Volume 7, Issue 6, 2025

Print ISSN: 2663-1024
Online ISSN: 2663-1016

EURASIA JOURNAL OF SCIENCE AND TECHNOLOGY

Eurasia Journal of Science and Technology

Volume 7, Issue 6, 2025

Published by Upubscience Publisher

Eurasia Journal of Science and Technology
Print ISSN: 2663-1024 Online ISSN: 2663-1016
Email: info@upubscience.com
Website: http://www.upubscience.com/

Table of Content

Eurasia Journal of Science and Technology
Print ISSN: 2663-1024
Online ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3117

INTELLIGENT FOLLOWING CAR BASED ON ULTRASONIC POSITIONING

WenFang Yang*, BuSheng Luo, Guang Luo
Power Supply Service Center, Huizhou Power Supply Bureau, Huizhou 516000, Guangdong, China.
Corresponding Author: WenFang Yang, Email: yangwenfang@gdhz.csg.cn

Abstract: Aiming at the problem of blind area of safety monitoring in electric power operation site, an intelligent following car based on ultrasonic positioning is designed in this paper. The target coordinates are calculated by dual-chip ranging to achieve ± 10 cm accuracy tracking, supporting a maximum identification distance of 10 meters and a dynamic response of 1.5 m/s. The system adopts STM32F103RBT6 main control board, integrates ultrasonic sensor and wireless communication module, and combines with base-station and tag positioning mechanism to solve the problem of monitoring failure caused by manual intervention lag of traditional monitoring equipment. The test shows that the car can last for 4 hours under the condition of slope ≤ 15 ° and load 30 kg, which effectively improves the efficiency of operation safety management and control.
Keywords: Intelligent following; Ultrasonic positioning; Safety monitoring; Electric power operation; Mobile robot

1 INTRODUCTION

In order to further improve the efficiency and accuracy of safety management at the work site, the project team developed an intelligent following visual recognition robot as part of the intelligent safety supervision system, aiming to solve the problem that the existing video surveillance equipment fails to effectively follow the operators and improve the safety supervision level at the work site[1].

In terms of the current research status both at home and abroad, in the field of intelligent small vehicles, many countries have conducted multi-angle research. For instance, foreign researchers have explored extensively in aspects such as automatic control, environmental perception, and path planning of intelligent small vehicles. Some research institutions and universities have conducted in-depth studies on positioning, navigation, and obstacle avoidance of intelligent small vehicles, proposing various algorithms and system design schemes. Additionally, significant progress has been made in hardware design, control algorithms, and sensor technology of intelligent small vehicles abroad. For example, designs such as intelligent clothes drying racks and intelligent curtain control systems based on STM32 microcontrollers have demonstrated the application prospects of intelligent small vehicles in smart homes and industrial automation.

Domestically, significant progress has also been made in the research of intelligent small vehicles. For instance, the design of intelligent following small vehicles based on ultrasonic positioning, which uses ultrasonic sensors to measure the distance to the target object and achieve positioning, thereby realizing the following function. This system mainly includes sensor modules, control modules, and execution modules, featuring a simple structure, low cost, and fast response, suitable for automatic following tasks in indoor environments. Moreover, designs such as intelligent clothes drying racks and intelligent curtain control systems based on STM32 microcontrollers have demonstrated the application prospects of intelligent small vehicles in smart homes and industrial automation. Additionally, the design and implementation of intelligent following small vehicles based on STM32 microcontrollers, through the design of ultrasonic positioning modules, motor drive modules, and LCD, etc., have achieved the automatic following function of the vehicle[2].

In terms of the hardware design of intelligent following small vehicles, researchers at home and abroad have proposed various design schemes. For example, the design of intelligent following small vehicles based on ultrasonic positioning adopts a four-wheel movement mode, controlled by stepper motors, and uses infrared distance sensors to achieve anti-collision functions. A mathematical model for tracking the target position was established, and the hardware design of the control system was completed, followed by experiments. The experimental results show that the vehicle can effectively follow the target within 5 meters and well achieve the anti-collision function. Additionally, the design and implementation of intelligent following small vehicles based on STM32 microcontrollers, through the design of ultrasonic positioning modules, motor drive modules, and LCD, etc., have achieved the automatic following function of the vehicle[3-4].

In terms of the control algorithms of intelligent following small vehicles, researchers at home and abroad have proposed various algorithms and system design schemes. For instance, the automatic following algorithm of intelligent small vehicles based on ultrasonic positioning measures the distance to the target object in real time through three ultrasonic modules, and determines the target position by combining the distance difference comparison method and the plane coordinate method, thereby achieving automatic following. At the same time, the vehicle is equipped with an infrared obstacle avoidance module, which can automatically bypass obstacles. A servo is used to scan the ultrasonic module to expand the coverage range. The system uses Arduino uno for servo control and realizes speed regulation and alarm functions through software programming. Additionally, the design and implementation of intelligent following small

vehicles based on STM32 microcontrollers, through the design of ultrasonic positioning modules, motor drive modules, and LCD, etc., have achieved the automatic following function of the vehicle.

In terms of the application fields of intelligent following small vehicles, researchers at home and abroad have proposed various application scenarios. For example, the design of intelligent following small vehicles based on ultrasonic positioning is suitable for the needs of automated material handling in industrial, commercial, and elderly assistance scenarios, improving work efficiency and safety. Additionally, the design and implementation of intelligent following small vehicles based on STM32 microcontrollers are applicable to education, industrial automation, and smart home fields, which can be used as teaching cases to enhance practical abilities and also for material transportation or monitoring[5].

In terms of the challenges and development trends of intelligent following small vehicles, researchers at home and abroad have proposed various challenges and solutions. For instance, the current automatic following carts mostly adopt active positioning technology, which requires the target to be followed to carry a signal transmitter. This brings about problems such as inconvenience in application and high cost, making it difficult to be widely adopted. This paper proposes a positioning and obstacle recognition method based on ultrasonic array, analyzes the possible situations in front of the cart and formulates control strategies to solve the problem of ultrasonic ranging error. In addition, the design and implementation of an intelligent following cart based on STM32 is presented. Through the design of functional modules such as ultrasonic positioning module, motor drive module and LCD, the automatic following function of the cart is realized.

2 TECHNICAL RESEARCH PROGRAMME

2.1 Positioning Process

The base station performs two times of ranging with the tag through two chips, establishes an equation according to the time phase difference and distance of the two times of ranging, calculates the X and y coordinates and distance values of the tag relative to the base station, and calculates the X and y coordinate values, which is equivalent to knowing the angle deviation from the tag to the base station. Taking the rectangular coordinate axis as an example, from the front direction of the base station, the negative half axis of the X axis is the right side of the base station, and the positive half axis is the left side of the base station. The positive half axis of the Y axis is the front of the base station, and the back of the base station cannot get the correct coordinates, see Figure 1.

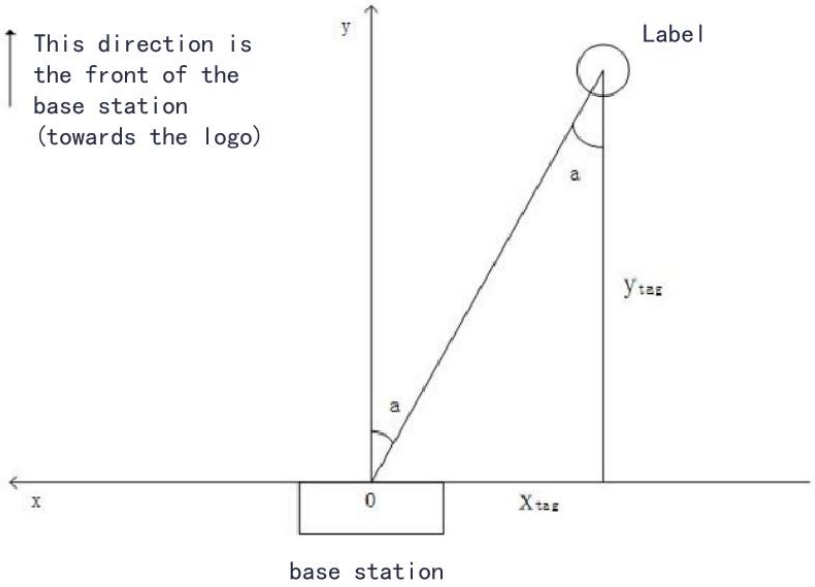


Figure 1 Analytic Diagram of Base Station Coordinate

2.2 Label List

A label list (Knownlist, referred to as K list) is stored inside the base station. The K list stores information such as IDs of tags with which the base station is to perform ranging and positioning. That is to say, the base station will only perform ranging and positioning with the tags stored in the K list. When the base station detects a tag, the base station firstly compares the ID of the tag with the ID existing in the K list of the base station, if the ID does not exist in the K list, the base station actively reports the information (NewTag) that a new tag is monitored once, and then the base station is required to send a command (AddTag) to the base station through a serial port from the outside. This ID information is stored in the K list of the base station. After successful addition, the base station will start ranging and positioning with the tag. The user can issue a Save command to save the current configuration of the base station to Flash, and the information in the K list can still exist after the power is turned off and the power is turned on again. If

the user powers off the base station without issuing a Save command, the tag needs to be powered on again when the base station locates the tag next time, see Figure 2.

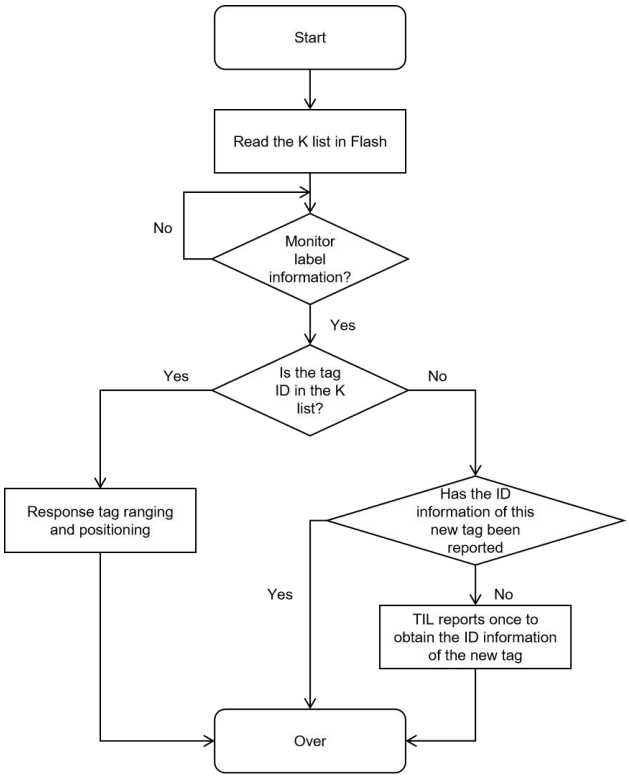


Figure 2 Flow chart of base station positioning

2.3 Description of Control Mode

The control board will always monitor the new tag information reported by the base station. If the base station is monitored to report new tag information, the base station will be ordered to add the tag for positioning. It will not send a message to the base station to save the tag, and after responding to a following tag, it will not respond to the newly powered tag.

The method comprises the following steps of: using a label which is not added to a base station K list; electrifying the trolley, and then electrifying the label; After hearing the buzzer, the representative has been added to the tag. Each successful communication will cause LED3 to flash. At this point, the car has begun to follow the label.

3 Conclusion and Prospect

3.1 Conclusion

1. The robot can automatically follow the person in charge of the work or the special guardian wearing the controller according to the needs of the work site, which solves the problem that the traditional video surveillance equipment can not move in time due to the transfer of the staff. The realization of this function relies on acoustic wave technology, which can accurately track the movement of workers wearing acoustic wave remote control machine and maintain the whole monitoring of the whole operation site.

2. Robots can implement real-time safety supervision covering 360 degrees of the operation site and the whole process to ensure that every work link is monitored and avoid potential safety hazards due to human negligence or inadequate monitoring equipment. Combined with intelligent safety supervision system, robots can become an important part of real-time supervision, thus improving the standardization and scientificity of safety production.

3.2 Prospect

As an important part of the on-site intelligent safety supervision and management system, the intelligent following visual recognition robot will greatly improve the safety management and control efficiency of the operation site. With the continuous development of technology, the application of this intelligent equipment will have a far-reaching impact on the safety supervision of all walks of life, and help to create a safer and more standardized working environment.

The development of intelligent small vehicles also faces some challenges. For instance, high technical costs, incomplete regulations, and low consumer acceptance have to some extent restricted the popularization and application of intelligent small vehicles. Additionally, the market development of intelligent small vehicles also encounters some challenges, such as high technical costs, incomplete regulations, and low consumer acceptance. These issues have to

some extent restricted the popularization and application of intelligent small vehicles. In the future, with the further maturation of technology and the in-depth expansion of the market, these problems are expected to be gradually solved.

In the future, the development of intelligent small vehicles will present multiple trends and directions, which are not only reflected in the technical aspect but also in the market, application, and policy. With the continuous progress of artificial intelligence, the Internet of Things, 5G communication, and other technologies, intelligent small vehicles will demonstrate their unique value and potential in more fields.

