management in severely congested urban environments.

Keywords: Oversaturated traffic; Evolutionary game theory; Dynamic signal control; Macroscopic fundamental diagram (MFD); Receding horizon control

Online publication: February 27, 2026

1. Introduction

Urban mobility in megacities is increasingly challenged by chronic oversaturation during peak periods. According to the 2024 China Urban Mobility Report, core districts in cities like Beijing and Shanghai experience network-wide saturation levels exceeding 0.9 during morning and evening peaks, with intersection queue spillback rates surpassing 35% and average speeds dropping below 20 km/h^[1]. Under such conditions, the fundamental assumption of classical traffic assignment models, that travelers act as perfectly rational agents minimizing personal travel time, becomes untenable^[2].

Instead, empirical studies confirm that drivers exhibit bounded rationality: their decisions are shaped by limited information, cognitive heuristics, and social imitation rather than global optimization^[3]. For instance, in heavily congested corridors, some drivers persistently wait in long queues due to familiarity or perceived reliability, while others impulsively detour onto parallel streets, often exacerbating congestion elsewhere, a phenomenon rarely captured by conventional control systems^[4].

Current traffic signal control paradigms suffer from two critical limitations. First, fixed-time and coordinated systems optimize signal parameters offline or with slow adaptation, ignoring the dynamic feedback between signal settings and evolving driver behavior^[5]. Second, while adaptive methods such as deep reinforcement learning (DRL) offer real-time responsiveness, they typically treat drivers as passive flow entities or assume full rationality, thereby neglecting the strategic interaction and imitation-driven evolution of route choices^[6].

Evolutionary game theory (EGT) provides a natural framework for modeling such boundedly rational populations. Originating from biology, EGT describes how successful strategies propagate through imitation in a population of non-optimizing individuals, eventually converging to an evolutionarily stable strategy (ESS), a state resistant to invasion by alternative strategies^[7]. When coupled with the Macroscopic Fundamental Diagram (MFD), a well-established empirical relationship between network-wide vehicle density and flow, EGT enables dynamic calibration of strategy payoffs based on real-time congestion levels^[8].

Despite promising early efforts, existing studies either decouple behavioral modeling from signal control or apply static signal plans despite dynamic user responses^[9]. Crucially, a closed-loop, real-time framework that integrates (1) EGT-based behavioral evolution, (2) MFD-mediated state feedback, and (3) online signal optimization remains absent—particularly for highly oversaturated networks (saturation > 0.9), where traditional equilibrium assumptions break down.

To bridge this gap, this study proposed a three-tier collaborative control framework:

(1) Behavioral layer

Drivers choose between “wait” and “detour” strategies; proportions evolve via replicator dynamics.

(2) State feedback layer

MFD quantifies network congestion and dynamically adjusts strategy payoffs.

(3) Control layer

A receding horizon controller minimizes total delay by optimizing signal timing in real time.

The main contributions are: MFD-modulated replicator dynamics model that captures congestion-dependent payoff adjustments; A real-time RHC signal optimization scheme solvable via PSO, enabling online implementation; Empirical validation on a high-saturation grid network, demonstrating significant improvements over state-of-the-art baselines.

2. Methodology

2.1. Modeling assumptions

In the research field of urban traffic flow modeling and management, constructing a network scenario that conforms to the actual characteristics of traffic operation serves as the fundamental premise for accurately analyzing traffic behaviors and optimizing signal control strategies. The homogeneous urban subnetwork focused on in this study, by virtue of its clear and reasonable core assumptions, provides a reliable experimental platform for exploring drivers' route choice behaviors (such as "Wait" and "Detour" strategies) and the application value of the Macroscopic Fundamental Diagram (MFD) in traffic management decision-making. These assumptions are not only mathematically operable but also highly consistent with the traffic operation rules of specific urban areas, thus laying a solid foundation for the subsequent model derivation and result validation.

First and foremost, the network topology and infrastructure parameters are clearly defined to ensure the homogeneity and comparability of the research object. Specifically, the subnetwork is designed as a regular 6×6 grid, which contains 36 intersections in total. Each intersection is equipped with a four-phase signal control system, a widely adopted signal mode in urban areas that can effectively separate conflicting traffic flows (such as straight and left-turning vehicles) and reduce the risk of traffic accidents. In terms of link attributes, each road segment has a uniform length of 200 meters and is configured with 4 lanes in each direction. This setting eliminates the interference of heterogeneous infrastructure (such as uneven road lengths or varying lane numbers) on traffic flow distribution, making it possible to focus on the impact of drivers' decision-making behaviors on network performance.

In terms of driver behavior, two typical route choice strategies are defined to simulate the bounded rational decision-making process in real traffic scenarios. Strategy A, referred to as the "Wait" strategy, means that when a driver encounters a queue at the current intersection, they choose to stay in the queue and wait for the signal phase to pass. Strategy B, known as the "Detour" strategy, involves the driver diverting to an adjacent parallel link to avoid the current queue, which will increase the total trip length by $\Delta L = 200$ meters. This 200-meter detour distance is not arbitrarily set; it corresponds to the length of a single link in the grid network, ensuring that the detour cost (in terms of distance and time) is quantifiable and comparable. More importantly, this setting reflects the bounded rationality of driver, unlike the "fully rational" assumption in traditional traffic assignment models, drivers in this subnetwork cannot obtain global optimal path information (such as the traffic status of all links in the entire network). Instead, they can only rely on real-time network-average density data (which is easily accessible through navigation apps or traffic information platforms in practice) to make decisions, which is highly consistent with the actual decision-making characteristics of drivers summarized by Di & Liu (2016) in their research.

From the perspective of network traffic flow characteristics, the subnetwork is assumed to exhibit a well-defined, unimodal Macroscopic Fundamental Diagram (MFD). The MFD is a key tool in macroscopic traffic flow theory, which describes the stable functional relationship between network-average flow, density, and speed. The "well-defined" and "unimodal" attributes mean that the MFD of the subnetwork has a clear peak flow (i.e., capacity), and there is a one-to-one correspondence between density and flow in the free-flow and congested flow stages, without the ambiguity caused by heterogeneous traffic conditions. This assumption is not only theoretically feasible but also has been verified through offline calibration methods proposed by Daganzo (2007). Offline calibration involves using historical traffic data (such as probe vehicle data or loop detector data) to fit the MFD curve of the subnetwork, ensuring that the assumed MFD is consistent with the actual traffic flow operation laws

of the network.

Finally, it is necessary to emphasize the reasonable adaptability of these assumptions to specific urban scenarios. In fact, the homogeneous grid subnetwork constructed by the above assumptions is highly consistent with the traffic characteristics of Central Business Districts (CBDs) in most cities. On one hand, CBDs usually have a relatively regular road network topology, with grid roads crisscrossing and forming a relatively uniform spatial structure, which matches the 6×6 regular grid assumption of the subnetwork. On the other hand, CBDs are typically equipped with dense probe data coverage, with a large number of taxis, ride-hailing vehicles, and private cars equipped with GPS positioning devices, real-time traffic data such as network-average density can be efficiently collected and released, which provides a data basis for drivers to obtain decision-making information and also supports the application of MFD-based traffic management strategies. Therefore, the assumptions of this homogeneous urban subnetwork not only simplify the research problem but also effectively capture the core characteristics of CBD traffic scenarios, making the research results have strong practical application value.

2.2. Evolutionary game model

Let $x(t) \in [0,1]$ denote the proportion of drivers adopting Strategy A at time t . Payoffs are defined as negative travel times:

$$u_A(t) = -T_A(t), u_B(t) = -T_B(t)$$

Travel times are computed as:

Strategy A:

$$T_A(t) = \frac{n_A(t) \cdot l_v}{q_{\text{sat}}} + t_{\text{signal},A}(t) + \frac{L}{v_0}$$

where $l_v = 7.5$ m, $q_{\text{sat}} = 1800$ pcu/h/ln, $v_0 = 40$ km/h.

Strategy B:

$$T_B(t) = \frac{L + \Delta L}{v_B(\rho_B(t))} + t_{\text{signal},B}(t), v_B(\rho_B) = \frac{Q(\rho_B)}{\rho_B}$$

To account for network-wide congestion effects, we introduce an MFD-based attenuation factor:

$$\alpha(\rho(t)) = \begin{cases} 1, & \rho(t) \leq \rho_c \\ 1 - \lambda(\rho(t) - \rho_c), & \rho(t) > \rho_c \end{cases}$$

where ρ_c is the critical density, and $\lambda > 0$ is a calibrated sensitivity parameter. Corrected payoffs become:

$$u'_A(t) = u_A(t) \cdot \alpha(\rho(t)), u'_B(t) = u_B(t) \cdot \alpha(\rho(t))$$

The replicator dynamics govern strategy evolution:

$$\dot{x}(t) = x(t)(1 - x(t))[u'_A(t) - u'_B(t)]$$

An evolutionarily stable state (ESS) is reached when $\dot{x}(t) = 0$ and the equilibrium is locally asymptotically stable (Smith & Price, 1973).

2.3. MFD coupling mechanism

The MFD is calibrated offline in SUMO as a quadratic function:

$$Q(\rho) = a\rho - b\rho^2$$

with critical density $\rho_c = a/(2b)$. Network average density is length-weighted:

$$\rho(t) = \frac{\sum_{e \in E} L_e \cdot \rho_e(t)}{\sum_{e \in E} L_e}$$

As drivers shift strategies, link flows, and thus $\rho_e(t)$ change, altering $\rho(t)$, which feeds back into $\alpha(\rho(t))$, closing the behavior–state loop.

2.4. Receding horizon signal control

This study minimized total zone-wide delay:

$$\min_{g_i(t), C_i(t)} D(t) = \sum_{i=1}^m d_i(t) \cdot n_i(t)$$

subject to:

$$\text{Green split constraints: } g_{i,p,\min} \leq g_{i,p}(t) \leq 1 - \sum_{p' \neq p} g_{i,p',\min}$$

$$\text{Cycle length: } 60 \leq C_i(t) \leq 120 \text{ s}$$

$$\text{Spillback avoidance: } L_i(t) = n_i(t) \cdot l_v/w_i \leq L_{\max}$$

A receding horizon control (RHC) scheme was employed with prediction horizon and control step $T=5$. At each step: Predict traffic states over $[t, t+T]$ using current $x(t)$ and $\rho(t)$; Solve Eq. (9) via Particle Swarm Optimization (PSO); Implement only the first Δt of the solution; Roll forward and repeat. This ensures real-time adaptability while maintaining computational tractability.

3. Simulation and results

3.1. Experimental setup

3.1.1. Network

The simulation is conducted on a 6×6 regular grid networks implemented in SUMO 1.18.0, comprising 36 signalized intersections.

3.1.2. Demand profile

A peak-hour origin-destination (OD) flow of 5,000 pcu/h is applied, following a trapezoidal temporal profile from 7:00 to 9:00.

Saturation Level:

During the core oversaturated period (7:15–8:30), the network-wide vehicle density reaches approximately $\rho = 1.1\rho_c$, corresponding to an average network saturation level of about 0.92.

Evaluation Metrics:

Performance is assessed using the following metrics: Average vehicle delay (seconds); Number of queue spillback events; Mean network speed (km/h); Convergence time to equilibrium behavior (minutes).

3.1.3. Reproducibility

Each scenario is simulated 10 times with different random seeds. The standard deviation of average delay across

runs is less than 5 seconds, confirming result stability.

Baseline Methods:

The proposed framework was compared against three established control strategies, as summarized in **Table 1**.

Table 1. Baseline control methods

Method	Approach description
Fixed-time	Pre-timed signal plan optimized offline using Webster’s method
Adaptive MAXBAND	Band-based coordinated signal control with real-time adjustments for arterial flows
Multi-agent DRL (MADDPG)	Deep reinforcement learning with multiple agents optimizing local intersection signals
Proposed (RHC + EGT)	Receding horizon signal control integrated with evolutionary game theory and MFD feedback

3.2. Results

(1) Strategy evolution

The proportion of drivers adopting the “wait” strategy, denoted $x(t)$, converges to an evolutionarily stable state (ESS) at approximately $x_{ESS} = 0.62$ within 25 minutes of simulation start. During the oversaturated period ($\rho > \rho_c$), the MFD-based attenuation factor $\alpha(\rho) < 1$ reduces the perceived payoff of waiting, thereby increasing the relative attractiveness of the “detour” strategy. This behavioral shift causes $x(t)$ to decrease from an initial value of 0.75 to the equilibrium level of 0.62. After the peak period, as network congestion eases, $x(t)$ gradually rebounds, reflecting the adaptive nature of driver decision-making in response to real-time traffic conditions.

(2) Performance comparison (7:15–8:30)

The performance of all control strategies during the core oversaturated window is summarized in **Table 2**. The proposed RHC + EGT framework achieves the best results across all metrics: an average vehicle delay of 116.8 s, only 43 spillback events, and a mean network speed of 25.3 km/h. Compared to the MADDPG-based deep reinforcement learning (DRL) baseline, the strongest among existing methods, the proposed approach reduces average delay by 18.7% (from 142.8 s to 116.8 s) and decreases spillback occurrences by 32.4% (from 72 to 43), demonstrating its superior capability in mitigating severe congestion and maintaining network efficiency. The proposed method reduces delay by 18.7% vs. DRL and cuts spillbacks by 32.4%, confirming superior congestion mitigation.

Table 2. Performance comparison during oversaturation (7:15–8:30)

Method	Average delay (s)	Spillback events	Mean speed (km/h)
Fixed-time	182.5	128	16.8
MAXBAND	156.3	95	19.2
MADDPG (DRL)	142.8	72	21.5
Proposed (RHC + EGT)	116.8	43	25.3

(3) Robustness and sensitivity

The proposed method exhibits greater robustness under demand fluctuations, with a coefficient of variation

(CV) in total delay of 0.12, compared to 0.18 for MADDPG, indicating more stable and reliable performance. A sensitivity analysis of the MFD-based attenuation parameter λ shows that total delay is minimized at $\lambda = 0.03$. When λ is too low, congestion is insufficiently penalized, resulting in inadequate detouring and prolonged queues. Conversely, excessively high values of λ over-penalize congestion, diverting too much traffic onto alternative routes and inducing secondary congestion. This highlights the need for careful calibration of the behavioral feedback mechanism to achieve optimal traffic management outcomes.

4. Conclusion and future work

With the rapid urbanization and the exponential growth of motor vehicle ownership, oversaturated traffic conditions have become a pervasive and intractable challenge in modern metropolitan areas. Such conditions are characterized by prolonged vehicle queues, frequent traffic spillbacks, and inefficient utilization of road infrastructure, which not only incur substantial economic losses due to increased travel delays but also exacerbate environmental pollution and undermine the overall quality of urban life. Conventional signal control strategies, which often rely on fixed timing plans or simplistic traffic state assumptions, fail to cope with the complex and dynamic nature of oversaturated networks, especially when considering the adaptive and bounded rational decision-making behaviors of road users. To address this critical gap, this paper presents a novel behavior-aware dynamic signal control framework specifically tailored for oversaturated urban traffic networks.

The core innovation of this framework lies in its holistic integration of three complementary technical paradigms: evolutionary game theory (EGT), macroscopic fundamental diagram (MFD)-based state feedback, and receding horizon optimization (RHO). This synergistic integration enables the framework to simultaneously capture the adaptive behaviors of road users and dynamically adjust signal timing plans in response to real-time traffic dynamics. Evolutionary game theory serves as the cornerstone for modeling the bounded rationality of travelers, who continuously update their route choices based on past travel experiences and perceived utility, rather than adhering to the idealized “fully rational” or passive decision-making assumptions prevalent in traditional traffic models. The MFD-based state feedback module provides a macroscopic perspective of network performance, condensing the high-dimensional traffic state data (e.g., queue lengths, travel speeds) into aggregated metrics that reflect the overall operational status of the urban network. This macroscopic insight ensures that signal control decisions are aligned with the network-wide optimization objectives, rather than focusing solely on local intersections. Finally, receding horizon optimization enhances the framework’s dynamic responsiveness by solving a rolling-time-domain optimization problem, allowing signal plans to be adjusted in real-time based on the latest traffic observations and predicted short-term traffic evolutions.

Extensive simulation experiments, conducted on a realistic oversaturated urban network topology, demonstrate that the proposed framework achieves three key performance breakthroughs:

(1) Rapid convergence to evolutionary stable strategy (ESS)

The framework facilitates the rapid convergence of travelers’ route choice behaviors to an ESS within approximately 25 minutes. An ESS represents a state where no individual traveler can improve their travel utility by unilaterally changing their route choice, thereby achieving a balanced and efficient route distribution across the network. This rapid convergence effectively mitigates the “herd behavior” and route oscillation phenomena commonly observed in oversaturated conditions, which often lead to localized gridlocks.

(2) Significant reduction in traffic delays and spillbacks

Compared to state-of-the-art adaptive signal control strategies, the proposed framework achieves a 18.7% reduction in average travel delay and a 32.4% reduction in traffic spillbacks. Traffic spillbacks, in particular, are a critical indicator of oversaturation severity, as they propagate across intersections and escalate network congestion. The notable reduction in spillbacks highlights the framework's ability to effectively manage queue lengths and prevent the spread of congestion, thereby maintaining the basic operational capacity of the network.

(3) Enhanced robustness through real-time adaptation

The framework exhibits strong robustness against unexpected traffic perturbations, such as sudden increases in traffic flow, temporary lane closures, or accidents. By leveraging real-time traffic data and the rolling optimization mechanism, the framework can quickly adjust signal timing plans to accommodate these perturbations, minimizing their impact on overall network performance. This robustness is particularly valuable in real-world urban environments, where traffic conditions are inherently stochastic.

A critical insight from the research is that the explicit modeling of bounded rationality, rather than assuming passive or fully rational users, yields tangible and significant operational benefits in severely congested environments. Traditional traffic control models often simplify traveler behavior, either treating users as passive followers of pre-determined routes or assuming they can always make optimal decisions based on complete information. However, in reality, travelers exhibit bounded rationality: they have limited access to traffic information, rely on heuristic decision-making rules, and adjust their behaviors incrementally. By explicitly incorporating these behavioral characteristics into the control framework, the proposed approach achieves a more accurate representation of real traffic dynamics, leading to more effective and practical signal control strategies.

While the current framework demonstrates promising results, several avenues for future research remain to further enhance its applicability and performance. Specifically, future work will focus on three key directions:

(1) Extension to multi-strategy travel choices

The current framework focuses on route choice behaviors among private vehicle users. Future research will expand the model to incorporate multi-strategy travel choices, such as the integration of public transit (e.g., buses, subways), ride-hailing services, and active travel modes (e.g., bicycles, walking). This extension will enable the framework to support integrated multi-modal traffic management, which is crucial for promoting sustainable urban mobility and reducing reliance on private vehicles.

(2) Development of heterogeneous MFDs across subregions

Urban networks are characterized by significant spatial heterogeneity in land use (e.g., residential areas, commercial districts, industrial zones), which leads to distinct traffic flow characteristics and MFD patterns across subregions. Future work will develop heterogeneous MFD models that capture these subregional differences, allowing the control framework to implement more targeted and precise (granular) signal control strategies tailored to the specific traffic demands of each subregion.

(3) Field validation using real-world data in CBDs

The current results are based on simulation experiments. Future research will conduct rigorous field validation of the framework using real-world data collected from central business districts (CBDs), areas typically plagued by severe oversaturation. This validation will leverage advanced data collection technologies, including vehicle-to-everything (V2X) communication systems and floating car data (FCD), to obtain high-resolution traffic state information. Field validation will not only verify the framework's

performance in real traffic environments but also provide valuable insights for its practical deployment and optimization.