In the technical aspect, the development of intelligent small vehicles will rely more on the deep integration of artificial intelligence, big data, and cloud computing. The application of artificial intelligence technology will enable intelligent small vehicles to have stronger autonomous learning and decision-making capabilities, and better adapt to complex and changeable environments. For example, through deep learning and machine learning algorithms, intelligent small vehicles can continuously optimize their path planning, obstacle avoidance, and navigation capabilities, improving their adaptability and safety in complex scenarios. At the same time, with the popularization of 5G communication technology, intelligent small vehicles will achieve more efficient communication and data transmission, enhancing their capabilities in real-time data processing and remote control.

In the application field, the application scope of intelligent small vehicles will further expand. Currently, intelligent small vehicles are widely used in logistics transportation, autonomous driving, security patrol, medical diagnosis, industrial automation, and other fields. In the future, with the continuous progress of technology and the reduction of costs, intelligent small vehicles will find application scenarios in more emerging fields, such as smart cities, smart homes, and agricultural monitoring. Moreover, the combination of intelligent small vehicles and the Internet of Things will promote their application in smart homes and intelligent transportation systems, achieving a more efficient and convenient lifestyle.

In the market and industry aspect, the intelligent small vehicle market will face more intense competition. With the continuous progress of technology and the expansion of the market, more and more enterprises will enter this field, promoting the innovation of intelligent small vehicle technology and the diversification of products. At the same time, the government and enterprises will increase their investment in the research and development of intelligent small vehicles, promoting their application in more fields. Additionally, with the continuous improvement of policies and the gradual establishment of laws and regulations, the market environment for intelligent small vehicles will become more standardized and orderly.

In terms of challenges, the development of intelligent small vehicles also faces some challenges. For instance, high technical costs, incomplete regulations, and low consumer acceptance have to some extent restricted the popularization and application of intelligent small vehicles. Moreover, the safety and privacy protection issues of intelligent small vehicles also need to be urgently addressed. Therefore, the future development requires joint efforts in technological innovation, policy support, and market regulation to promote the healthy development of intelligent small vehicle technology.

In the future, the development of intelligent small vehicles will move towards a more intelligent, networked, collaborative, and secure direction. With the continuous progress of technology and the expansion of the market, intelligent small vehicles will demonstrate their huge potential and value in more fields, bringing more convenience and innovation to society[6-8].

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Yan R, Li P, Gao H, et al. Car-following strategy of intelligent connected vehicle using extended disturbance observer adjusted by reinforcement learning. CAAI Transactions on Intelligence Technology, 2023, 9(2): 365-373. DOI: 10.1049/CIT2.12252.

[2] Song Y C, Xie Z J. Control method of three-wheel intelligent tracking logistics car based on OpenMV4. Journal of Physics: Conference Series, 2022, 2246(1): 1-5. DOI: 10.1088/1742-6596/2246/1/012033.

[3] Liu L, He D, Zheng J, et al. Design and implementation of intelligent tracking car based on machine vision. Journal of Intelligent & Fuzzy Systems, 2020, 38(5): 5799-5810. DOI: 10.3233/JIFS-179667.

[4] Xiao J C, Yin Y X, Zhang Y Z. A car following model considering driver memory effect in intelligent transportation environment. Chongqing Vocational College of Transportation (China), 2024, 13018, 130180R-130180R-7. DOI: 10.1117/12.3024023.

[5] Zhang C L, Zhang Y Z. Design of a car-following driving system based on machine vision.Internet of Things Technologies, 2024, 14(03): 118-119+122. DOI: 10.16667/j.issn.2095-1302.2024.03.028.

[6] Jian H. Research on indoor active obstacle avoidance of intelligent following car.Chongqing University of Technology, 2025(3). DOI: 10.27753/d.cnki.gcqgx.2024.001206.

[7] Zhou N X. Research on intelligent vehicle platoon control in warehouse environment based on UWB positioning. Zhejiang University, 2025(5). DOI: 10.27461/d.cnki.gzjdx.2024.000854.

[8] Dong J H, Cai X Y, Deng J Y, et al. Intelligent car following driving system based on MSP430. Science and Technology & Innovation, 2024(04): 56-58. DOI: 10.15913/j.cnki.kjycx.2024.04.014.

Eurasia Journal of Science and Technology
Print ISSN: 2663-1024
Online ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3118

RAIL DAMAGE DETECTION SYSTEM BASED ON FUSION OF PLANAR CAPACITIVE SENSING AND VISION

YuYang Xia, ZhongFu Liu*
School of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, Liaoning, China.
Corresponding Author: ZhongFu Liu, Email: lzhongfu@dlnu.edu.cn

Abstract: This research addresses the specific and crucial application scenario of railway track damage detection by designing and implementing a comprehensive intelligent detection system solution. At the hardware architecture level, the system meticulously selects the high-performance STM32F103 microcontroller as the core control unit. This unit offers fast processing speed, low power consumption, and high reliability, laying a solid foundation for the system's stable operation. Centered around this core control unit, the system integrates various functional modules, including high-precision detection modules like planar capacitive sensors for real-time and accurate collection of track status data. Simultaneously, the OpenMV visual processing module is utilized to achieve precise machine vision recognition of track surface defects, significantly enhancing the intuitiveness and accuracy of detection. Additionally, the system is equipped with a motor drive module to enable precise movement of the detection device. The system not only features excellent hardware configuration but also possesses powerful data processing capabilities, enabling real-time and efficient analysis and processing of the collected massive data, thereby significantly improving detection accuracy and timeliness. In terms of material selection, the system innovatively adopts new energy-saving and environmentally friendly materials, which not only substantially reduce the overall energy consumption of the system but also effectively extend the equipment's service life, markedly enhancing its durability. This system integrates numerous advantages such as efficient detection, energy saving, environmental protection, and low-cost operation, demonstrating broad application prospects and significant market promotion potential in the field of railway operation and maintenance. It is expected to provide a solid and powerful guarantee for the safe and stable operation of railways.
Keywords: Microcontroller; Sensor; Rail damage; Planar capacitance; High efficiency and energy saving; OpenMV visual processing

1 INTRODUCTION

In recent years, China's rail transportation industry has achieved rapid development. High-speed rail networks crisscross the country, and subway lines extend in all directions. These modern rail transportation facilities have deeply integrated into our daily lives, becoming the preferred mode of transportation for people. At the same time, as a critical national infrastructure, rail transportation plays an irreplaceable key role in promoting regional economic development, driving industrial upgrading, and stimulating employment. However, with the continuous growth of rail transit operational mileage and the rapid increase in transport volume, the track system faces increasingly severe challenges. Due to the long-term high-speed operation of trains, intense friction occurs between the wheels and rails. Coupled with the tracks' long-term exposure to complex and variable external environments such as wind, sun, rain, and snow erosion, a series of safety hazards requiring urgent solutions arise. For instance, minor cracks may appear on the rail surface or internally, the track geometry may deform, or uneven wear may occur on the rail contact surface. These seemingly minor damage issues, if not detected and addressed promptly, will seriously affect the train's running stability. In mild cases, this leads to increased train vibration and reduced passenger comfort; in severe cases, it may trigger major safety accidents such as train derailments, resulting in casualties, property losses, and even causing a chain reaction of social problems. In response to this situation,various rail damage detection technologies have been developed. Zhang et al. proposed an enhanced blind separation method for rail defect signals using a time-frequency separation neural network and smoothed pseudo Wigner-Ville distribution [1]. Mao et al. introduced a novel similarity measure based on a dispersion-transition matrix and Jensen-Fisher divergence for detecting rail short-wave defects [2]. Gong et al. optimized magnetic flux leakage detection equipment for rail surface inspection through finite element simulation [3]. Ding et al. developed a rail defect detection framework under class-imbalanced conditions based on an improved YOLO network [4]. Additionally, the Transportation Technology Center expanded the rail flaw library to enhance rail integrity [5], while Chang et al. proposed a defect detection method for ferromagnetic rails using an EMAE-based peak-to-peak method and confidence probability indicator [6]. Santur et al. adopted 3D laser cameras combined with ResNet50 for rail surface defect detection [7], and Zhao et al. applied the hybrid laser ultrasonic method in rail inspection [8]. Drawing on these research foundations, this design innovatively proposes a rail dama this design innovatively proposes a rail damage detection system solution based on the fusion of planar capacitive sensing and vision. Through multi-sensor data fusion and intelligent analysis algorithms, this system can achieve precise detection and intelligent diagnosis of rail damage, offering significant advantages such as high detection accuracy, fast response speed, and strong adaptability. It will provide robust technical support for the promotion and application of rail damage detection technology.

2 SYSTEM OVERALL DESIGN SCHEME

In this system design, the system primarily uses the STM32F103C8T6 as the main control board. Planar capacitive sensors serve as the acquisition module to collect and process information related to track damage. The collected relevant information is uploaded to a WeChat mini-program. Within the mini-program, a pre-trained rail damage detection system is used to perform preprocessing, feature extraction, and classification of the rail damage severity. The condition of the rail damage is then displayed on the WeChat mini-program interface, facilitating real-time understanding of the rail damage status by relevant personnel. Using a 4G module, the collected rail damage information can achieve data communication between the device end and the WeChat mini-program via a serial port. If the rail condition is damaged, relevant personnel can use the information viewing function pushed by the mini-program to understand the rail damage situation and make corresponding judgments, thereby enabling more convenient and accurate detection of rail damage. Simultaneously, relevant personnel, after authorizing login via WeChat to the mini-program, can encrypt and manage user information, protecting user privacy rights and enhancing the security and credibility of user data. The overall system framework structure diagram is shown in Figure 1.

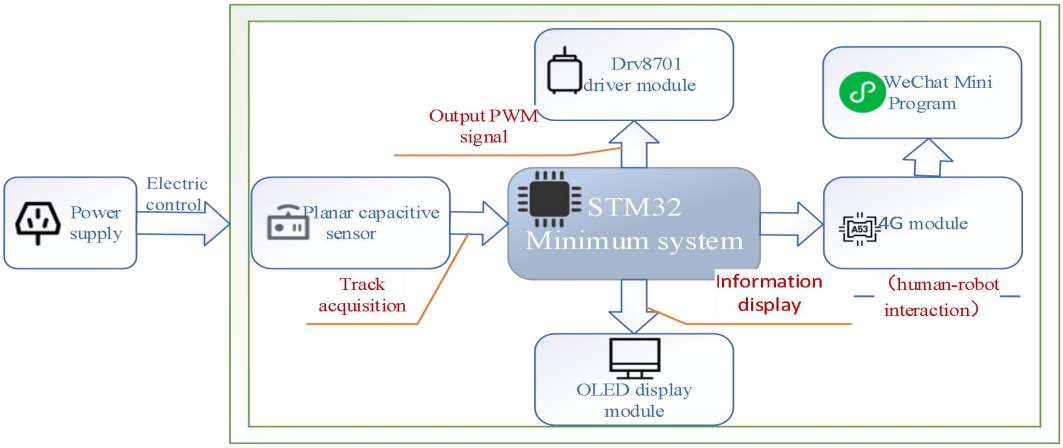


Figure 1 Overall System Framework Diagram

The rail damage detection car based on the BiFPN-YOLOv8 algorithm model and the WPD-CNN compensation capacitor fault diagnosis method deeply integrates sensor technology, embedded technology, and wireless communication technology with a WeChat mini-program to build an intelligent Internet of Things (IoT) system. This system consists of three main structural layers: the perception layer, the transmission layer, and the application layer. The perception layer is composed of the STM32 main controller and planar capacitive sensors; the transmission layer consists of the 4G module; and the application layer is the WeChat mini-program end. Relevant personnel at the application layer receive the rail damage data acquired by the perception layer and uploaded by the transmission layer. Subsequently, they use the WeChat mini-program's information recognition and analysis functions to better judge the damage status of the rails, thereby achieving accurate identification and real-time monitoring of rail damage. The overall system architecture block diagram is shown in Figure 2.