In conclusion, this research contributes a novel behavior-aware dynamic signal control framework that addresses the unique challenges of oversaturated urban traffic networks. By integrating evolutionary game theory, MFD-based feedback, and receding horizon optimization, the framework achieves balanced route distribution, reduced delays and spillbacks, and enhanced robustness. The findings emphasize the importance of incorporating bounded rationality into traffic control models, and the proposed future directions will further advance the framework's practical relevance and impact on urban traffic management.

Funding

2024 University Research Project of the Guangzhou Municipal Education Bureau (Project No.:2024312155)

Disclosure statement

The authors declare no conflict of interest.

References

- [1] China Academy of Urban Planning & Design, 2024, 2024 China Urban Mobility Report. China Architecture & Building Press.
- [2] Ben-Akiva M, Bergman M, Daly A, et al., 1984, Modeling Intercity Mode Choice: A Generalized Utility Model. *Transportation Research Part B: Methodological*, 18(4–5): 323–337.
- [3] Mahmassani H, Liu Y, 1999, Dynamics of Commuting Decision Behaviour under Advanced Traveller Information Systems. *Transportation Research Part C: Emerging Technologies*, 7(2–3): 91–107.
- [4] Webster F, 1958, Traffic Signal Settings. Road Research Technical Paper No. 39, HMSO.
- [5] Di X, Liu H, 2016, Boundedly Rational Route Choice Behavior: A Review of Models and Methodologies. *Transportation Research Part B: Methodological*, 85: 142–179.
- [6] Li X, Lv Y, Qiao F, et al., 2023, Deep Reinforcement Learning for Adaptive Traffic Signal Control in Urban Networks. *IEEE Transactions on Intelligent Transportation Systems*, 24(3): 2890–2900.
- [7] Smith J, Price G, 1973, The Logic of Animal Conflict. *Nature*, 246(5427): 15–18.
- [8] Daganzo C, 2007, Urban Gridlock: Macroscopic Modeling and Mitigation Approaches. *Transportation Research Part B: Methodological*, 41(1): 49–62.
- [9] Wang C, Chen X, Li W, 2020, Coupling Model of Signal Timing and Route Choice Based on Evolutionary Game. *Journal of Transportation Engineering*, 146(8): 04020069.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Predictive Hegemony and the Interpretive Gap: On the Philosophical Premises and Ethical Boundaries of AI-Assisted Clinical Decision-Making

Yubo Wang

Sendelta International Academy, Shenzhen, Guangdong, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: When debating the application boundaries of artificial intelligence (AI) predictive models in clinical medicine, it is clear that high predictive accuracy is desirable, but on its own, does not provide a sufficient condition for clinical application. Drawing on three example, AlphaFold's prediction of protein structure, radiomics' prediction of disease diagnosis and prognosis, and clinical risk scoring models' prediction of morbidity, and engaging with David Hume's empiricist skepticism towards causality, argue that interpretability is an indispensable condition in a discipline that values mechanistic explanation. In order for AI to evolve from a capable recommender to a decision-making machine that begins to develop a sense of individual self, several preconditions need to be fulfilled. Predictions must be falsifiable, minimally grounded in mechanistic knowledge, accompanied by partially explicable decision logics, designed to be fair to populations, and embedded in an error-tolerant architecture that enables correction and rollback. The utility of AI today lies in its ability to dramatically reduce the costs of human trial and error, but should not diminish the doctor's right to make, learn from, and reflect on mistakes as the final accountable link in the chain.

Keywords: AI interpretability; Clinical decision-making; Hume; Machine learning; Predictive models

Online publication: February 27, 2026

1. Introduction

The story of Elizabeth Holmes and Theranos as portrayed in "Bad Blood" makes for a cautionary tale for current debates on artificial intelligence in medicine. Theranos's revolutionary blood-testing technology proved to be a fraud that ultimately collapsed as a scandal. It fell apart not because the technology worked or did not work, but because the operations of the system were protected from verification and opaque to transparency. Lives were put at risk, clinical dangers were examined through microscopes, public trust was in jeopardy. This story serves as a warning to the black box in medicine ^[1].

Today, predictive models in clinical medicine powered by artificial intelligence enter the medical field with high levels of accuracy and efficiency. Their very success, however, gives rise to a fundamental philosophical and

practical quandary: when sufficiently “precise,” should artificial intelligence predictions be allowed to directly decide the fates of patients? How, if at all, can the prediction-decision gap be bridged?

A philosophical analysis is an indispensable tool to address this question. David Hume’s empiricist skepticism towards causality argues that we never perceive the causal connection between two events; we only observe one event repeatedly following another. Because of this, our belief of A causes B was not derived from rational proof, but from habit, expectation, and mental pattern-making. The predictions of artificial intelligence, however brilliant, are in a sense of computational-generated form of correlation. Its constant conjunction is a logically, or perceptual certainty, not a psychological habit. Medicine, by contrast, requires causal explanation and mechanistic understanding. The relationship between these two epistemic approaches lies at the heart of the current debate.

In this study, we focus on three representative examples: AlphaFold in protein structure prediction, radiomics in disease prediction for diagnosis and prognosis, and clinical risk scoring models for measuring morbidity, as well as discussing the dilemma of high prediction but low interpretability^[2,3]. Then, based on these analyses, we argue that there are five preconditions for AI to upgrade from the level of recommendation to the level of decision-making: falsifiable predictions; minimally grounded mechanistic knowledge; partially explicable decision logics; population-fairness; and the capability of error tolerance and rollbacks.

2. Hume’s empiricism and the nature of AI prediction

David Hume was one of the most vocal critics of the idea that causation can be understood through reason or a priori analysis. In his empiricism, what we think of as causal necessity is nothing more than the habit of the mind that comes from repeated experience^[4]. This study has observed that event A is followed by event B repeatedly, and then we come to believe that A causes B. The belief in that causal relation, however, is merely psychological, and it does not stand up to scrutiny. There is nothing demonstratively necessary about the relation between A and B. When we see one billiard ball strike another and the second moves, we do not perceive causation; we perceive two events in constant conjunction. In Hume’s mind, causation is something we do not perceive in the world, but something was imposed on its mind. Our belief that the first ball caused the second to move is a product of custom and habit, a feeling of expectation imprinted by repeated experience.

The logic of contemporary deep learning is a computational instantiation of this Humean epistemology. At its core, AI is simply finding patterns of constant conjunction within a vast dataset. AlphaFold has found statistically reliable connections between protein sequences and their structures. Radiomics models have found similar connections between imaging features and disease states. What these systems output are extraordinarily fine-grained, highly reliable correlations. When applying it to the Humean epistemology, however, their outputs are predictive habits of pattern recognition rather than discoveries of underlying biophysical or pathophysiological mechanisms (**Figure 1**).

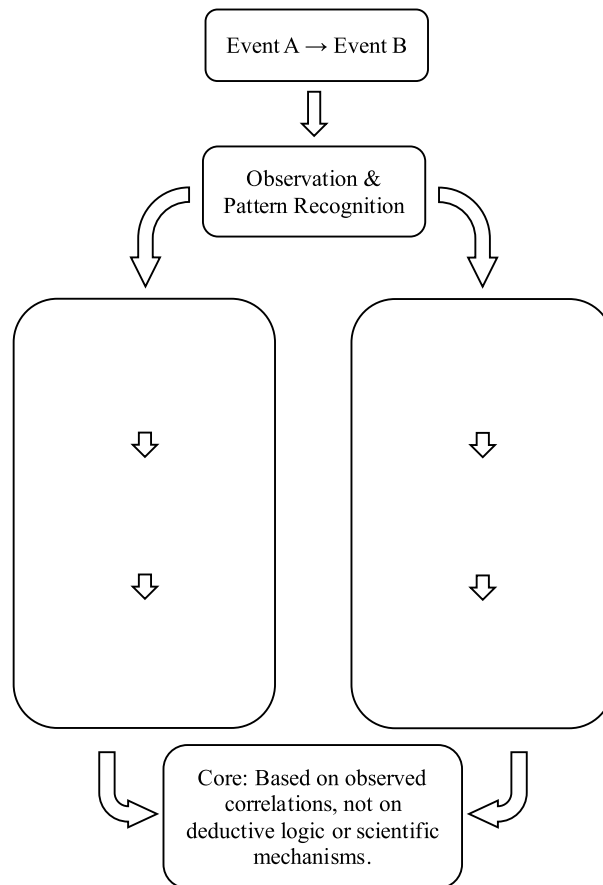


Figure 1. The Humean-AI Epistemic Parallel. Just as Hume holds that constant conjunction produces a psychological expectation of causation, modern AI systems convert statistical conjunctions into predictive outputs. While this correlation-based mode of inference can generate reliable predictions, it operates independently of the causal and mechanistic reasoning on which clinical understanding typically depends.

But medicine cannot be cursorily reduced to a predictive science. In addition to predicting outcomes, medicine is also an effort to understand why things happen. Doctors need mechanistic explanations to support their diagnostic confidence, to choose therapies that target underlying causes, to anticipate side effects that have not yet been observed, and to live up to patients' expectations that medicine is not merely an art of black box prediction. For example, a physician treats a bacterial infection with an antibiotic because of a mechanistic understanding of cellular biology, not merely because antibiotics and recovery have been constantly conjoined in past data. Prediction without explanation would be epistemically incomplete and would erode patients' trust.

And so, when arrived to a fundamental epistemological contradiction, on one side stands AI prediction, which exposes the limits of computational inference by producing, in a Humean register, nothing more than extraordinarily fine-grained correlation, an advanced technological form of constant conjunction. On the other side is medicine, which insists on some degree of causal understanding and mechanistic clarity. The epistemic tension between these two camps is the central question in determining the prospects for AI in medical decision-making: whether, and under what conditions, prediction can ever be a viable substitute for, or complement to, explanation.