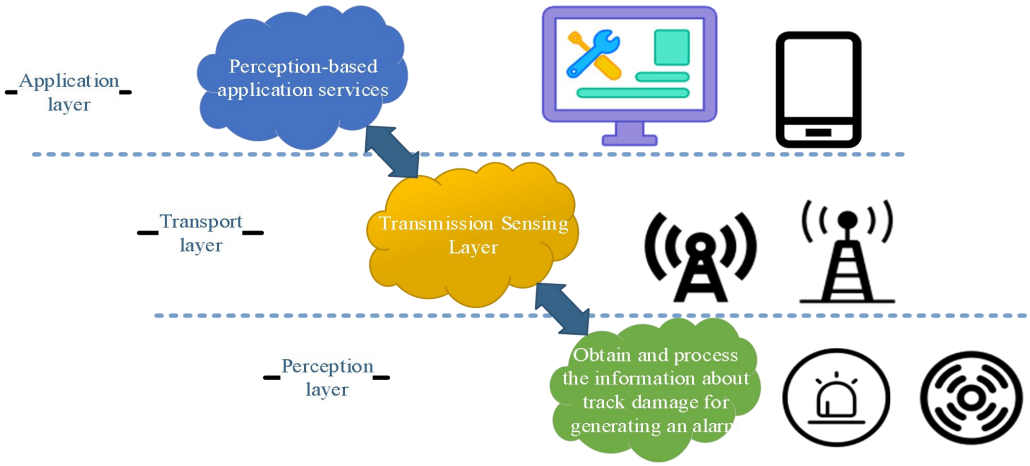
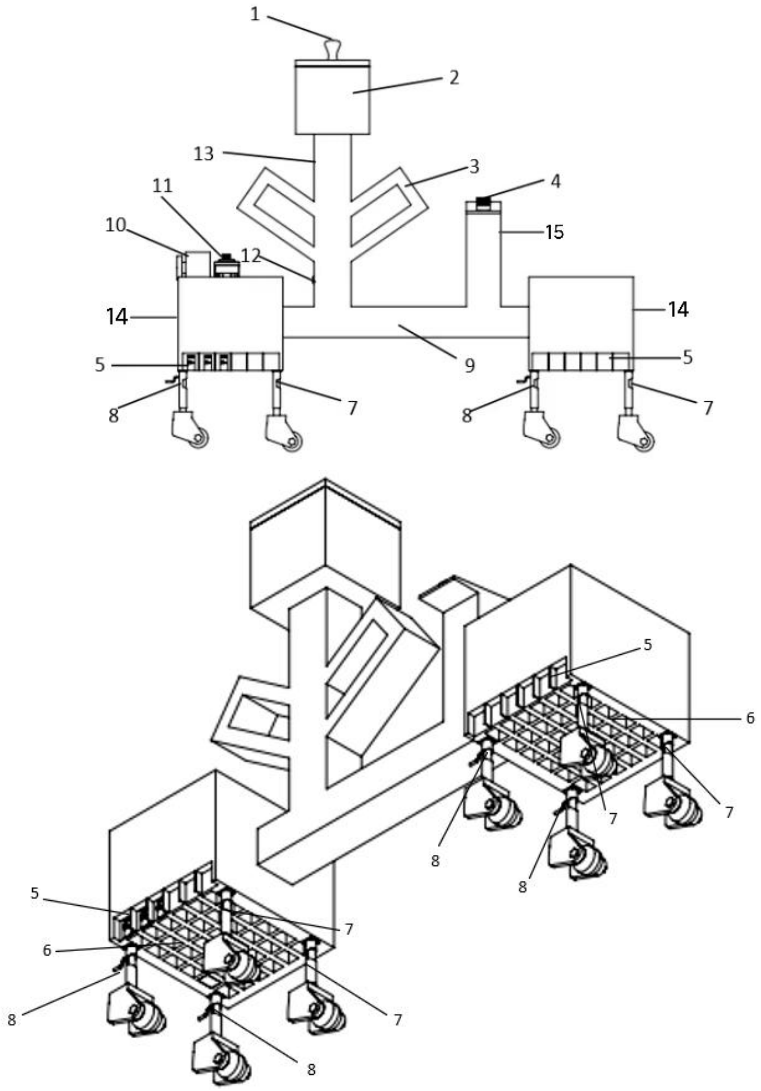


Figure 2 System Overall Architecture Diagram

The system model design is shown in Figure 3. The entire system consists of two 6x6 planar capacitive sensor matrices, two storage racks placed on connecting rods, two fixed wheels and two manually adjustable wheels, one OLED display, etc. The outer skin of the car is made of zinc-iron sheet with strong anti-rust capability, allowing the car to adapt to

various conditions. The front bottom of the car also has a row of tracking sensors, enabling the car to better control its direction. The bottom ends on both sides of the car feature two fixed wheels and two manually adjustable wheels, which help the car adapt to different rail conditions, making detection more comprehensive and efficient.



1.Ellipsoidal alarm; 2. OLED display; 3. Storage rack; 4. Control terminal; 5. Infrared sensor; 6. Capacitive sensor matrix; 7. Wheel; 8. Hand-cranked lifting wheel; 9. Connecting rod one; 10. Servo motor; 11. OpenMV camera; 12. Rechargeable power module; 13. Connecting rod two; 14. Car body; 15. Connecting rod three.

Figure 3 System Model Design Diagram

3 HARDWARE OVERALL DESIGN

The system control circuit consists of electrical components such as the OLED display module, planar capacitive sensors, and tracking sensors.

3.1 Microcontroller Circuit

This system selects the STM32F103C8T6 produced by STMicroelectronics as the main control chip. This is a high-performance 32-bit enhanced microcontroller. As a representative product of the STM32F1 series, this microcontroller adopts the ARM Cortex-M3 core architecture, with a maximum operating frequency of 72 MHz (Note: Corrected from potentially erroneous "2MHz" in original, standard for Cortex-M3 in STM32F103 is up to 72MHz), capable of meeting real-time data processing requirements. In terms of storage resources, the chip has built-in 64 KB of SRAM for data caching and is equipped with 256 KB of Flash memory for program storage. These resource configurations provide the hardware foundation for implementing complex control algorithms. It is particularly worth mentioning that this microcontroller possesses relatively strong image processing capabilities, enabling it to handle video data acquisition and processing tasks. To ensure the stable operation of the microcontroller, the system designs a complete minimum system circuit, including essential peripheral circuits such as the crystal oscillator circuit, reset circuit, and power supply circuit. Its specific circuit connection method is shown in Figure 4. This minimum system circuit provides the basic working environment for the microcontroller and is the foundation for subsequent functional

expansion.

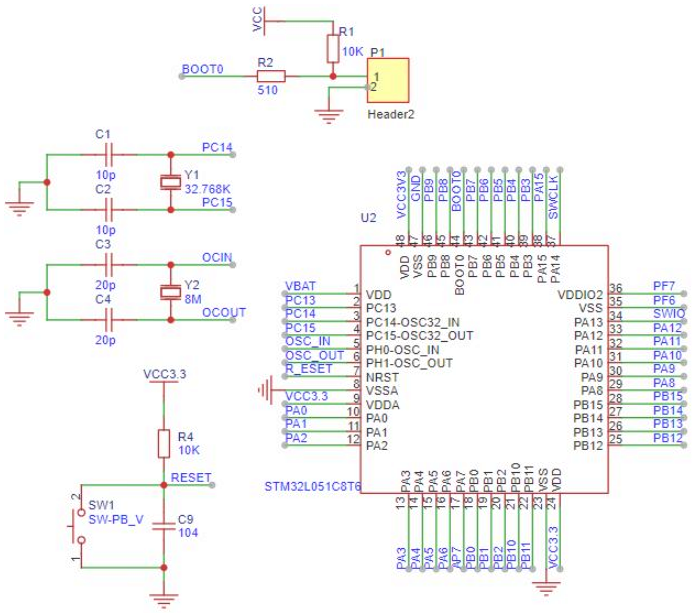


Figure 4 STM32F103C8T6 Minimum System Circuit Diagram

3.2 Power Module Circuit

This protection circuit uses the widely adopted TP4056 charging management chip as the core control device. This chip features stable performance and high reliability. The entire peripheral circuit design is concise and efficient, requiring only a few components to achieve complete lithium battery charging protection functions. It can effectively prevent overcharging, over-discharging, overcurrent, and other abnormal conditions during the charging process, thereby ensuring charging safety and battery lifespan. The circuit uses a standard +5V external power supply. The positive terminal of the lithium battery is connected to pin 5 of the TP4056 chip, and the negative terminal is connected to the GND of the chip. Regarding charging status indication, the red LED indicator remains steadily lit when the battery is charging. When the battery is fully charged, the green LED indicator automatically lights up. This intuitive indicator design facilitates users to grasp the charging status at any time. The PROG pin (pin 2) of the chip is used for charging current detection; the charging current can be precisely set by an external resistor. The TEMP pin (pin 1) is the battery temperature detection terminal. In this design, this pin is directly connected to GND, simplifying the circuit design and making the circuit more practical. The complete circuit connection schematic is detailed in Figure 5 of this paper, which clearly shows the connection relationships and working principles of the components.

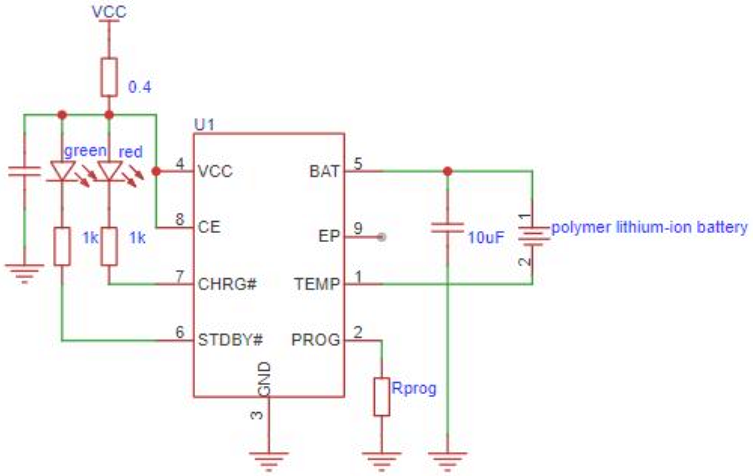


Figure 5 Power Module Circuit

3.3 Detection Sensor Circuit

Working principle of planar capacitive sensor detection is A planar capacitor consists of two parallel plates. The capacitance value follows the formula $C = \frac{\varepsilon S}{d}$, where ε is the dielectric constant of the medium between the plates, S is

the opposing area of the plates, and d is the distance between the plates. When track damage occurs, it indirectly changes one or more of these three parameters, causing a change in capacitance. The detection circuit converts the capacitance change into a readable electrical signal, which is then used to judge the track damage condition. If the track deforms or settles, it causes the distance d between the sensor plates to change; for example, track subsidence brings the plates closer, d decreases, capacitance increases. If cracks or wear appear on the track, it may cause misalignment of the plate installation position, changing the opposing area S; for example, cracks cause plate misalignment, S decreases, capacitance decreases. If impurities enter the damaged area of the track, they change the equivalent dielectric constant of the medium between the plates, also causing capacitance changes.

3.4 OLED Display Circuit

This circuit uses a 3.3V DC power supply. To ensure power stability, a filter capacitor C9 is set in the power supply loop to effectively filter out high-frequency noise and interference signals from the power source. Simultaneously, a voltage divider circuit composed of resistors R8 and R9 allows for precise regulation and stable control of the operating voltage. The OLED display uses the I²C communication protocol for data interaction with the main control chip. The SCL clock signal line is connected to the main control's IO22 pin, and the SDA data signal line is connected to the IO21 pin. When the system detects track damage, the main control chip sends the damage data to the OLED display via the I²C bus. The dedicated driver circuit integrated within the OLED then accurately controls the light emission state of each pixel based on the received data information, clearly displaying the specific location and severity of the track damage. The complete circuit connection method and component layout design of the entire OLED display module are detailed in the circuit schematic shown in Figure 6.

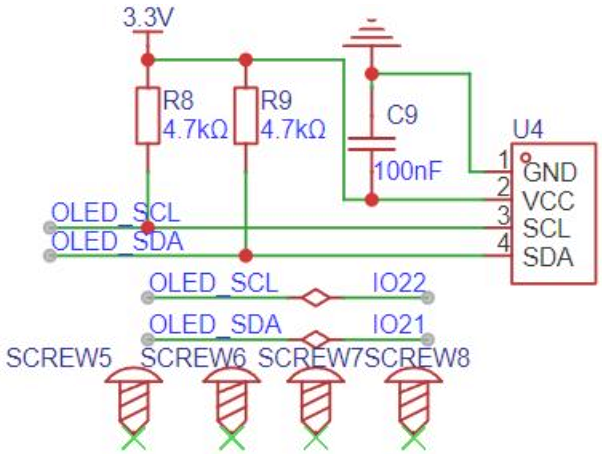


Figure 6 OLED Display Circuit Diagram

3.5 Alarm Module Circuit

In the daily maintenance of rail transit, the rail damage car plays a crucial role. It is equipped with an advanced alarm module circuit. When the car is running on the track and detects track damage, it quickly triggers the alarm mechanism. Upon receiving the signal, the precise vibration device inside the alarm immediately activates, generating an audible alarm through its internal vibration, promptly alerting surrounding personnel. The alarm module circuit diagram is shown in Figure 7, which details the circuit connections and working principle of the entire alarm system, providing a clear and accurate reference for technicians performing system maintenance and troubleshooting.

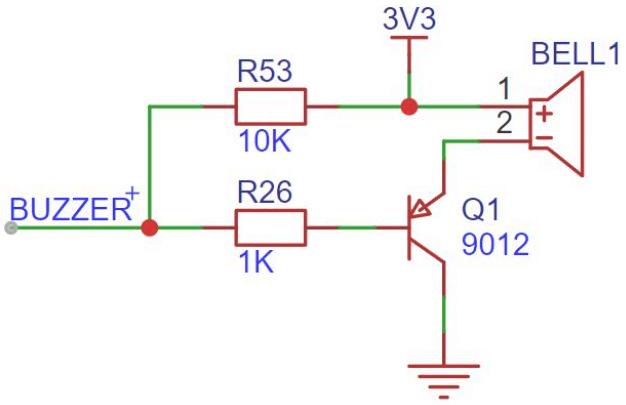


Figure 7 Alarm Module Circuit Diagram

4 SOFTWARE OVERALL DESIGN

In the software architecture developed for this project, we adopted a multi-language collaborative development strategy. Specifically, in the visual processing module part, we chose Python as the primary development language. Based on the PyQt5 framework, we conducted in-depth programming development for the OpenMV embedded vision module, achieving efficient image acquisition and processing functions. Simultaneously, by introducing the YOLOv8 algorithm model and leveraging the hardware acceleration capability of the psc module (Note: Assuming 'psc' refers to a specific hardware accelerator, possibly a typo or project-specific term), the performance of OpenMV in object detection and recognition was significantly enhanced. In the underlying control module aspect, we used C language for firmware development on the STM32 microcontroller under the Keil5 integrated development environment, achieving precise motor control and sensor data processing. Furthermore, we integrated advanced planar capacitive sensor technology, combined with the WPD-CNN algorithm, greatly enhancing the STM32's capabilities in non-contact measurement and environmental perception. Finally, through communication protocols and system integration solutions, these functional modules work together to drive the track inspection car to achieve multiple intelligent monitoring functions, including track identification, obstacle detection, and automatic navigation, see Figure 8.

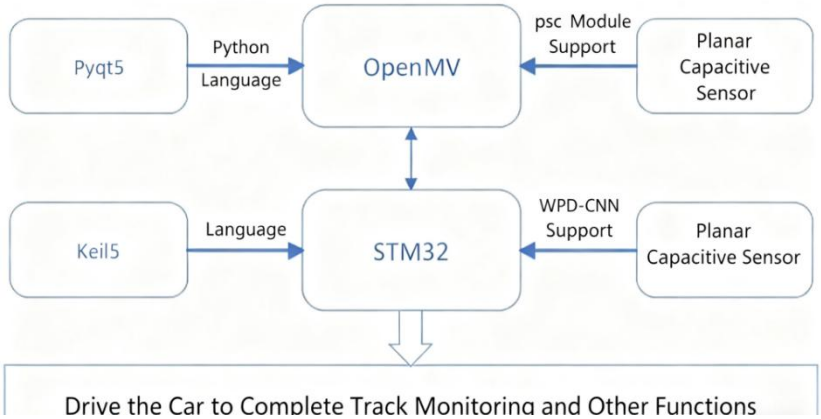


Figure 8 Software Overall Framework

4.1 YOLOv8n Algorithm Improvement

In the original C2f module of YOLOv8n, due to the fixed size and stride of the convolution kernels, the C2f module could only capture local image information and could not fully utilize the entire image's information to improve feature representation and detection accuracy. This limited receptive field problem negatively impacted the performance of the object detection model. In rail damage detection, the shape and size of damage vary widely, and there might be occlusion by foreign objects. To better understand the structure and features of rail damage, the model needs a broader receptive field and stronger contextual perception capabilities to improve detection accuracy and robustness. To solve this problem, this paper reconstructs this module to improve its ability to capture geometric deformations of damage in rail damage images. Specifically, Dynamic Snake Convolution is used to replace the ordinary convolution operation in the original C2f module. Dynamic Snake Convolution can generate convolution kernels of different shapes and sizes according to the shape and size of the rail damage, enhancing the receptive field and feature extraction capability of the C2f module. Meanwhile, after introducing Dynamic Snake Convolution into the C2f module, the shape information of the rail damage can be integrated into the convolution operation, enhancing the expressive power and discriminability of rail damage features, and improving the performance and adaptability of the object detection model. The improved part of the C2f module is shown in Figure 9.

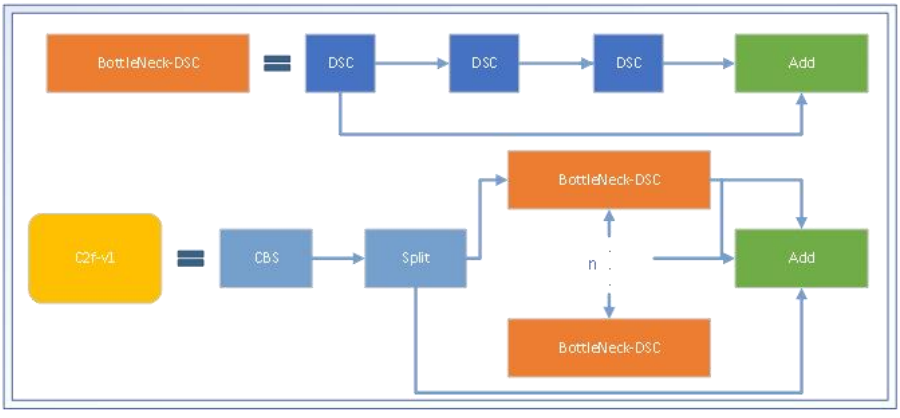


Figure 9 C2f-v1 Module Based on Dynamic Snake Convolution

4.2 WPD-CNN Compensation Capacitance Method

To effectively extract compensation capacitor feature information and achieve effective identification of compensation capacitor faults at different track locations, the WPD-CNN method is adopted for fault diagnosis of compensation capacitors. The overall approach is shown in Figure 10. In the dataset construction stage, power spectrum analysis is used to find the frequency band range that can effectively reflect the characteristics of the compensation capacitor. Then, the wavelet packet transform is used to decompose the original signal, extracting the low-frequency trend component and the wavelet packet coefficients corresponding to the characteristic frequency band of the compensation capacitor to form a feature matrix. In the model training stage, the generated training dataset is input into the CNN network for learning. The model is optimized through the backpropagation algorithm to obtain a model with relatively good performance, which is then validated on the test set. Finally, the trained model is used to achieve fault diagnosis for the compensation capacitors of the rail detection car.

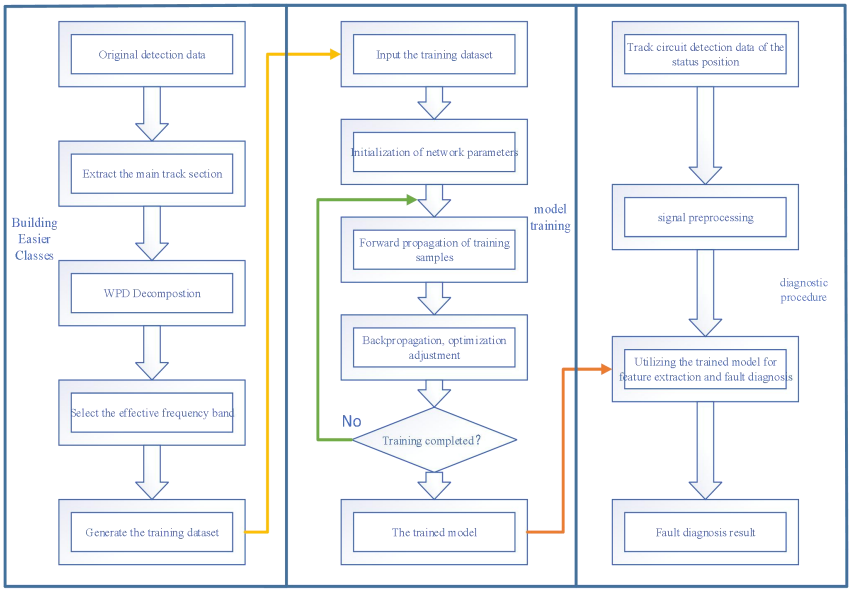


Figure 10 Overall Approach of Compensation Capacitor Diagnosis Method Based on WPD-CNN

5 FUNCTION TEST

Data provided according to simulation experiments are shown in Table 1.

Table 1 Simulation Experiment Data Table

Research Subject	Urban/Rural Rail	Provincial Capital	Detection Rail Efficiency
Group 1	Rail Crack	Rail Breakage	92%
Group 2	Rail Deformation	Rail Wear	95%
Group 3	Rail Wear	Rail Rust	93%
Average Detection Efficiency			94%

As this experiment did not consider other external conditions, the damage conditions for both urban/rural rails and provincial capital rails were mostly limited to rail cracks, rail deformation, rail breakage, and rail rust. Testing with this system revealed that the most common rail damage conditions were rail wear and rail deformation. The average detection efficiency was 94%.

6 CONCLUSION

This project's research innovatively integrates planar capacitive sensing technology and machine vision detection technology to construct an intelligent rail damage detection system. The core of the system lies in using a 6x6 planar capacitive sensor matrix to achieve non-contact identification of internal rail defects through changes in the electric field distribution. Simultaneously, paired with the OpenMV high-definition camera module, it can accurately capture various types of damage features on the rail surface. The system employs the STM32F4 series high-performance microcontroller as the main control chip (Note: Original text mentioned F4 here, but F103 earlier; kept as in original), achieving intelligent detection functions through optimized circuit design, effectively reducing the labor intensity and energy consumption of manual inspections. It is worth mentioning that the system is equipped with a hand-cranked adjustable lifting wheel mechanism, enabling flexible adaptation to track detection needs at different heights. This largely addresses the industry pain points of traditional detection equipment, such as insufficient precision and poor

adaptability. This detection system has wide applicability and can be used in various rail transit scenarios such as high-speed rail, subways, and intercity railways. The main innovations of the system are reflected in three aspects. First, it achieves a composite working mode combining intelligent automatic detection and manual adjustment. Second, it innovatively adopts a dual-algorithm fusion architecture comprising the BiFPN-YOLOv8 object detection algorithm and the WPD-CNN feature extraction algorithm, significantly improving the accuracy of rail damage identification. Finally, the system adopts a modular design, where core components such as capacitive sensors support regular calibration or quick replacement, ensuring long-term reliability. The entire system has significant advantages such as high cost-effectiveness, strong portability, and easy maintenance, demonstrating broad application prospects and market promotion value in the field of rail transit operation and maintenance.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

We are grateful for the financial support provided by the Sunbird Project of Dalian Minzu University(TYN2025761), the Innovation and Entrepreneurship Training Program for College Students in Liaoning Province(202512026184).

REFERENCES

[1] Zhang M X, Wang K W, Yang Y L, et al. Enhanced Blind Separation of Rail Defect Signals with Time–Frequency Separation Neural Network and Smoothed Pseudo Wigner–Ville Distribution. Applied Sciences, 2025, 15 (7): 3546. DOI: 10.3390/APP15073546.

[2] Mao X G, Xia C L, Liu J Z, et al. A novel similarity measure based on dispersion-transition matrix and Jensen–Fisher divergence and its application on the detection of rail short-wave defects. Chaos, Solitons and Fractals: the interdisciplinary journal of Nonlinear Science, and Nonequilibrium and Complex Phenomena, 2025, 192, 115988. DOI: 10.1016/J.CHAOS.2025.115988.

[3] Gong W D, Akbar M F, Jawad G N, et al. The Optimization Method of Magnetic Flux Leakage Detection Equipment in Rail Surface Inspection Based on Finite Element Simulation. Integrated Ferroelectrics, 2024, 240 (8-9): 1205-1219. DOI: 10.1080/10584587.2024.2328859.

[4] Ding Y, Zhao Q, Li T, et al. A rail defect detection framework under class-imbalanced conditions based on improved you only look once network. Engineering Applications of Artificial Intelligence, 2024, 138(PA): 109351. DOI: 10.1016/J.ENGAPPAI.2024.109351.

[5] Anonymous. Enhancing Rail Integrity: Expansion of the Rail Flaw Library at the Transportation Technology Center. Railway Track & Structures, 2024, 120(9): 4-7.

[6] Chang Yongqi, Shen Yi, Zhang Xin, et al. Defect detection of ferromagnetic rail using EMAE-based peak-to-peak method and confidence probability indicator. Measurement Science and Technology, 2024, 35(1). DOI: 10.1088/1361-6501/AD006B.

[7] Santur Y, Yilmazer M, Karakose M, et al. A New Rail Surface Defects Detection Approach Using 3D Laser Cameras Based on ResNet50. Traitement du Signal, 2022, 39(4): 1339-1345. DOI: 10.18280/TS.390427.

[8] Zhao Y, Sun J H, Ma J, et al. Application of the hybrid laser ultrasonic method in rail inspection. INSIGHT, 2014, 56(7): 360-366. DOI: 10.1784/INSI.2014.56.7.360.

Eurasia Journal of Science and Technology
Print ISSN: 2663-1024
Online ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3119

OPTIMAL PLANTING STRATEGY BASED ON GOAL PROGRAMMING

Yang Rong*, ShunYu Li, YuXin Wang
School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, Liaoning, China.
Corresponding Author: Yang Rong, Email: rongyang467@gmail.com

Abstract: In order to maximize the income of agriculture, this paper established linear programming models for the two cases of more than half selling and more than half selling with the constraints of plot area, continuous cropping restrictions, legume crop rotation, planting plots should not be dispersed and crop planting adaptability restrictions. Then it was solved by genetic algorithm, and the optimization results were divided into four categories according to the category of crops: grain, vegetables, edible fungi and beans. The uncertainty risk is measured by calculating the standard deviation of the expected sales volume, yield per mu, planting cost and sales price of crops, and the objective function of minimizing the risk is determined by the weighted sum of the four risk variances. Then, under the condition that more than half price is sold, a double objective optimization model with the objective function of maximizing income and minimizing risk is established, which is solved by NSGA-II algorithm The substitution and complementarity are incorporated into the risk model through the risk weighting mechanism, and the objective function is updated. Then, the Pareto frontier graph is generated by using NSGA-II algorithm. Compared with the results of problem 2, it is found that after considering the substitution and complementarity of agricultural products, the optimal solution set becomes more replaceable, more selective, and more extensive search space.
Keywords: Double objective optimization; Genetic algorithm risk model; NSGA-II algorithm; Pareto frontier

1 INTRODUCTION

The development of rural agriculture has promoted the revitalization of rural economy, the protection of ecological environment and the progress of rural society. It is an important pillar for the realization of Rural Revitalization and sustainable development. Selecting appropriate crops and optimizing planting strategies make field management easier, reduce labor intensity and production costs, and reduce planting risks caused by uncertain factors such as climate and market fluctuations, so as to stabilize farmers' income.