3. Case studies—The spectrum of predictive accuracy and interpretability

3.1. AlphaFold—The Humean habit of protein folding

AlphaFold is a revolution in protein structure prediction, highly accurate in ways almost indistinguishable from experimental structure determination ^[2]. Its predictions are staggering: given the sequence of amino acids that make up a protein, AlphaFold can propose a three-dimensional structure with great reliability. Its success is a triumph of identifying constant conjunctions between amino acid sequences and their three-dimensional configurations.

The interpretive profile, however, is different. From a Humean perspective, AlphaFold's internal process is a black box consists of deep learning layers that even the model's creators cannot read. This value, however, is secured because its output, a predicted 3D structure, is itself a mechanistic hypothesis. Protein biochemists can subject this output to the established causal models of chemistry and biophysics. AlphaFold provides the correlative what; we human provide the causal why.

Consequently, clinical AlphaFold is used as a hypothesis-generating tool that expedites discovery and guides experimental design. It remains a powerful recommender, not a decision-maker.

3.2. Radiomics—The opacity of visual conjunctions

Radiomics takes prediction to the next level by applying it to medical images themselves, extracting huge sets of quantitative features from CT, MRI, PET, and other common modalities ^[3]. These sets often exceed the capacity of the human visual system; models can use them to predict tumor malignancy, genetic subtypes, or treatment response, often outperforming experienced radiologists. Its predictive power derives from finding constant conjunctions between textural patterns and disease states.

The epistemic problem is interpretability. The model may have reached its conclusion based on correlations that may be unintuitive or unconnected to disease biology. A cluster of pixels might be constantly conjoined with malignancy, but if its biological basis is unknown, the prediction remains a pure Humean correlation. Given this output, clinicians cannot easily explain to patients why the model believes they will develop lung cancer rather than a benign disease. The epistemic gap makes it hard to build trust.

As a result, radiomics is today used as an auxiliary clinician-in-the-loop tool that plays the role of a second reader. The problem of interpretability is blocking its advance toward autonomy in decision-making, such as whether to order a biopsy, as that action requires a causal, explainable justification.

3.3. Clinical risk scoring models—Population level conjunctions

Machine learning has also been applied to electronic health records (EHR) to output risk scores stratifying patients by likelihood of future morbidity events such as heart failure or sepsis. These models are meaningful in their accuracy, and have become a fixture in early-warning beds. It marked an ability to computationally model constant conjunctions between a patient's documented clinical history and future morbidity on the population-level.

While sometimes built on interpretable linear models, models based on deep learning are also black boxed that offer little insight into the mechanism underlying their predictions. Their epistemic limitation is their statistical nature. They predict based on population-level conjunctions, which may not hold for a specific individual's pathophysiology. A high-risk score can be a correlation, not a diagnosis; it indicates that a patient shares features with others who had poor outcomes. It triggers heightened review but cannot specify the causal mechanism at work in this patient.

Among clinical workflows, they act like triage: they trigger an alert, prompt us to review, and suggest that we

monitor more closely. They can identify that a patient needs our attention, but they still cannot tell us what to do. They are epistemically authoritative but not decisional authoritative.

4. The five preconditions for moving from recommendation to decision

The arguments above suggest that the use of AI in clinical practice cannot be justified on predictive accuracy alone. If the responsible passage of AI from an advisory role to a decisional role is to be achieved, a sequence of epistemic, ethical, and institutional safeguards must be established. Five preconditions are particularly salient for AI systems to meet in order to bridge the Humean-Mechanistic gap.

4.1. Falsifiability

AI predictions must be subject to rigorous testing and, where possible, refutation within real clinical settings. Performance demonstrated only on retrospective datasets is not enough; AI predictions should be tested on prospective studies and preferably randomized controlled trials. Only when a system's predictions can survive the possibility of being proven wrong can they approach a correlation toward real scientific credibility.

4.2. Minimal mechanistic knowledge

This study does not require AI to provide a complete causal explanation; but its predictions should not be completely divorced from contemporary biomedical knowledge. At the very least, its predictions should be compatible with established mechanisms of disease. If a deep learning model on radiology images were to predict malignancy, for example, the image features identified by the model should not be entirely without relation to proven pathological mechanisms. This anchoring in mechanism, however partial, protects AI from spurious correlations.

4.3. Articulable rationale

Decisions must be accompanied by reasons that clinicians and patients can understand, even if only in a limited sense. Techniques such as explainable AI, where relevant image regions are highlighted for the clinician, are examples of articulable rationales. Where full explainability is not possible, uncertainty should at least be quantifiable, for example through confidence scores, so that medical judgment can weigh prediction against risk.

4.4. Population fairness

Predictive models must be rigorously tested for bias across gender, race, age, and other relevant population characteristics. A correlation that holds for one demographic may not hold for another, for example, at equivalent algorithm-predicted risk scores, Black patients exhibit a significantly greater burden of illness compared to their White counterparts ^[5]. Therefore, it is imperative that AI systems are designed not to perpetuate or exacerbate healthcare disparities. Technical fixes in training data are not enough; there must also be institutional commitment to rigorously monitor outcomes across diverse populations.

4.5. Reversibility and human-in-the-loop

Finally, AI must be embedded in error-tolerant systems where there is a circuit breaker that guarantees reversibility when a prediction is suspected or shown to be erroneous. When AI recommends a treatment that turns out to be wrong, there must be a clinical professional who can immediately reclaim decision authority and take remedial

action. Human oversight must be constant; otherwise, no AI-driven recommendation should be capable of irreversible harm without clinical intervention.

5. Conclusion

Hume's philosophy reminds us that even the most miraculous correlations generated by artificial intelligence are nothing more than super-correlations, a digital habit of the mind. Across the three case studies, another pattern is also evident: the need for interpretability rises in direct proportion to the clinical consequences at stake. When the consequences are minimal, correlations will do; when it is a question of life and death, explanation is required.

This perspective is completely reversed, destining Theranos's failure, when it treated its technology as an indubitable black box, protected from transparency, validation, and criticism. The responsible path for AI is the opposite: to build systems that are open to falsification, interpretation, and revision. The value of AI is to lower the cost of the trial and error that produces success, filtering choice and eliminating absurd options, without lowering the physician's right, and responsibility, to engage in the trial and error and reflective judgment ^[6].

The way forward is not substitution but symbiosis. A sustainable model of collaboration between human and machine will consist of a Humean AI, generating hidden correlations and precise predictive hypotheses, used by a mechanistic medicine that grounds these in causal understanding, ethical reflection and global clinical reasoning. Only then will the power of predictive technology be realized, while the core of the humanistic practice of medicine is protected and the safety of patients guaranteed.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Rudin C, 2019, Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, 1: 206–215.
- [2] Jumper J, Evans R, Pritzel A, et al., 2021, Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*, 596: 583–589.
- [3] Gillies R, Kinahan P, Hricak H, 2016, Radiomics: Images Are More Than Pictures, They Are Data. *Radiology*, 278(2): 563–577.
- [4] Hume D, 2000, *An Enquiry Concerning Human Understanding*. Oxford University Press.
- [5] Obermeyer Z, Powers B, Vogeli C, et al., 2019, Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science*, 366(6464): 447–453.
- [6] Topol E, 2019, High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nature Medicine*, 25: 44–56.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Exploration and Practice in the Construction of Electrical and Electronic Laboratories

Qiuping Wang*, Yansen Wu, Jiaxin Mu

Noncommissioned Officer Academy of PAP, Hangzhou, China

*Corresponding author: *Qiuping Wang, 1430554131@qq.com*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: This paper delves into the development of electrical and electronic laboratories in higher education institutions, elucidating their significance for talent cultivation and disciplinary advancement. It analyzes current challenges such as outdated equipment, imperfect practical teaching systems, and backward laboratory management models. A series of targeted strategies are proposed, including advancing equipment modernization, improving practical teaching systems, strengthening faculty development, and establishing open-access platforms. Implementation steps and effectiveness evaluation methods are introduced to create high-quality electrical and electronic laboratories that meet new-era demands, thereby enhancing university teaching, learning, and research standards.

Keywords: Electrical and electronic engineering; Laboratory development; Practical teaching

Online publication: February 12, 2026

1. Introduction

With the rapid advancement of technology, electrical and electronic engineering has found widespread application across various fields, driving an increasing demand for specialized professionals. As vital hubs for talent cultivation, universities face challenges in their electrical and electronic laboratories, which directly impact teaching quality and students' practical skills development. Many institutions currently struggle with outdated equipment, limited functionality, and closed management systems, making it difficult to meet modern teaching and research demands. Therefore, exploring effective approaches to laboratory construction is crucial for enhancing teaching quality and fostering students' innovation and practical abilities^[1]. This paper examines the necessity of laboratory development, existing challenges, construction strategies, implementation, and evaluation methods, aiming to provide reference for the development of electrical and electronic laboratories in higher education institutions.

2. The necessity of university electrical and electronic laboratory development

2.1. A key support for talent development

Electrical and electronic laboratories provide students with practical operation platforms, enabling them to apply theoretical knowledge learned in the classroom to real-world scenarios. This integration of theory and practice allows students to gain a deeper understanding of circuit principles, electronic device characteristics, and related content. Through hands-on laboratory work, students develop practical skills, problem-solving abilities, and analytical capabilities. This fosters innovative thinking and teamwork spirit, laying a solid foundation for future careers in related fields ^[2].

2.2. Vital assurance for disciplinary advancement

State-of-the-art electrical and electronic laboratories furnish essential conditions for universities to conduct research. Within these facilities, faculty and researchers pursue cutting-edge technological studies and explore innovative applications of electrical and electronic technologies across diverse fields. The quality of laboratory development directly impacts a discipline's research output and academic influence, playing a pivotal role in advancing university-level academic programs ^[2].