Heydari and Mosadegh developed a fuzzy GP model to optimize crop cultivation patterns in Iran, integrating goals for profit maximization, water conservation, and labor efficiency[1,2]. Their approach reduced water usage by 18% while increasing net profit by 12% compared to traditional practices. Similarly, Arani and Sadeghieh applied GP to dynamic cell formation in apple orchards, achieving a 15.08% yield improvement through optimized planting material allocation and intercropping strategies[3]. These studies underscore GP's versatility in reconciling economic and ecological targets. Jia Chen et al. proposed a linear programming-GP hybrid model for North China's mountainous regions, incorporating variables like planting area, seasonal rotations, and market prices. Their framework reduced fertilizer use by 22% and irrigation water by 19% while maintaining yield levels, demonstrating GP's role in promoting sustainable intensification. Complementing this, Alabdulkader et al. optimized Saudi Arabia's cropping patterns using GP, achieving a 14% increase in gross revenue with 30% less land degradation risk. These findings align with Pastori et al.'s work in Africa, where GP-based strategies enhanced food security by 25% in drought-prone zones.

This paper first needs to judge the possible substitution and complementarity between crops. Based on the analysis of endogenous risk, a bi objective programming model with the goal of maximum return and minimum risk is established. NSGA-II algorithm was used to find the optimal planting scheme, and the sensitivity of the model was tested by introducing exogenous economic risk. The yield and price of each crop can be obtained, and the price cross elasticity of each agricultural product can be calculated. According to the price cross elasticity, we can judge whether crops are alternative or complementary, and form the table of alternative relationship and complementary relationship. The substitution and complementarity conditions of agricultural products are added into the constraint conditions to establish the relationship among planting cost, substitute price and sales volume. The substitution and complementarity will be incorporated into the risk model through the risk weighting mechanism, and the objective function will be updated. For the solution using non dominated sorting genetic algorithm.

2 PRELIMINARY

2.1 Assumption

It is assumed that the expected sales volume, planting cost, yield per mu and sales price of various crops in the future will remain stable relative to 2023. That is to say, the sales data of 2023 will be used as the sales target for each year in the future. The crops planted each season are only sold in the current season, and there is no retention. The recovery

cost and labor, equipment use, transportation and other related expenses are not considered.When beans are planted on different plots for 3-year rotation, in order to simplify the model, only one kind of beans is planted on this plot and no other crops are planted.

2.2 Notations

The symbols used in the paper are listed in Table 1.

Table 1 Symbols Notations

Symbols	Notation
P_j	Sales price of the jth crop
C_j	Planting cost of the jth crop
Y_j	Yield per mu of the jth crop
S_j	Expected sales of crop J
$C_{i,j}$	Cost per kilogram of the j-th crop in the i-th plot
$Y_{i,j}$	Yield per mu of the j-th crop in the i-th plot
A_i	Area of plot I

3 PLANTING COST AND YIELD PROFIT

3.1 Basic Analysis of The Crops

In order to formulate reasonable planting strategies, the village should accurately understand the profit distribution of different crops. By observing the profit differences of different crops, we can better evaluate their economic benefits, market demand and the utilization rate of land resources. By analyzing the expected sales volume, we can identify the crops with high market demand, so as to increase the planting area of these crops, avoid unsalable and waste of resources caused by planting too many crops with low demand, and maximize profits. The profit histogram of 41 crops per mu is shown in Figure 1.

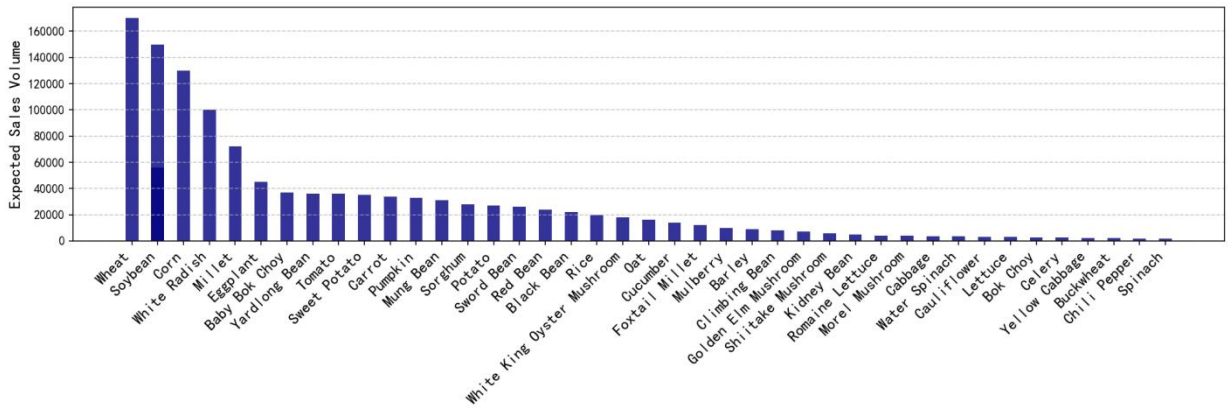


Figure 1 Expected Sales Volume

3.2 Optimal Planting of Crops

The planting area of crops J in each plot or greenhouse shall not exceed the maximum area of the plot. In addition, in order to facilitate farming operations and field management, the area of crops planted in a single plot (including greenhouse) should not be too small. This article defines that the planting area of each crop in a single plot (including greenhouse) shall not be less than 0.2% of the plot area. For irrigated land, ordinary greenhouses and smart greenhouses that can be planted in two seasons, only cowpeas, beans and kidney beans can be planted. According to common sense, cowpea and concanavalis usually prefer a warm climate and are suitable for planting in the first season of spring or summer; Kidney bean has a wide range of adaptability. It is common to plant in the first season in warm areas, but it can also be planted in the second season in high temperature and frost free areas. Because the village is located in the northern mountainous area and the temperature is low throughout the year, only beans are planted in the first season.

The flat dry land, terrace and hillside are suitable for planting one season of grain crops every year, that is, no other crops except one season of grain crops are planted. Irrigated land is suitable for planting one season of rice or two crops of vegetables every year. If two seasons of vegetables are planted on a irrigated land, a variety of vegetables (except Chinese cabbage, white radish and red radish) can be planted in the first season for the convenience of management; In

the second season, only one of Chinese cabbage, white radish and red radish can be planted. In addition, according to seasonal requirements, Chinese cabbage, white radish and carrot can only be planted in the second season of irrigated land.

4 GENETIC ALGORITHM FOR THE MAXIMUM PROFIT

Genetic algorithm (GA) is an optimization algorithm simulating natural selection and genetic mechanism, which is used to solve complex optimization and search problems. The core idea of genetic algorithm is to gradually find the optimal solution by simulating the evolution process in nature. By simulating the process of natural selection, crossover and mutation, genetic algorithm gradually optimizes the solution in the solution space to find the approximate optimal solution of the problem. Its flexible search mechanism and global optimization ability make it have a wide application prospect in dealing with complex optimization problems.

4.1 Visualization and Solutions

In order to intuitively present the optimization characteristics of the "initial seed+local greedy search" algorithm and the cooperative shielding effect of three smoke bombs. After 17.5 iterations, it reaches 6.7677s and is stable, which verifies the effective mining and convergence reliability of the algorithm for the optimization space, and also provides support for parameter optimization. The interval is closely connected without blank, and the total shielding interval is [6.0869,12.8546] s (6.7677s), which not only verifies the synergy effect of "less overlap and no blank" and the rationality of the duration of a single bomb (both<20s), but also provides a theoretical basis for the actual launch.Precise timing reference.

For grain crops, sweet potato with the largest profit per mu is mainly planted after optimization. For vegetables, after optimization, Chinese cabbage, lettuce, lettuce and cucumber are mainly planted. For edible fungi, the yield of Yuhuanggu is the largest, so Yuhuanggu is mainly planted after optimization. After optimization, legumes are mainly planted with red beans, soybeans and black beans. It is not difficult to see that the crops planted after optimization are those with larger profits per mu. Analysis of grains (excluding beans) shows that wheat, millet and corn account for the largest proportion before optimization, while sweet potato accounts for much higher proportion after optimization than other crops.

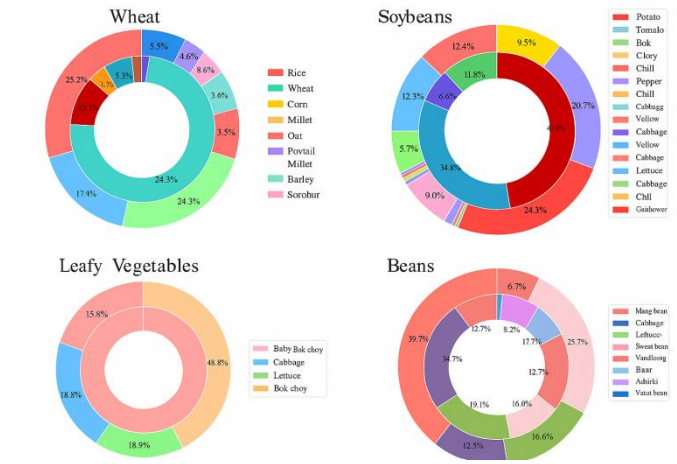


Figure 2 Proportion of Crops before and after Optimization

According to Figure 2, sweet potato has the highest profit per mu, so it is reasonable that the proportion of sweet potato is much higher than that of other crops. According to the analysis of vegetable crops, before optimization, Chinese cabbage and potato accounted for the most, while after optimization, lettuce and cucumber accounted for the most, and the profit per mu in line with both was the largest. Analysis of edible fungi crops can be obtained. After optimization, only Pleurotus nebrodensis and Pleurotus eryngii, the two most profitable edible fungi, are planted. Analysis of grain (excluding beans) shows that before optimization, soybeans and mung beans are mainly planted, and after optimization, black beans are mainly planted, which has the largest profit in grain beans. In conclusion, the results of this paper are accurate and reliable.

4.2 Endogenous Economic Risk

Agricultural production is vulnerable to natural disasters, climate change and market fluctuations. Therefore, it is very important to consider the risk and uncertainty factors in the model. In the existing studies, most farmers' attitudes towards risk are risk averse, so this paper establishes a dual objective programming model with the maximum rate of return and the minimum value of risk to optimize the planting area of crops and the planting structure proportion of different crops.

In actual agricultural production, if the yield of a crop exceeds the market demand, price reduction is a common way to deal with it, and price reduction can reduce the waste and loss caused by slow sales. In the double objective optimization model, in addition to maximizing profits, it is also necessary to minimize risks. The treatment of 50% price reduction for the excess part can balance these two objectives to a certain extent and provide a more robust planting scheme. Therefore, for the second question, this article only considers that the excess part is sold at a 50% price reduction.

4.3 Endogenous Economic Risk

After solving the double objective optimization model, we considered the potential planting risk. Through literature review, we found that there are many risks in the process of crop planting, mainly including natural disaster risk, food safety risk, logistics risk, policy risk and information asymmetry risk [3]. Among them, logistics risk is considered to be one of the most important risks. According to the AHP evaluation results of literature, the weight of logistics risk is as high as 0.393, ranking first among all risk factors, indicating that it has a great impact on the planting and sales of agricultural products.

Logistics risks are mainly manifested in the processing, transportation, storage and distribution of agricultural products, especially vegetable agricultural products, which have high requirements for logistics speed and conditions. If the conditions in the logistics process are not up to standard (such as improper temperature control), the products will rot and deteriorate, thus reducing sales.

Because there are many links involved in the logistics of agricultural products, including harvesting, transportation, storage and distribution, the logistics cost of each link may rise due to risks. Especially in the rural areas of this paper, the infrastructure is not perfect, which increases the difficulty and cost of logistics. These costs are finally transferred to the market price, weakening the market competitiveness of products and affecting the sales volume.

In addition, the logistics risk makes agricultural products unable to stabilize the supply market, which may lead to a situation of short supply or oversupply. When the supply exceeds the demand, the price will drop and the farmers' income will be damaged; If the supply exceeds the demand, the demand of consumers may be restrained due to the high price, resulting in the reduction of sales [4-10].

4.4 Alternative and Complementary Constraints

The substitution relationship means how to adjust the planting area and sales volume of another crop as a substitute product when the price or sales volume of one crop changes. A_{jk} are the price cross elasticity of alternatives. If the two crops J and K are complementary, their sales and prices usually change at the same time. When the price of one crop fluctuates greatly, the sales volume or price of another crop will also be affected. This complementary effect can be expressed by similar correlation coefficients in the calculation of standard deviation. The stronger the complementarity, the greater the fluctuation correlation between the two crops.