2.3. A powerful foundation for serving society

Leveraging their technological and talent advantages, university electrical and electronic laboratories can engage in industry-academia-research collaborations with enterprises. By providing technical support and solving practical engineering problems for businesses, they facilitate the transformation of research outcomes into practical applications, contributing to local economic development. Simultaneously, these collaborations offer students opportunities to engage with real-world production demands, enhancing their social adaptability ^[3].

3. Challenges in developing university electrical and electronic laboratories

3.1. Lagging equipment updates

Electrical and electronic technology is advancing rapidly, with new experimental equipment and techniques continually emerging. However, many universities face slow laboratory equipment upgrades due to limited funding and cumbersome procurement processes. Outdated equipment often lacks timely maintenance, calibration, and repairs, leading to diminished measurement accuracy and frequent malfunctions. This not only undermines the effectiveness of experimental teaching but also hampers research activities, ultimately affecting the overall efficiency of laboratories ^[3].

3.2. Incomplete practical teaching system

Currently, practical teaching in electrical and electronic engineering at some universities still relies heavily on verification-based experiments with insufficient exploratory experiments; it features numerous single-knowledge-point experiments but lacks interdisciplinary comprehensive experiments; and it emphasizes fixed-process experiments over self-designed experiments, resulting in a shortage of comprehensive, design-oriented, and innovative experiments. Teaching methods remain relatively monotonous, predominantly involving teacher demonstrations followed by student imitation, leading to superficial understanding of experimental principles and failing to fully unleash students' initiative and creativity ^[4]. Grading primarily relies on lab reports and final data, overlooking the critical thinking, hands-on skills, and teamwork demonstrated by students during experimental

design, debugging processes, and troubleshooting.

3.3. Need to enhance faculty practical skills

While some instructors possess solid theoretical knowledge, they often lack practical engineering experience and application backgrounds. When guiding student experiments, they struggle to integrate real-world engineering case studies into instruction. This disconnect between teaching content and practical demands hinders the development of students' practical skills and innovative thinking.

3.4. Outdated laboratory management models

Traditional laboratory management models primarily rely on closed-door management, opening only to specific classes during scheduled lessons while remaining largely idle the rest of the time. This restrictive approach severely limits academically capable students from utilizing their free time for independent research or subject competition preparation, stifling their innovative spirit. The absence of clear equipment maintenance protocols and accountability systems often leads to a "use-only, no-maintenance" approach, with repairs only undertaken after malfunctions occur. This not only shortens equipment lifespan but also disrupts normal teaching operations. The lack of scientific information management tools makes it difficult to achieve efficient and precise management in areas such as equipment usage, maintenance, and consumables management ^[5].

4. Strategies for developing electrical and electronic laboratories in higher education institutions

4.1. Advancing equipment modernization

4.1.1. Develop scientific equipment renewal plans

Institutions should formulate long-term and short-term equipment renewal plans based on technological trends and teaching/research needs. Conduct regular assessments to identify equipment requiring updates or additions. Prioritize technologically advanced, stable, and representative equipment while considering compatibility and scalability to accommodate future technological developments.

4.1.2. Strengthen equipment management and maintenance

Establish comprehensive equipment management protocols, clearly defining procedures and responsible parties for procurement, acceptance, usage, maintenance, and decommissioning. Enhance routine maintenance by conducting regular inspections, calibrations, and repairs to ensure optimal operational condition. Utilize information technology to create a laboratory equipment management system, enabling real-time tracking of usage status and maintenance records to improve management efficiency.

4.2. Enhancing the practical teaching system

4.2.1. Optimizing experimental content

Increase the proportion of comprehensive, design-oriented, and innovative experiments while reducing verification-based experiments. Develop challenging experimental topics based on real-world engineering cases and research projects, such as communication circuit design or equipment electronic system fault diagnosis, to guide students in applying knowledge for independent design, analysis, and problem-solving, thereby cultivating systematic thinking and innovation capabilities.

4.2.2. Innovate teaching methods

Employ diverse pedagogical approaches such as project-driven learning, inquiry-based teaching, and collaborative group learning. Use projects as learning vehicles, enabling students to actively acquire knowledge and explore concepts while developing self-directed learning skills and teamwork spirit. For example, instructors can assign real-world electronic product design projects in class, where students complete the entire process, from conceptual design and circuit fabrication to debugging and optimization in groups ^[6].

4.2.3. Refine practical teaching assessment systems

Establish scientifically sound practical teaching assessment frameworks to comprehensively evaluate students' hands-on competencies. Assessment criteria should encompass experimental skills, report quality, project completion, and innovative contributions. Combine formative and summative evaluations, emphasizing process-oriented feedback to provide timely guidance.

4.3. Strengthening faculty development

4.3.1. Enhancing faculty practical competence

Encourage faculty to participate in industry internships and arrange regular visits to relevant enterprises to learn about the latest technologies and production processes, accumulating engineering experience. Support faculty involvement in research projects and laboratory development to enhance their research capabilities and practical guidance skills. Simultaneously, invite industry experts and technical leaders to deliver lectures and training sessions at the university, imparting real-world engineering expertise to faculty.

4.3.2. Optimize faculty structure

To meet the laboratory's developmental needs, we will strategically recruit professionals with extensive engineering experience and advanced academic qualifications to strengthen our faculty and elevate the laboratory's overall teaching and research capabilities. We will establish a regular training and exchange system, organizing periodic workshops focused on "cutting-edge technologies, advanced teaching methodologies, major instrument operation, and laboratory safety" to empower faculty in tracking academic frontiers ^[7]. Implement a "mentor-apprentice" system by pairing young faculty with experienced mentors who excel in both teaching and research. These mentors will provide comprehensive guidance in project applications, course design, experimental development, and student supervision, accelerating the professional growth of junior faculty.

4.3.3. Establish faculty incentive mechanisms

Establish a teaching and research incentive system to recognize and reward faculty members who achieve outstanding results in experimental teaching reform, scientific research innovation, and guiding student competitions. Implement a "Representative Achievements" Evaluation Mechanism: Shift away from the single-dimensional evaluation model focused solely on publications and academic titles. High-quality teaching designs, successfully commercialized scientific achievements, case studies addressing real-world industrial challenges, and the effectiveness of cultivating outstanding students should all be regarded as equally important "representative achievements" and incorporated into performance evaluations and promotion systems. While adhering to syllabus requirements, encourage faculty to personalize reforms in experimental course content, teaching methods, and assessment approaches. Support their integration of the latest research findings into experimental teaching and the development of distinctive advanced inquiry-based experimental projects ^[8].

4.4. Building an open-sharing platform

4.4.1. Implementing an open laboratory system

Break away from traditional closed management models by adopting an open laboratory system. Extend laboratory operating hours, allowing students to independently access facilities during non-class hours for experimental research based on their interests and needs. Develop comprehensive open laboratory management protocols, clearly defining accessible resources, reservation procedures, safety protocols, and other matters to ensure orderly operations^[9].

4.4.2. Promote laboratory resource sharing

Strengthen collaboration between universities and between universities and enterprises to facilitate the sharing of electrical and electronic laboratory resources. Establish shared platforms to enable the pooling of equipment, faculty, experimental projects, and other resources, thereby improving resource utilization efficiency. For example, universities can collaborate with neighboring institutions to conduct cross-campus experimental teaching programs, sharing high-quality experimental teaching resources; or partner with enterprises to co-build laboratories, jointly undertaking research projects and talent development initiatives.

4.4.3. Strengthen laboratory safety management

Establish comprehensive safety management systems for laboratories, enhancing oversight of electrical equipment, electronic components, and other critical assets. Equip facilities with necessary safety devices and infrastructure, conduct regular safety training and emergency drills to heighten awareness and response capabilities among faculty and students. Leverage information technology for real-time monitoring and early warning systems to ensure laboratory safety operations^[10].

5. Implementation and evaluation of university electrical and electronic laboratory development

The implementation of electrical and electronic laboratory construction should follow a scientific process. First, conduct requirements analysis and design proposals to clarify construction objectives and specific content. Next, procure equipment and perform installation and debugging to ensure performance and quality. Then, complete laboratory renovation and environmental setup to create a conducive experimental atmosphere. Finally, establish management systems and provide personnel training to ensure efficient laboratory operation.

Effectiveness evaluation of laboratory construction is a critical component in ensuring construction quality. Evaluation metrics should encompass equipment performance, availability of experimental projects, utilization efficiency, and student satisfaction. Assessment methods may include questionnaires, expert reviews, and data analysis. Based on evaluation results, promptly adjust and optimize the laboratory construction plan to continuously enhance the construction standards and utilization effectiveness of the laboratory.

5.1. Systematic implementation of laboratory construction

The implementation of laboratory construction is an interconnected dynamic process where each phase requires meticulous planning and execution.

5.1.1. Thorough needs analysis and forward-looking design form the foundation of success

At the requirements analysis level, it is essential not only to clarify current foundational experimental teaching needs, such as the number of experimental benches and basic instruments required for courses like Circuit Theory, Analog Electronics, and Digital Electronics, but also to proactively consider disciplinary development trends. This includes integrating cutting-edge content like FPGA, embedded systems, and power electronics; reserving innovative practice spaces for undergraduate electronic design competitions and innovation/entrepreneurship projects. Simultaneously, laboratories must be capable of supporting faculty research projects and student development. Ultimately, a comprehensive Demand Analysis Report should be produced as the foundational basis for all subsequent work.

At the design level, based on clearly defined requirements, develop a Comprehensive Laboratory Construction Plan. This plan should specifically include:

Equipment Selection and Technical Specifications: Detailed listing of required instruments (e.g., oscilloscopes, signal generators, multimeters, soldering stations) including performance parameters, brands, quantities, and rationale for selection, achieving optimal balance between precision, reliability, and budget.