The optimal planting plan is shown in Figure 3. The abscissa represents profit and the ordinate represents risk, which are the two objective functions of the model. The purple dot is case 1, which represents the optimal solution set of problem 2. The green point is case 2, which represents the optimal solution set of problem 3. It can be seen from the model that the greater the profit of the model, the smaller the risk. It can be seen from the figure that the optimal solution set of case 2 is larger than that of case1, and more inclined to the lower right. Through comparison, it can be judged that after the optimization of substitutability and complementarity among various crops, the optimal solution set of case 2 has greater substitutability, more selectivity and wider search space.

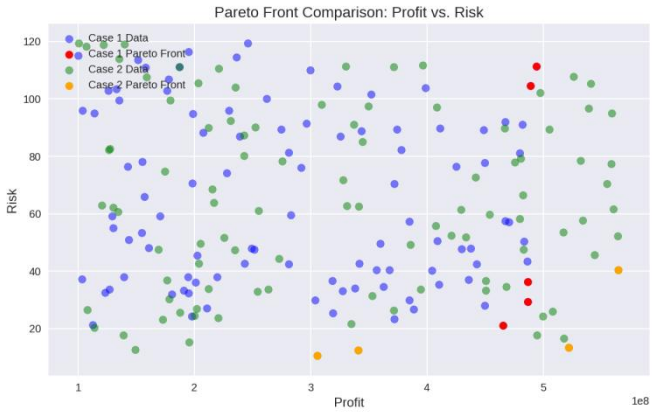


Figure 3 Pareto Front Comparison Chart

5 CONCLUSION

In this study, we focused on optimizing crop - planting strategies for a rural area in the mountainous region, aiming to enhance production efficiency and mitigate planting risks. Through comprehensive data analysis and model

construction, we addressed three key issues and achieved meaningful results. We conducted meticulous data pre - processing and visualization, confirming the absence of significant errors in the provided data. By formulating linear programming models under two scenarios (excess produce being unsellable and sold at half - price), with constraints such as land area, crop rotation restrictions, non - dispersion of planting plots, and crop adaptability, we maximized revenue. this study has successfully developed a comprehensive and practical approach to optimize crop - planting strategies in rural areas. By incorporating multiple factors such as revenue maximization, risk minimization, and crop relationships, we have provided valuable decision - making tools for farmers.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Sow S, Kumar N, Rana L, et al. Suitable Planting Material with Integrated Nutrient Management Can Enhance Productivity, Nutrient Use Efficiency and Profitability of Sugarcane. Journal of Soil Science and Plant Nutrition, 2025, 25(2): 3652-3675.

[2] Jimoh A, Beath H, Markides C N, et al. Enhancing solar mini-grid utilisation in farming communities: crop strategies to reduce costs and improve energy access. IOP Publishing Ltd, 2025.

[3] Dong M, Hong Q I, Zhang Q, et al. The Impact of Sowing Methods on the Seed Germination Environment and Cotton Emergence and Growth. Scientia Agricultura Sinica, 2025, 58(12): 2346-2357.

[4] Kubesch J O C, Simon K J, Kennedy D W, et al. Optimizing late winter plantings of cool-season annual forages in the southern United States. Crop, Forage & Turfgrass Management, 2025, 11(1).

[5] Tom W, Mark S, Andrew S, et al. Catchment woodland planting and the benefits to flood reduction. Forestry: An International Journal of Forest Research, 2025.

[6] Talla A, Swain D K, Srivastava R K, et al. Integrating Climatic Factors and Agronomic Management for Optimal Hybrid Rice Yield and Quality. Communications in Soil Science and Plant Analysis, 2025(5/7): 56.

[7] Nishimura K. Dairy Cattle Feed Design Support Program Linked to Optimal Planting Plan for Self-Supplied Feed. Japan Agricultural Research Quarterly, 2025, 59(1).

[8] Mrope F M, Chilinga F, Nyerere N. Mathematical modeling of powdery mildew disease in cashew plants with optimal control and cost-effectiveness analysis. Modeling Earth Systems and Environment, 2025, 11(4): 1-17.

[9] Shen F, Jiang Y, Yang Y, et al. Long-term impacts of stand density on soil fungal and bacterial communities for targeted cultivation of large-diameter Larix olgensis. Forest Ecology and Management, 2025: 591.

[10] Brown A, Soutor M, Ellison S. Impact of plant density on stem diameter, plant height, and branching in hemp (Cannabis sativa L.). Agrosystems, Geosciences & Environment, 2025, 8(2).

Eurasia Journal of Science and Technology
Print ISSN: 2663-1024
Online ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3120

OPTIMAL CROP PLANTING STRATEGIES BASED ON INTEGER LINEAR PROGRAMMING AND MONTE CARLO MODELS

CongCong Wang[1*], YuLu Ding[2], XuZhe Chen[1], WeiHao Zhou[1], TianLe Cheng[1]
[1]School of Mechanical and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, Shandong, China.
[2]School of Management Engineering, Shandong Jianzhu University, Jinan 250101, Shandong, China.
Corresponding Author: CongCong Wang, Email: w19015318994@163.com

Abstract: This study investigates optimal crop planting strategies for a mountainous village in North China during 2024–2030 to maximize profits. Two optimization models were developed for different market scenarios. First, assuming stable market and production parameters, an integer linear programming model is developed to maximize profits while incorporating core agricultural constraints: field type compatibility, crop rotation to avoid replanting, and legume rotation. This model solves for optimal planting schemes under two scenarios: (1) surplus production leading to unsold waste, and (2) surplus sales at a 50% discount. Second, to address uncertainties and potential risks in projected sales volume, yield per mu, planting costs, and sales prices over the coming years, a Monte Carlo model was introduced for analysis. This method simulates fluctuations in parameters (e.g., annual sales growth rate of 5%–10%, yield variation of ±10% per mu) through random sampling, calculates maximum average returns, and thereby determines the optimal planting area allocation scheme with risk resilience in uncertain environments.
Keywords: Crop planting strategy; Integer linear programming; Monte Carlo simulation; Discount

1 INTRODUCTION

With China's rapid socioeconomic development and rising living standards, agriculture—a vital component of the primary sector—must actively embrace technological innovation to pursue efficient production. Particularly in regions constrained by natural conditions, achieving efficient land utilization, organic cultivation, enhanced agricultural product value, and increased farmer income holds significant practical importance for advancing the rural revitalization strategy. However, planning crop planting schemes must account for crop growth patterns, land suitability, crop rotation requirements, and uncertainties arising from market fluctuations. Traditional experience-based decision-making models struggle to meet these complex demands. To address this challenge, this study focuses on developing a long-term planting plan for the village from 2024 to 2030. It aims to construct a scientific and systematic mathematical optimization model to overcome resource allocation difficulties in both stable and uncertain market environments. This approach will integrate integer linear programming models to determine optimal profit schemes under stable parameters, while incorporating Monte Carlo simulations to address random fluctuations and potential risks in sales volume, yield per mu, costs, and prices. Ultimately, it will provide the village with an optimal planting strategy that balances both economic and social benefits.

The core of this research lies in solving two key optimization problems in crop planting: First, determining the long-term profit-maximizing planting allocation under deterministic constraints. This involves constructing an integer linear programming model under the assumption of stable market parameters while satisfying strict agricultural constraints such as land suitability, crop rotation avoidance, and legume rotation. The goal is to identify the optimal planting area distribution over the next seven years under two risk scenarios: surplus production leading to unsold waste and discounted sales, thereby achieving long-term profit maximization. Second, developing an optimal revenue strategy under market and production uncertainties and potential risks. This involves introducing a Monte Carlo model to effectively simulate random fluctuations of core parameters—such as expected sales volume, yield per mu, and cost-sales price—within defined ranges. A mixed-integer programming model incorporating uncertainty factors is constructed, and extensive simulation experiments calculate the system's statistical characteristics and performance metrics to ultimately determine the optimal planting strategy achieving maximum average revenue.

2 LITERATURE REVIEW

Optimizing crop planting strategies represents a core issue at the intersection of agricultural systems engineering and agricultural economic management. Its objective is to achieve synergistic maximization of economic, ecological, and social benefits through scientific planning within limited arable land resources [1]. Amid intensifying global climate change, increasing market uncertainty, and growing demands for sustainable agricultural development, research in this field is shifting from traditional experiential decision-making toward model-based and data-driven precision decision-making [2]. Scholars worldwide have conducted multidimensional, multi-method explorations on planting strategy optimization, providing a robust theoretical foundation and methodological references for this study.

2.1 Macro Context and Risk Factors in Planting Strategy Optimization

The formulation of planting strategies is not an isolated economic decision but is profoundly influenced by natural environments and macro policies. Li Maoxun elevated the perspective to the water-soil-energy-food (WLEF) coupled system, highlighting that under climate change, the insecurity and misallocation of agricultural resources constitute systemic risks constraining the sustainability of cropping strategies. This underscores the necessity of identifying and regulating multi-factor coupled risks [3]. Zhang Yaoyao focuses on the impacts of multiple coupled meteorological disasters on micro-level actors. Using corn farmers in Shaanxi as a case study, the research reveals how the frequent occurrence and accumulation of extreme weather events constrain farmers' adaptation choices by affecting both yield and risk perception, highlighting the complexity of planting decisions under uncertainty [4]. Tian Pengpeng further empirically analyzed the impact of climate change and farmland water supply on grain output in the Yellow River Basin, demonstrating that infrastructure—as a key means of addressing climate risks—directly influences the stability of planting strategies and output efficiency through its effective provision [5]. Collectively, these studies indicate that a robust planting strategy optimization model must fully account for external uncertainties such as climate change, resource constraints, and disaster risks.

2.2 Evolution of Models and Methods for Crop Strategy Optimization

Mathematical models and intelligent algorithms form the technical backbone of optimization approaches in this field.
Deterministic optimization models represent traditional and core methodologies. Wang Lin and Guo Yaxin applied a linear programming model targeting yield maximization while integrating crop planting patterns and economic constraints, demonstrating its efficacy in supporting scientifically informed planting strategies [6]. Li Yujin et al. introduced dynamic programming and Markov decision models to address multi-stage decision problems, enabling optimization strategies to adapt to the dynamic patterns of crop growth [7]. While these models provide optimal solutions under the assumption of stable parameters, they struggle to handle real-world uncertainties.
To address uncertainty, researchers have adopted risk quantification and stochastic optimization techniques. Shi Jinghao et al. developed a planting strategy optimization model incorporating Conditional Value at Risk (CVaR), solved using the DEGA algorithm. This model integrates risk control objectives with profit maximization, providing a framework for robust decision-making under uncertain market conditions [8]. Following a similar approach, Monte Carlo simulations are widely employed to model random fluctuations in parameters such as sales volume, yield per acre, costs, and prices, thereby evaluating the expected returns and risks of strategies within a stochastic programming framework [1, 8].
Intelligent optimization algorithms demonstrate significant advantages in addressing high-dimensional, nonlinear-constrained crop optimization problems due to their robust global search capabilities. Wang Gangping et al. employed a genetic algorithm to solve a multi-constraint optimization model with total profit as the objective function. Results indicated that cultivating crops with high unit price and high yield per acre constitutes a dominant strategy for profit maximization [9]. Xia Wen et al. employed a particle swarm optimization algorithm to establish an optimization model for scenarios involving multiple terrains and crops. They distinguished between market scenarios characterized by yield surplus waste and discounted sales, providing concrete planting schemes for rural agriculture [1]. Tan Zhiyi and Xiao Junwen, in reviewing mathematical modeling competitions, affirmed the value of mathematical programming and heuristic algorithms in solving complex agricultural operations research problems [2].

2.3 Foundational Applications of Remote Sensing and Information Technology in Planting Strategies

Accurate and timely crop condition information is essential for optimized decision-making. Dong Tianci utilized GF-6 WFV remote sensing data combined with machine learning (random forest, support vector machine) and deep learning (CNN, CNN-LSTM) methods to achieve fine-grained classification of major summer crops in Hengshui City [10]. Such research provides technical support for rapidly acquiring large-scale spatial distribution data of crops, forming a crucial data foundation for future regional-scale precision agricultural management and dynamic adjustment of planting structures.

2.4 Relationship and Innovation of This Study Compared to Existing Literature

In summary, existing research has yielded substantial achievements in model construction, algorithmic solutions, and risk mitigation for optimizing planting strategies. However, most studies either focus on planning under deterministic conditions or concentrate on single risk factors [6-8]. Integrated research that comprehensively considers multiple uncertainties—such as market volatility, yield variations, and price fluctuations—over extended planning periods spanning several years, while systematically comparing different risk mitigation strategies (e.g., surplus product disposal or discounted sales), remains relatively scarce.
Building upon this foundation, this study aims to construct a two-stage integrated optimization framework. First, under deterministic conditions, an integer linear programming model is established to derive the optimal planting plan for the baseline scenario [2, 6]. Subsequently, Monte Carlo simulation is introduced to capture uncertainties in core parameters such as sales, yield, costs, and prices [1, 8], extending the model into a stochastic programming problem to seek robust strategies that maximize expected returns under risk. Through this approach, this study aims to combine the precision of deterministic optimization with the inclusiveness of stochastic optimization toward uncertainty. It seeks to provide a scientifically sound and practically applicable long-term decision support scheme for crop planting in mountainous rural areas of North China, thereby offering a valuable supplement to existing research.

3 MODEL ESTABLISHMENT AND SOLUTION

3.1 Establishment and Solution of the Integer Linear Programming Model

3.1.1 Establishment of the integer linear programming model

Integer linear programming is a mathematical optimization method used to maximize or minimize a linear objective function under a set of linear constraints.

Before establishing the model, it is necessary to define decision variables, parameters, objective functions, and constraints, which are detailed in the following table: $i \in \{1, \ldots, N\}$: represents the i-th plot. $j \in \{1, \ldots, M\}$: represents the j-th crop. $t \in \{1,2\}$: $t = 1$ for the first season and $t = 2$ for the second season. Definition of the Linear Programming Model is shown in table 1.

Table 1 Definition of the Linear Programming Model

Type	Form	Specific Meaning
Decision Variable	$y_{i,j,t}$	1: Crop j is planted on Plot i in Season t; 0: Crop j is not planted on Plot i in Season t
Parameter	p_j	Sales price of the crop
	$q_{i,j,t}$	Yield of Crop j planted on Plot i in Season t
	$D_{i,j,t}$	Expected sales volume of the j-th crop
	c_j	Cost of planting Crop j on Plot i in Season t
	A_j	Area of Plot i

Case 1: The part of the total output exceeding the corresponding expected sales volume is not included in the revenue. The mathematical expression for the total objective, i.e., the maximum profit to be obtained, is as follows:

The constraints are as follows:

Constraint 1: Flat dry land, terraced fields, and sloping land are suitable for single-crop planting of food crops (excluding rice) every year, leading to the following constraint: No replanting in the second season:

$$\sum_{j=1}^{15} y_{i,j,1} = 1, \forall i \in \{A^{1-6}, B^{1-14}, C^{1-6}\} \tag{1}$$

Constraint 2: Paddy fields can be planted with rice in a single season every year, leading to the following constraint:

$$\sum_{j=16} y_{i,j,1} = 1, \forall i \in \{D^{1-6}\} \tag{2}$$

$$\sum_{j=16} y_{i,j,2} = 0, \forall i \in \{D^{1-6}\} \tag{3}$$

Constraint 3: Paddy fields can also be planted with double-crop vegetables. A variety of vegetables (excluding white radish, carrot, and Chinese cabbage) can be planted in the first season; only one of Chinese cabbage, white radish, and carrot can be planted in the second season (for ease of management), leading to the following constraints:

$$\sum_{j=17}^{34} y_{i,j,1} = 1, \forall i \in \{D^{1-8}\} \tag{4}$$

$$\sum_{j=17}^{34} y_{i,j,2} = 1, \forall i \in \{F^{1-4}\} \tag{5}$$

$$\sum_{j=35}^{37} y_{i,j,2} \leq 1, \forall i \in \{D^{1-8}\} \tag{6}$$

Constraint 5: Ordinary greenhouses can be planted with two crops every year. A variety of vegetables (excluding Chinese cabbage, carrot, and white radish) can be planted in the first season, and only edible fungi can be planted in the second season, leading to the following constraint:

$$\sum_{j=17}^{34} y_{i,j,1} = 1, \forall i \in \{E^{1-16}\} \tag{7}$$

Constraint 6: Smart greenhouses can be planted with two crops of vegetables (excluding Chinese cabbage, white radish, and carrot) every year, leading to the following constraint:

Constraint 7: No continuous replanting of the same crop on the same plot (including greenhouses), leading to the following constraint:

Constraint 8: All land of each plot (including greenhouses) must be planted with leguminous crops at least once every three years, leading to the following constraint:

Integrating the above mathematical expressions, the problem is constructed into the following integer linear programming model:

$$\max V_1 = \sum_i \sum_j \sum_t \left(p_j \times \min(Q_{j,t}, D_{j,t}) \times y_{i,j,t} - c_j \times A_j \times y_{i,j,t}\right) \tag{8}$$

Case 2: Compared with Case 1, the unsold part in Case 2 is not treated as waste but sold at a 50% discount of the 2023 sales price. Therefore, it is only necessary to add this scenario to the total objective function. The specific model is established as follows:

$$\max V_2 = \sum_{i=1}^{n} \sum_{j=1}^{M} \sum_{t=1}^{2} \left\{p_j \times \min(Q_{j,t}, D_{j,t}) + 0.5p_j \times (Q_{j,t} - D_{j,t}) - c_j \times x_{i,j,0}\right\} \tag{9}$$

The constraints for this case are consistent with those in Case 1. Finally, all expressions are summarized as follows:

$$\max V_2 = \sum_{i=1}^{n} \sum_{j=1}^{M} \sum_{t=1}^{2} \left\{p_j \times \min(Q_{j,t}, D_{j,t}) \times y_{i,j,t} + 0.5p_j \times (Q_{j,t} - D_{j,t}) - c_j \times A_i \times y_{i,j,t}\right\} \tag{10}$$

3.1.2 Solution of the integer linear programming model

Step 1: Problem Understanding and Variable Definition

First, clarify the problem requirements. It can be known from the problem that the optimal crop planting plan for the village from 2024 to 2030 needs to be provided. Determine the decision variables and other relevant parameters in the problem, such as the planting area of crops in different plots in different seasons, and assign symbolic representations to each variable.

Step 2: Establish the Objective Function

According to the objective of the problem, construct the objective function. Since it is required to provide the optimal crop planting plan for the village from 2024 to 2030, the problem can be converted into finding the maximum benefit under the optimal planting plan. The objective function can be expressed as an expression of total income minus total cost, involving a linear combination of decision variables.

Step 3: Determine Constraints

Analyze various constraint factors in the problem and convert them into linear inequalities.

Step 4: Organize the Model

Integrate the objective function and constraints to form a complete integer linear programming model. Ensure all variables are non-negative and conform to the standard form of integer linear programming.

Step 5: Select the Solution Method

Use Python's PuLP library, etc., and configure corresponding parameters for software solution.

Step 6: Solve the Problem

Run the Python code to obtain the optimal solution to the integer linear programming problem. The solver will return the optimal decision variable values and the corresponding optimal objective function value.

3.1.3 Analysis of the results

The planting strategy obtained according to the planning model fully considers the constraints given in the problem. Partial Crop Planting Strategy for Q1 2024) shows that in plots such as flat dryland (A1-A6), terraced fields (B1-B14), and hillside land (C1-C5), the planting areas for soybeans, black beans, red beans, mung beans, and winged beans are all 0. Wheat, corn, sorghum, pumpkin, and sweet potato have identical planting areas across all plots (e.g., 16 mu in plot A1 and 11 mu in plot A2, with gradient variations based on plot size). The planting areas for millet, foxtail millet, and buckwheat are also zero. This indicates that during the first quarter of 2024, under the scenario of "excess production leading to unsold waste," the village prioritized uniformly planting wheat, corn, sorghum, pumpkins, and sweet potatoes across different plot types, while refraining from planting legumes, millet, sorghum, and buckwheat. The allocation of planting areas within the same plot type remained consistent, adhering to constraints such as plot type suitability (e.g., flat dryland plots only for grain crops) and crop rotation practices. The planting area of wheat has decreased in the past few years, and food crops are planted in rotation. Partial Crop Planting Strategy for Q1 2025) shows significant adjustments in crop types: Black beans became the primary crop, occupying exactly the same acreage as wheat and corn across all plots (e.g., 16 mu in Plot A1, 11 mu in Plot A2); The planting areas for wheat, corn, sorghum, pumpkin, and sweet potato become zero. The planting areas for millet and foxtail millet are the same as black beans, while soybeans, red beans, mung beans, winged beans, and buckwheat remain unplanted. This demonstrates that in the first quarter of 2025, under identical constraints and market stagnation conditions, the village adjusted its planting scheme to prioritize black beans, millet, and foxtail millet. It maintained the planting areas for each plot consistent with the core crop areas of the corresponding plots from the previous year while satisfying the legume rotation requirement (planting legumes at least once within a three-year cycle.

On flat dryland (A1-A6), terraced fields (B1-B14), and hillside land (C1-C5), the planting areas for soybeans, black beans, red beans, mung beans, winged beans, millet, foxtail millet, and buckwheat are all 0; Wheat, corn, sorghum, pumpkin, and sweet potato are planted across all plots with identical acreage, distributed in a gradient based on plot size (e.g., 16 mu in Plot A1, 11 mu in Plot A2, 3 mu in Plot C1). This indicates that under the "discounted sales for unsold surplus" revenue model, crops like wheat and corn remained the priority choices in Q1 2024. Core crop categories were not adjusted; excess production was absorbed solely through price mechanisms. The allocation of planting areas complied with plot suitability constraints (e.g., grain crops only planted on flat dryland) and crop rotation restrictions.

Black beans, millet, and sorghum occupy identical acreage across all plots (e.g., 16 mu in Plot A1, 12 mu in Plot B1, 3 mu in Plot C1). Soybeans, red beans, mung beans, cowpeas, wheat, corn, sorghum, pumpkin, sweet potatoes, and

buckwheat are not cultivated. This indicates that in Q1 2025, even when adopting the "discounting unsold portions" model, the planting plan remains centered on black beans (satisfying legume rotation constraints), millet, and foxtail millet. The core crops and planting areas per plot perfectly match the "unsold waste" scenario. Profit optimization occurs solely through revenue calculation logic (including discounting), without altering crop types or area allocation strategies.

4 MODEL ESTABLISHMENT AND SOLUTION

4.1 Establishment and Solution of the Model

The Monte Carlo model is a method for simulating the behavior of complex systems through random sampling. Based on probability and statistics theory, it uses random number generators to simulate uncertain factors in the system, thereby conducting multiple simulation experiments on the system to obtain the statistical characteristics and performance indicators of the system.

4.2 Introduction of Random Disturbance Variables

The traditional integer linear programming model assumes that all parameters (such as planting costs and expected sales volumes) are deterministic and remain unchanged throughout the planning period. However, in this problem, many parameters are uncertain. For example, the future expected sales volumes of wheat and corn are expected to grow at an annual growth rate of 5% to 10%, while the sales volumes of other crops may fluctuate by approximately ±5% every year compared to 2023. The yield per mu of crops is affected by climate and other factors and may change by ±10% every year. The planting cost is expected to increase by about 5% every year. The prices of food crops are basically stable, the prices of vegetable crops are expected to increase by about 5% every year, and the prices of edible fungi may decrease by 1% to 5% every year, especially the price of morel mushrooms, which is expected to decrease by 5% every year. Therefore, instead of using the traditional integer linear programming for solving this problem, we use the Monte Carlo model to simulate various planting scenarios through generating a large number of random samples, calculate the maximum average profit, and determine the optimal planting area allocation plan to cope with uncertainty.
The following table 2 defines the decision variables, objective function, and constraints:

Table 2 Definition of the Monte Carlo Model

Type	Form	Specific Meaning
Decision Variable	$y_{i,j,t,T}$	Whether to plant on the i-th plot in the t-th quarter of the T-th year; 1 indicates planting, 0 indicates not planting
Parameter	$p_{j,T}$	Sales price of the j-th crop in the T-th year
	$q_{j,T}$	Yield of the j-th crop in the T-th year
	$D_{j,T}$	Expected sales volume of the j-th crop in the T-th year
	$c_{j,T}$	Planting cost of the j-th crop in the T-th year
	A_j	Area of Plot j

Where $T \in \{2024, \ldots, 2030\}$
The following are the mathematical expressions for various uncertain factors:
The expected sales volumes of wheat and corn will have an annual growth rate of 5%-10%:

$$r_{i,j,T} \in [0.05, 0.1] \tag{11}$$

Where $T \in \{2023, \ldots, 2030\}$

$$D_{i,j,T+1} = D_{i,j,T} \times (1 + r_{i,j,T}), \forall i \in \{z_6, z_7\} \tag{12}$$

The sales volume of other crops changes by ±5%:

$$\ell_{i,j,T} \in [-0.05, 0.05] \tag{13}$$

$$D_{i,j,T+1} = D_{i,j,T} \times (1 + \ell_{i,j,T}), \forall i \in \{z_{1-5}, z_{8-41}\} \tag{14}$$

The yield per mu of crops changes by ±10% every year:
The prices of food crops are basically stable:

$$p_{i,j,T+1} = p_{i,j,T}, \forall j \in \{1 \sim 16\} \tag{15}$$

The prices of vegetable crops increase by 5% every year:

$$p_{i,j,T+1} = p_{i,j,T} \times 1.05, \forall i \in \{z_{17-37}\} \tag{16}$$

The price of morel mushrooms decreases by 1%~5%:

$$\gamma_{i,j,T} \in [0.01, 0.05] \tag{17}$$

The planting cost increases by an average of 5% every year:

$$c_{i,j,T} = c_{i,j,T} \times 1.05 \tag{18}$$

Potential risks (extreme drought or severe diseases and pests):

$$\eta_{j,T} \in [-0.1, 0.1] \tag{19}$$

$$\mu_T \in [0.75, 0.85] \tag{20}$$

Integrating the above uncertain factors and constraints, a mixed-integer programming model with uncertain factors is constructed, and the Monte Carlo method is used to simulate the optimal planting plan:

4.2.1 Solution of the Monte Carlo model

Step 1: Define Decision Variables

As in the previous model construction, determine the decision variables representing the planting plan, such as the planting area of various crops.