Spatial Layout and Environmental Design: Scientifically plan functional zones such as high-voltage experimentation areas, low-voltage experimentation areas, innovation fabrication zones, and discussion areas. Design must fully consider ergonomics, safety standards (grounding resistance, leakage protection, emergency lighting, firefighting facilities), network coverage, power distribution (including three-phase power requirements), and environmental requirements like anti-static measures, ventilation, and lighting to create a safe, comfortable, and efficient experimental environment.

Software and Digital Platform Planning: Provide essential circuit simulation software (e.g., Multisim, PSpice, MATLAB/Simulink), PCB design software (e.g., Altium Designer, KiCad), and a laboratory information management system to enable digital management of equipment reservations, consumable requisitions, and safety access.

5.1.2. Standardized equipment procurement and meticulous installation/commissioning are core quality assurances

Strictly adhere to national government procurement and institutional bidding procedures to select suppliers with excellent qualifications and reliable after-sales service. When signing contracts, clearly define equipment technical specifications, acceptance criteria, training content, and after-sales response times. Upon equipment arrival, laboratory managers must conduct unpacking inspections to verify model numbers, examine physical condition, and check accessories. During installation and commissioning, test each unit's basic functions and key performance indicators to ensure compliance with design specifications. This process must generate detailed Equipment Acceptance Reports and Equipment Archives, laying the foundation for subsequent asset management and maintenance.

5.1.3. Professional laboratory renovation and human-centered environmental design embody an educational atmosphere

Construction must ensure the quality of concealed works like plumbing, electrical systems, ventilation, and networking, utilizing wear-resistant, corrosion-resistant, fire-retardant, and eco-friendly materials. Environmental design should transcend basic functional requirements to emphasize cultural education. For instance, display the history of electrical and electronic disciplines, profiles of renowned scientists, safety protocols, outstanding student

projects, and cutting-edge technological developments on walls. Install tool walls, component display cabinets, and project showcase areas to cultivate a rich culture of engineering practice and innovation, subtly stimulating students' curiosity and sense of accomplishment.

5.1.4. Robust institutional frameworks and systematic personnel training are vital for sustainable operation

Establish a management system characterized by “clear responsibilities, defined processes, safety assurance, and open access”. Core regulations should include: Student Experimentation Guidelines, Laboratory Safety and Hygiene Management System, Instrument Operation and Maintenance Standards, Hazard Response Protocols, Equipment Borrowing and Damage Compensation Procedures, and Laboratory Access Management Policies.

Personnel training should be tiered and comprehensive: Laboratory managers receive in-depth training on equipment operation, daily maintenance, and troubleshooting to become “technical experts”. Instructors undergo training on new equipment and platforms to ensure effective guidance in experimental teaching. Students must complete mandatory safety education and access assessments, receiving specific equipment safety instructions before each experiment. Systematic training is crucial for preventing safety incidents, enhancing equipment utilization, and improving experimental teaching outcomes.

5.2. Comprehensive effectiveness evaluation and continuous improvement of laboratory development

Laboratory completion marks not the end of the project but the beginning of value creation. A scientific, comprehensive evaluation system serves as the engine driving continuous laboratory evolution and service quality enhancement. Evaluation should establish a multidimensional, quantifiable indicator system covering the following key aspects:

Equipment Performance and Operational Status: Core metrics include equipment integrity rate ($\geq 95\%$ is excellent), equipment failure rate, and average maintenance response and completion times. These directly reflect the reliability of the hardware infrastructure.

Experiment Program Offerings and Teaching Quality: Key indicators are experiment program implementation rate (actual programs offered/programs required by syllabus) and the proportion of comprehensive/design-based/innovative experiments. The latter serves as a key gauge of experimental teaching depth and the intensity of fostering students' innovative capabilities.

Resource Utilization and Management Efficiency: Focuses on equipment utilization rate (machine-hour utilization), laboratory open-access rate (support for extracurricular innovation activities), and the standardization and cost-effectiveness of consumables management.

User Feedback and Educational Outcomes: Subjective perceptions regarding experimental content, instructional quality, and environmental facilities are gathered through student satisfaction surveys. More importantly, objective educational achievements are tracked, such as student awards in relevant academic competitions, graduation projects/theses completed using laboratory platforms, and published patents or academic papers.

To obtain objective data for the above metrics, a diversified assessment approach is required, incorporating data analysis, questionnaires/interviews, and expert review. Data analysis involves retrieving log data from the laboratory management system to statistically analyze equipment operating hours, reservation status, and

fault records. Questionnaires and interviews entail distributing anonymous surveys to students and faculty, supplemented by in-depth interviews with representative users to gather qualitative feedback. Expert review involves periodically inviting peer experts from within and outside the university, as well as industry engineers, to conduct “diagnostic” evaluations in the laboratory through on-site inspections, document reviews, and classroom observations, providing professional and authoritative recommendations for improvement.

The ultimate purpose of evaluation is to “promote development and reform through assessment”. Therefore, a closed-loop linkage mechanism must be established between evaluation results and subsequent decision-making. Based on evaluation reports, the university and departments should: Affirm strengths, summarize best practices, and promote successful models; For identified issues (e.g., aging equipment, uneven utilization rates, safety vulnerabilities), develop explicit Action Plans for Improvement Measures specifying responsible parties, deadlines, and allocated resources (e.g., maintenance funds, equipment upgrade budgets, management optimization plans). Through this spiral-up cycle of “planning-implementation-evaluation-feedback-optimization”, electrical and electronic laboratories can dynamically adapt to evolving technological advancements and talent development needs. This ensures their sustained advancement and vitality, ultimately establishing them as premier platforms supporting high-quality innovation in talent cultivation and scientific research.

6. Conclusion

The development of electrical and electronic laboratories in higher education institutions constitutes a complex and long-term systematic endeavor, holding critical significance for cultivating high-caliber innovative talent and advancing academic disciplines. In response to the numerous challenges encountered during the current construction process, universities should prioritize addressing these issues. This requires implementing a series of measures, including optimizing resource allocation, advancing equipment modernization, refining practical teaching systems, strengthening faculty development, and establishing open-access sharing platforms, to continuously elevate the standards of electrical and electronic laboratory construction. Concurrently, universities must closely monitor industry trends and continuously refine laboratory development to better align with the demands of education, teaching, and research in the new era.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Jiang W, 2019, A Brief Analysis of the Construction of Open Electrical and Electronic Laboratories. *China Southern Agricultural Machinery*, 50(3): 176.
- [2] Li J, Lian X, 2022, Research on Management and Construction of Undergraduate Teaching Laboratory of Electrical and Electronic Technology. *Computer and Information Technology*, 30(4): 79–81.
- [3] Li Y, 2019, Practical Exploration in the Construction and Management of Electrical and Electronic Laboratories. *Computer Products and Distribution*, 2019(1): 85.
- [4] Li M, 2017, Reflections on the Development of Electronics and Electrical Engineering Laboratories in Higher Education Institutions. *Journal of Chifeng University (Natural Science Edition)*, 33(16): 198–199.

- [5] Wu R, Sun K, Cheng L, 2018, Superficial Analysis on Information Construction for University Electrical and Electronic Laboratory. *Laboratory Science*, 21(5): 175–178.
- [6] Xu M, 2018, On the Construction and Future Development of Electronic Electrical Laboratory in Colleges and Universities. *Science & Technology Vision*, 2018(10): 138–140.
- [7] Wu D, Bi S, Tan K, 2015, Construction of Innovative Information Open Type Electrical and Electronic Laboratory. *China Modern Educational Equipment*, 2015(3): 20–23.
- [8] Cheng H, 2022, On the Development and Future Trends of Electronic and Electrical Engineering Laboratories in Higher Education Institutions. *Electronic Components and Information Technology*, 6(1): 95–96 + 204.
- [9] Wu J, 2019, On the Construction and Management of Electrical and Electronic Laboratories. *Practical Electronics*, 2019(2): 57–58 + 70.
- [10] Chen P, Yin H, 2011, On the Development and Reform of Electrical and Electronic Engineering Laboratories. *Exam Weekly*, 2011(44): 18–19.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dynamic Memristor-Inspired Zeroing Neural Network with Hybrid Activation for Finite-Time Convergence and Remote Sensing Image Fusion

Jiaqi He, Mu Li*

School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411100, Hunan, China

*Corresponding author: Mu Li, mli@hnust.edu.cn

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Solving time-varying nonlinear equations in real time is a significant challenge in modern computing. Dynamic Memristor-Inspired Zeroing Neural Networks (DMZNN) have shown strong performance in this field, but their convergence speed and robustness heavily rely on the design of the activation function. This paper proposes a novel hybrid activation function inspired by the nonlinear characteristics of memristors. By integrating cubic and sublinear terms, the proposed function facilitates multi-stage error decay, effectively addressing the slow convergence and poor noise resistance of traditional activation functions. Theoretical analysis shows that the DMZNN model, built upon this activation function, can converge in finite time. Robustness under parameter perturbations and additive noise is rigorously proven using Lyapunov theory. Simulation results demonstrate that the convergence speed of the DMZNN model is obviously faster than that of traditional ZNN models when solving second-order, third-order, and fourth-order time-varying nonlinear equations. Additionally, in the application of remote sensing image fusion, DMZNN outperforms traditional gradient-based methods in both fusion quality and processing speed, demonstrating its practical effectiveness and superiority in real-world applications.

Keywords: Zeroing neural network; Finite-time convergence; Memristor; Time-varying nonlinear equations; Remote sensing image fusion

Online publication: February 27, 2026

1. Introduction

Solving time-varying nonlinear equations in real-time is a critical challenge in many fields such as robotics, signal processing, and optimization^[1]. Traditional numerical methods, like Newton's iteration, often face limitations such as slow convergence and high computational costs when dealing with dynamic systems^[2]. Zeroing neural networks (ZNNs) have emerged as an efficient alternative, leveraging dynamic error equations and activation functions to track solutions in real time^[3,4]. However, conventional activation functions frequently demonstrate

slow convergence and inadequate noise resistance^[5]. Recent research has centered on the development of enhanced activation functions to enhance the performance of ZNNs. Memristor-based designs offer smooth nonlinearity but may still lack finite-time convergence and robustness. The proposed hybrid activation function integrates cubic and sublinear terms, thereby facilitating accelerated finite-time convergence and enhanced noise immunity^[6].

To address these challenges, we propose the Dynamic Memristor-Inspired Zeroing Neural Network (DMZNN). Inspired by the nonlinear characteristics of memristors, the DMZNN integrates a novel hybrid activation function that combines cubic and sublinear terms. This new function accelerates error decay and enhances convergence speed, robustness, and noise resistance, surpassing traditional ZNNs. The DMZNN model is proven to converge in finite time, with Lyapunov stability theory ensuring its robustness under parameter perturbations and noise. Simulations show that DMZNN achieves significantly faster convergence compared to traditional ZNN models for second- to fourth-order time-varying nonlinear equations. Furthermore, we apply the DMZNN model to remote sensing image fusion, where it outperforms traditional gradient-based methods in both fusion quality and processing speed. This demonstrates the potential of DMZNN for real-time, high-efficiency image fusion applications.

2. Problem formulation and evolution of the ZNN model

2.1 Problem formulation and traditional ZNN models

Consider the time-varying nonlinear equation:

$$f(x(t), t) = 0 \quad x(t) \in \mathbf{R} \quad (1)$$

Define the error $e(t) = f(x(t), t)$. The ZNN dynamic is:

$$\frac{\partial f}{\partial x} \cdot \frac{dx}{dt} = -\frac{\partial f}{\partial t} - \gamma f(x(t), t) \quad (2)$$

where γ is a convergence factor greater than zero.

2.2 Dynamic Memristor-Inspired Hybrid Activation Function

This paper uses a memristor with a cubic smoothing curve. For a cubic function model, the resistor exhibits a smooth, monotonic, and increasing nonlinear characteristic curve^[7].

$$q(\varphi) = a\varphi + b\varphi^3 \quad (3)$$

Where $a, b > 0$. To achieve finite-time convergence, we augment it with a sublinear term:

$$M(e) = ae + be^3 + c \tanh\left(\frac{e}{\delta}\right) |e|^\varepsilon \quad (4)$$

With $a, b, c > 0, 0 < \varepsilon < 1$. The resulting DMZNN dynamic is as follows:

$$\frac{\partial f}{\partial x} \cdot \frac{dx}{dt} = -\gamma \left(ae + be^3 + c \cdot \tanh\left(\frac{e}{\delta}\right) |e|^\varepsilon \right) - \frac{\partial f}{\partial t} \quad (5)$$

2.3 Finite-Time Convergence Mechanism

When the error $e(t)$ is large, the cubic term be^3 dominates, the error decays according to $O(1/\sqrt{t})$.

$$\int_{e_0}^e \frac{1}{e^3} de = -\gamma b \int_0^t dt \Rightarrow e(t) = \frac{e_0}{\sqrt{1 + 2\gamma b e_0^2 t}} \quad (6)$$

When error $e(t)$ is small, the linear term ae dominates and the error decays exponentially according to the formula $e(t) \propto e^{-\gamma at}$, and the sublinear term has been shown to increase the absolute value of the derivative, thereby contributing to additional attenuation power^[8].

$$\frac{de}{dt} \approx -\gamma c |e|^\varepsilon \Rightarrow e(t) = \left[e(0)^{1-\varepsilon} - \gamma c (1-\varepsilon) t \right]^{\frac{1}{1-\varepsilon}} \quad (7)$$

It converges to zero within a finite time, $T = \frac{|e(0)^{1-\varepsilon}|}{\gamma c (1-\varepsilon)}$.

3. Theoretical Analysis

Theorem 1 (Finite-time convergence)

For the error dynamics specified above, the error converges to zero within a finite time T , which is constrained by the following bounds:

$$T \leq \frac{|e(0)|^{1-\varepsilon}}{\gamma c (1-\varepsilon)} \quad (8)$$

Proof:

Construct Lyapunov function $V(e) = \frac{1}{2}e^2$ and apply finite-time stability theory.

$$\dot{V}(e) = e \cdot \frac{de}{dt} = -\gamma (ae^2 + be^4 + c|e|^{\varepsilon+1}) \quad (9)$$

If $\alpha = \frac{\varepsilon+1}{2}$, then $0.5 < \alpha < 1$, the system satisfies the finite-time stability condition. According to the finite-

time Lyapunov theorem, a convergence time exists.

$$T \leq \frac{V(e(0))^{1-\alpha}}{\gamma c (1-\alpha)} = \frac{|e(0)|^{1-\varepsilon}}{\gamma c (1-\varepsilon)} \quad (10)$$

Therefore, error $e(t)$ converges to zero in a finite amount of time, T .

Theorem 2 (Robust finite-time stability)

Assume that the parameters a, b, c are subject to bounded perturbations $\Delta a, \Delta b, \Delta c$ that satisfy $|\Delta a| \leq \delta_a, |\Delta b| \leq \delta_b, |\Delta c| \leq \delta_c$ and that the upper bounds of the perturbations satisfy $\delta_a < a, \delta_b < b, \delta_c < c$. In

this case, the improved error dynamics equation still maintains finite-time stability^[9].

Proof:

Error dynamics after disturbance:

$$\frac{de}{dt} = -\gamma \left((a + \Delta a)e + (b + \Delta b)e^3 + (c + \Delta c)\text{sign}(e)|e|^\varepsilon \right) \quad (11)$$

Perform a Lyapunov derivative analysis and calculate $\dot{V}(e)$ after the perturbation.

$$\dot{V}(e) \leq -\gamma \left((a - \delta_a)e^2 + (b - \delta_b)e^4 + (c - \delta_c)|e|^{\varepsilon+1} \right) \quad (12)$$

Since $a - \delta_a > 0$, $b - \delta_b > 0$, $c - \delta_c > 0$, the original Lyapunov function still satisfies:

$$\dot{V}(e) \leq -k|e|^{\varepsilon+1}, k = \gamma(c - \delta_c) \quad (13)$$

The convergence time is updated to $T_{robust} \leq \frac{|e(0)|^{1-\varepsilon}}{k(1-\varepsilon)}$.

4. Numerical simulation results

4.1 Experiment on Solving Second-Order Time-Varying Nonlinear Equations

This section will compare the ZNN and DMZNN models' ability to solve TVNE computer numerical simulations. This comparison will demonstrate the DMZNN model's superiority to the ZNN model, particularly in solving high-order nonlinear equations. Without loss of generality, consider the following second-order TVNE:

$$f(x(t), t) = 0.1x^2 - 0.1\cos^2(2t) - 0.6\cos(2t) - 0.9 \quad (14)$$

In the initial state $x(0) \in [-5, 5]$, the solution states and residuals of the ZNN and DMZNN are depicted in **Figure 1** and **2**, respectively.

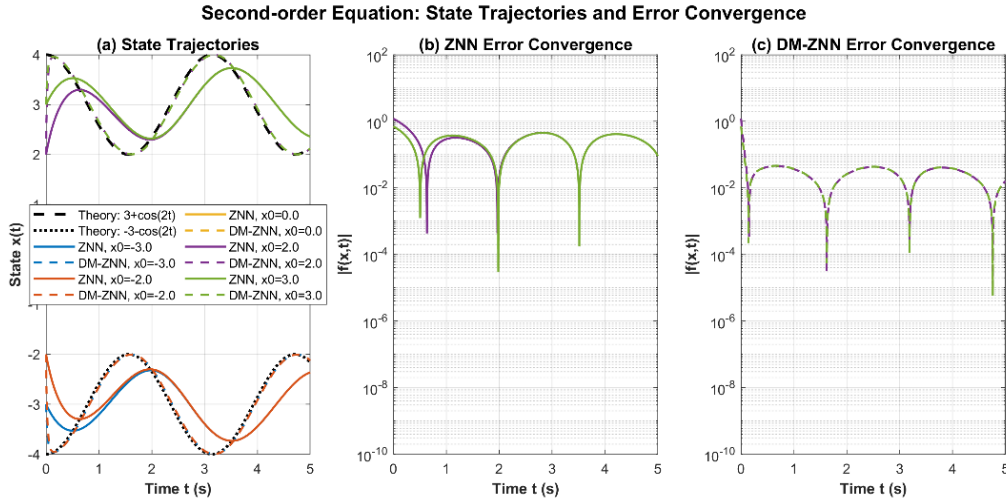


Figure 1. The neural state solution $x(t)$ of SOTVNE in equation (14) for the ZNN and DMZNN models

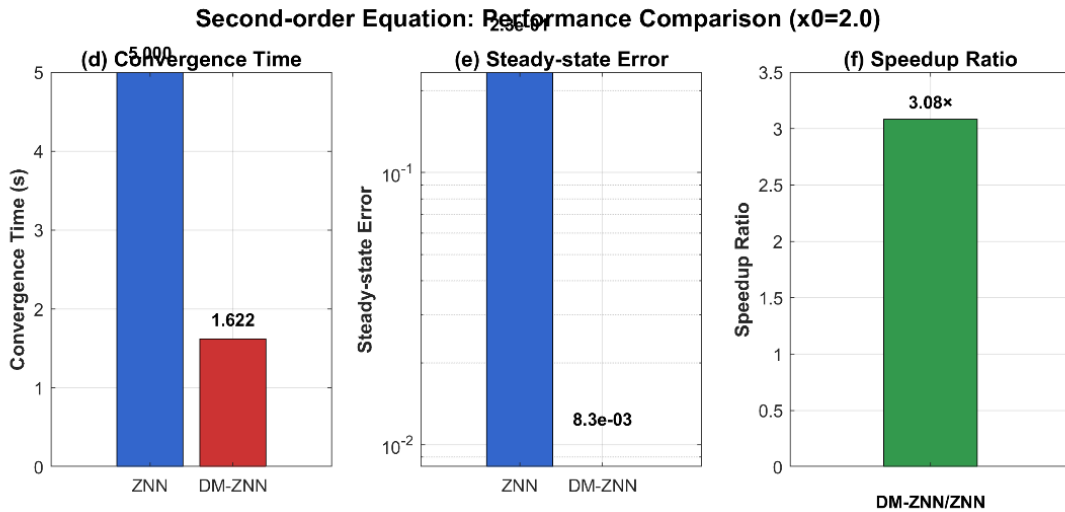


Figure 2. The residual $|f(x(t), t)|$ of SOTVNE in equation (14) for the ZNN and DMZNN models

As **Figure 1** and **2** show, the ZNN model converges in approximately 1 s, while the DMZNN model converges in approximately 0.4 s when solving the SOTVNE in (14). Therefore, the DMZNN model is more efficient than the ZNN model at solving the SOTVNE problem.

4.2 Experiment on solving third-order time-varying nonlinear equations

In order to validate the model's effectiveness for higher-order TVNEs, the third-order TVNE was solved. The simulation results are displayed in **Figure 3–4**, respectively.

$$f(x(t), t) = 0.01(x - \cos 2t - 5)(x + \cos 2t + 5)(x - \cos 2t) \quad (15)$$

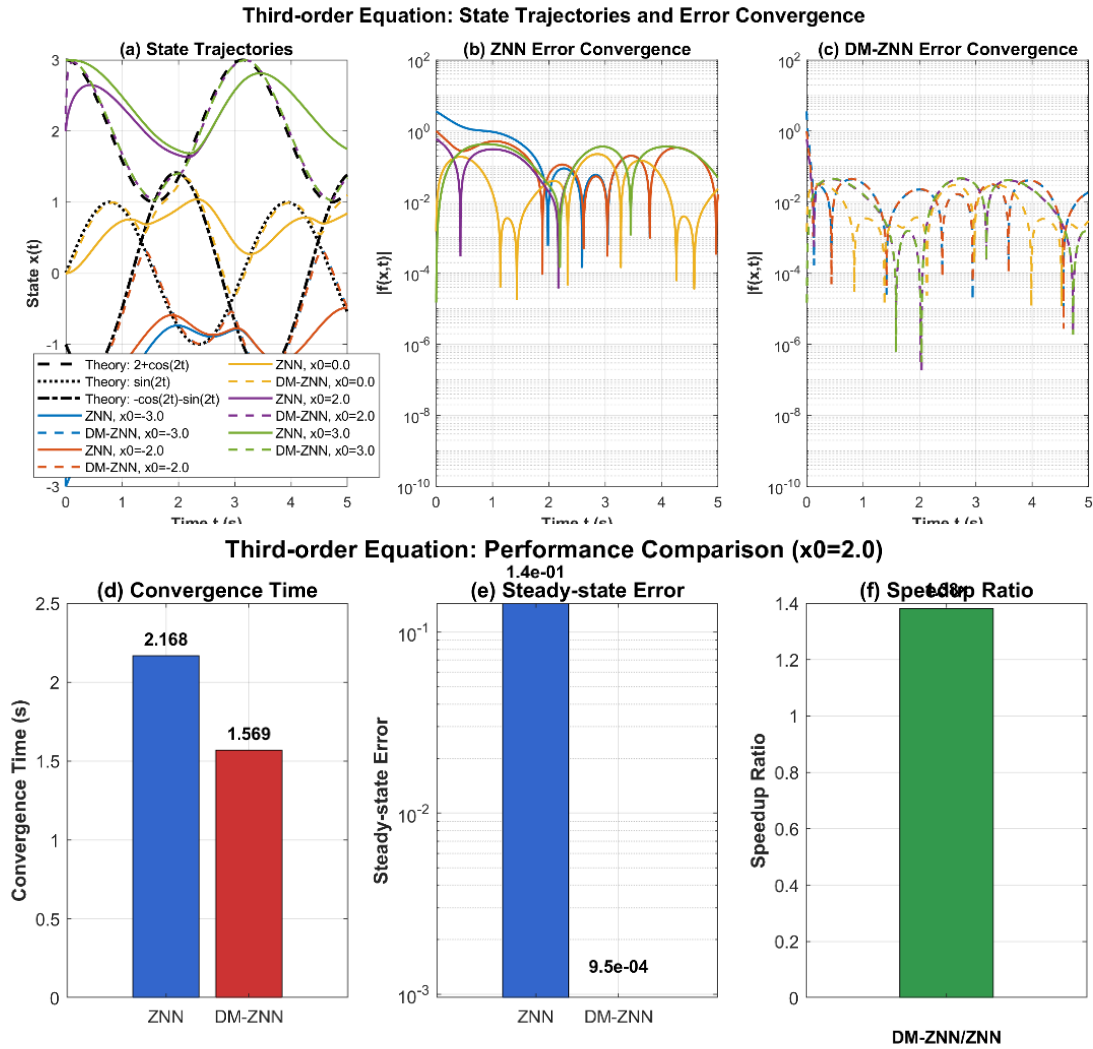


Figure 4. The residual $|f(x(t), t)|$ of SOTVNE in equation (15) for the ZNN and DMZNN models

As demonstrated in the case of traditional ZNN models, significant oscillations and path overlap are exhibited during the solution of high-order problems. Additionally, residuals demonstrate an inability to converge to zero. Conversely, the DMZNN model demonstrated rapid and stable convergence, devoid of oscillations in all instances. Due to space limitations, only the solution data for the fourth-order equation is provided.

In summary, the comparative simulation experiments confirm the efficiency of the proposed DMZNN model for solving time-varying nonlinear equations, particularly high-order ones.

The aim of remote sensing image fusion is to integrate complementary information from multiple images in order to generate a composite image with high spatial and spectral fidelity^[10].

This paper uses the QuickBird remote sensing dataset from the National Cryosphere Desert Data Centre. Subjective evaluation is predicated on visual comparison, while objective evaluation employs the calculation of multiple metrics by referencing reference images. Due to space constraints, the main text only presents a set of typical fusion results (**Figure 5**).

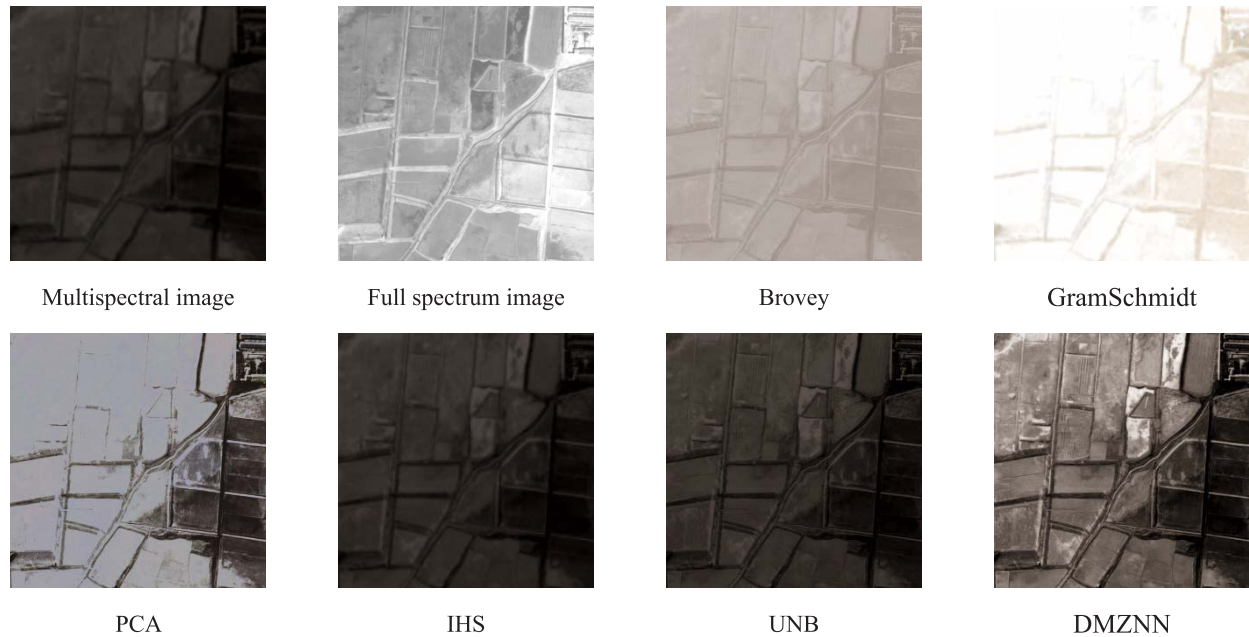


Figure 5 Simulation Experiment Fusion Results Based on the QuickBird Remote Sensing Dataset

The results of the fusion experiment, which are based on the QuickBird remote sensing dataset, are presented in **Table 1**. A review of the extant literature reveals that traditional methods generally exhibit suboptimal performance. While certain methodologies exhibited superior performance in specific metrics, they did not consistently demonstrate effectiveness in overall evaluation indicators. In contrast, the DMZNN method proposed in this paper achieved optimal or near-optimal results across all key metrics, exhibiting comprehensively superior fusion performance.

Table 1. Evaluation of the Fusion Results of the Simulation Experiments Based on the SPOT Remote Sensing Datasets

Method	SSIM ↑	PSNR ↑	CC ↑	ERGAS ↓	SAM ↓	RMSE ↓
Brovey	0.4167	4.82	0.9551	76.76	54.44	17.22
Gram-Schmidt	0.1247	2.59	0.6434	99.18	66.70	19.16
PCA	0.3870	6.75	0.8844	61.97	13.35	11.82
IHS	0.8834	42.31	0.9706	3.71	1.29	9.42
UNB	0.9023	38.32	0.9540	6.59	1.15	9.12
DMZNN	0.8974	41.65	0.9826	3.62	1.04	8.89