Step 2: Determine the Objective Function

To provide the optimal crop planting plan for the village from 2024 to 2030, convert this problem into finding the maximum benefit.

Step 3: Establish Constraints

The constraints for this problem are the same as 3.1.

Step 4: Conduct Monte Carlo Simulation

Determine the number of simulations: According to the complexity of the problem and computing resources, the number of simulations is determined to be 100 times.

Random Sampling: For each uncertain factor, perform random sampling according to its probability distribution. For example, for changes in crop yield per mu, randomly select a value from the interval [-0.1, 0.1]; for price changes, sample according to the corresponding distribution.

Calculate the Objective Function Value: For each set of uncertain factor values obtained by sampling, substitute them into the objective function and constraints to calculate the corresponding objective function value and decision variable values. This step is equivalent to simulating a possible actual situation.

Repeat Simulation: Repeat Steps 2 and 3 to conduct multiple simulations, obtain a large number of combinations of objective function values and decision variables, and finally obtain the planting plan with the maximum benefit.

4.2.2 Analysis of the results

Regarding the constraint in the problem that leguminous crops must be planted at least once every three years, we consider optimizing only the content of the most recent three years. As long as each plot is planted with leguminous crops every three years, the soil can be kept fertile, and the final objective model is measured with the goal of maximizing profit. Therefore, for the next n years, we only need to add this random disturbance and fill in the optimization results in turn.

The core crop mix shifts from “wheat + corn + sorghum + pumpkin + sweet potato” to “corn + sorghum + pumpkin + sweet potato,” with wheat acreage reduced to zero across all plots. The planting areas for corn, sorghum, pumpkin, and sweet potato in each plot increased compared to the corresponding plots in Question 1 (e.g., Plot A1 increased from 16 mu to 20 mu, Plot A2 increased from 11 mu to 13.75 mu, Plot C1 increased from 3 mu to 3.75 mu); Legumes, millet, sorghum, and buckwheat remain unplanted. This reflects the strategy of prioritizing increased planting areas for crops like corn in Q1 2024 while eliminating wheat to balance risk and reward, accounting for parameter fluctuations (e.g., uncertainty in wheat sales growth). This aligns with the Monte Carlo model's logic for addressing uncertainty. the crop mix shifts from “Black Soybeans + Millet + Sorghum” to “Black Soybeans + Wheat + Millet + Sorghum.” Wheat planting is newly introduced with an area equal to that of black soybeans, millet, and sorghum; (e.g., Plot A1: 20 mu, Plot A2: 13.75 mu, Plot C1: 3.75 mu); Soybeans, red beans, mung beans, winged beans, corn, sorghum, pumpkin, sweet potatoes, and buckwheat remain unplanted. This indicates that under uncertainty constraints, optimizing the planting structure in Q1 2025 by adding wheat and retaining black beans (to satisfy legume rotation) while maintaining core crop acreages consistent with the previous year's solution for each plot simultaneously addresses parameter fluctuations and complies with plot suitability and rotation constraints.

5 CONCLUSIONS

This study successfully constructed a two-stage optimization model for crop cultivation in the target village. Under the ideal scenario of stable parameters, the integer linear programming model provides optimal planting schemes satisfying all agricultural constraints (including crop rotation and plot limitations), explicitly indicating how to allocate resources to maximize profits when facing yield surpluses, unsold inventory, or discounted sales. For the more challenging scenario of uncertain market conditions, this study incorporates Monte Carlo simulation. By introducing random perturbations and simulations of expected sales volume, yield per mu, and price fluctuations, potential risks are effectively integrated into the decision-making process. The resulting optimal planting strategy is a robust solution measured by maximum average profit. The successful solution of this model provides scientific and reliable decision-making support for the village to fully utilize its limited arable land resources, develop organic farming tailored to local conditions, and achieve sustainable long-term economic growth.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Xia Wen, Shi Yufei, Zou Shuai, et al. Optimization of Multi-terrain Multi-crop Planting Strategies Based on Particle Swarm Optimization. Journal of Smart Agriculture, 2025, 5(17): 49-53.

[2] Tan Zhiyi, Xiao Junwen. Review of the “Crop Planting Strategy” Competition Problem. Mathematical Modeling and Its Applications, 2025, 14(02): 28-36.
[3] Li Maoxun. Identification of Risk Characteristics and Regulation Mechanisms in Heilongjiang Province's Water-Soil-Energy-Food Coupled System under Climate Change. Northeast Agricultural University, 2025.
[4] Zhang Yaoyao. Selection of Adaptation Measures by Corn Farmers under Multiple Coupled Meteorological Disasters: A Case Study in Shaanxi. Northwest A&F University, 2025.
[5] Tian Pengpeng. Study on the Impact of Climate Change and Farmland Water Supply on Grain Output: A Case Study of the Yellow River Basin. Northwest A&F University, 2025.
[6] Niraj K, Vikas R, Mukesh K, et al. Assessing soil water dynamics and wheat growth under different irrigation and planting strategies using HYDRUS-2D. Hydrology Research, 2025, 56(9): 878-901. DOI: 10.2166/NH.2025.026.
[7] Adebayo O, Thapa R V, Ulery A, et al. Optimizing inorganic nitrogen extraction method to evaluate alternative cropping strategies. Agronomy Journal, 2025, 117(4): e70120. DOI: 10.1002/AGJ2.70120.
[8] Neelipally R K T R, Chhetri A, Saha D, et al. Agronomic responses of transitioning organic grain rotations employing multi‐tactic tillage and cover cropping strategies. Agronomy Journal, 2025, 117(3): e70095. DOI: 10.1002/AGJ2.70095.
[9] Riely C C, MacMillan W R, Janowiak K M, et al. Field Note: Learning from Early Application of a Transition Forest Climate Adaptation Planting Strategy Incorporating Assisted Migration in Southern New England. Journal of Forestry, 2025, 123(3): 1-16. DOI: 10.1007/S44392-025-00019-Y.
[10] Untung N, Cahyono H, Marbun P, et al. Church planting strategies in the context of religious moderation in multicultural societies. HTS Teologiese Studies/Theological Studies, 2025, 81(1): e1-e7. DOI: 10.4102/HTS.V81I1.10498.

Eurasia Journal of Science and Technology
Print ISSN: 2663-1024
Online ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3121

NEURAL-NETWORK-ASSISTED MODEL PREDICTIVE CONTROL FOR ACTUATORS IN NUCLEAR POWER PLANT DRIVE SYSTEMS

Yuan Zhang*, YanKun Li, Chao Si
Electrical Department, China Institute of Atomic Energy, Beijing 102400, China.
**Corresponding Author: Yuan Zhang*

Abstract: This study develops a unified dynamic modeling framework for nuclear-grade actuators, such as fans, pumps, and valves, critical for reactor cooling, feedwater regulation, and ventilation. These actuators exhibit nonlinearities, cross-coupling, and operational constraints, challenging conventional controllers to maintain performance and safety under fluctuating conditions. To address this, a neural-network-assisted model predictive control (NN-MPC) architecture is proposed. The framework incorporates a mask-based input encoding strategy to represent all actuators within a single model, and a fully differentiable neural network is trained to approximate actuator dynamics. A bias-compensation mechanism is introduced to address multi-step prediction drift caused by process variations and aging. The resulting nonlinear MPC integrates actuator constraints, output limitations, rate restrictions, and real-time optimization, ensuring millisecond-level performance required in nuclear applications. The implementation workflow includes real-time execution analysis, anomaly detection, safety fallback logic, and hardware-in-the-loop validation. Experimental results demonstrate that the NN-MPC framework provides accurate modeling, high-performance control, and compliance with nuclear safety requirements. This research offers a robust pathway for advancing intelligent, data-driven control systems in next-generation nuclear power plants.
Keywords: Nuclear power plant actuators; Neural network prediction model; Model predictive control; Bias compensation; Real-time optimization

1 INTRODUCTION

Nuclear power plants are highly coupled, safety-critical energy systems whose operational performance depends strongly on the reliable and stable behavior of drive actuators such as fans, pumps, and valves. These electrically driven devices support essential functions including reactor coolant circulation, feedwater control for steam generators, plant ventilation and heat removal, chemical volume regulation, and emergency mitigation under accident scenarios. They constitute a vital part of the nuclear unit's capability to maintain steady-state operation and preserve safety margins [1]. The actuators are not only numerous but also distributed across various physical subsystems, forming complex dynamic chains that span multiple regions and loops. Their control performance directly influences thermohydraulic stability, operational boundaries of key equipment, and the coordination efficiency among reactor-side and turbine-side systems. Consequently, enhancing the dynamic performance, robustness, and predictability of these actuators has long been a core research focus in nuclear engineering control [2].

Compared with conventional industrial facilities, actuator control in nuclear power plants exhibits several distinctive characteristics. First, control strategies must conform to strict nuclear safety regulations governing allowable input ranges, actuation speeds, overshoot limits, steady-state accuracy, and behavior under abnormal or accident conditions [3]. Second, due to the structural complexity of nuclear systems, actuators are strongly coupled through hydraulic flow, differential pressure, thermal effects, and environmental interactions, resulting in nonlinear, time-varying, and tightly coupled dynamics [4]. Third, although the plant typically operates under relatively stable conditions, it is extremely sensitive to disturbances—such as load fluctuations, cooling water temperature variations, and system resistance changes—which may significantly affect actuator behavior; stability and robustness therefore remain primary design objectives. Furthermore, nuclear facilities cannot perform frequent online experiments or deliberate perturbations on actuators as in ordinary industrial systems, implying that control methods must achieve high reliability under limited data availability and restricted testing conditions [5]. Overall, the modeling, control design, and safety assurance of nuclear actuators involve challenges far exceeding those of typical industrial scenarios.

To date, proportional–integral–derivative (PID) control remains the dominant approach for nuclear actuator control due to its simplicity, ease of implementation, and long-standing operational experience in safety-critical environments. However, PID is inherently linear and therefore insufficient for compensating complex nonlinear dynamics. Moreover, PID cannot explicitly handle physical constraints such as stroke limits, maximum rotational speed, pressure-drop boundaries, or ramp-rate limits. Under strong disturbances, dynamic variations, or intensified coupling, PID controllers often exhibit overshoot, sluggish responses, or oscillatory behavior [6]. Additionally, the large number of actuators in nuclear power plants—with diverse operating characteristics—makes empirical PID tuning labor-intensive and inconsistent, compromising verifiability and maintainability. As modern nuclear power plants evolve toward digitalized and intelligent control systems, PID increasingly struggles to meet new demands related to fine regulation, energy

efficiency, dynamic performance enhancement, and auditability.

Model predictive control (MPC), an optimization-based control framework that employs explicit handling of system constraints and multivariable coupling, is widely regarded as a promising pathway for upgrading actuator control in nuclear power plants [7]. MPC offers excellent predictability and auditability: control actions are derived from real-time optimization with a transparent mathematical structure suitable for rigorous verification, aligning well with nuclear safety requirements. However, MPC heavily relies on accurate prediction models. The nonlinear dynamics of fans, pumps, valves, and their hydraulic networks are difficult to represent using traditional mechanistic models; conversely, high-fidelity physical models impose excessive computational burdens incompatible with the sub-100-ms sampling periods commonly required in nuclear applications [8]. Therefore, obtaining models that are both accurate and computationally efficient remains a key barrier to large-scale MPC deployment for nuclear actuators.

Recent advances in neural networks (NNs) have demonstrated strong potential in nonlinear system modeling, fluid-system prediction, and actuator-level data-driven modeling [9]. Unlike mechanistic modeling, neural networks do not require complex thermohydraulic equations and can learn input–output relationships directly from operational data. This makes them particularly suitable for characterizing fan pressure–flow behavior, pump head–flow curves, and valve flow characteristics as a function of opening and differential pressure [10]. In nuclear power plants, much of the available data originates from normal operations, cold-state testing, and mode transitions—typically stable operating regimes—allowing neural networks to achieve high-accuracy approximations of actuator dynamics. Additionally, their flexible architecture enables unified input encoding, facilitating shared modeling across different actuator types and thereby supporting a unified control framework [11].

Nonetheless, neural networks alone are not sufficient for nuclear-grade control. Their lack of interpretability, uncertain generalization outside the training distribution, sensitivity to bias accumulation, and dependence on data make them unsuitable for direct use in closed-loop safety-critical control [12]. In particular, multi-step NN prediction errors may accumulate rapidly, causing predicted trajectories to deviate significantly from physical behavior—an unacceptable risk in nuclear applications without additional safeguards. Therefore, using NNs as stand-alone controllers does not satisfy nuclear safety, auditability, or performance predictability requirements.

Against this backdrop, combining neural network–based nonlinear modeling with the constraint optimization capabilities of MPC has emerged as an ideal solution—namely neural-network-assisted MPC (NN-MPC). In this framework, NNs provide accurate nonlinear prediction models, while MPC computes control actions under explicit constraints, thereby achieving high prediction fidelity and robust enforcement of nuclear safety requirements [13]. Importantly, NNs serve solely as predictive components, while MPC governs the final control decisions, preserving interpretability and compliance with nuclear standards. Through structured input encoding and a mask-based mechanism, the unified NN model accommodates fans, pumps, and valves within a single predictive architecture, significantly enhancing engineering maintainability [14].

Based on these considerations, this study proposes a unified NN-MPC control framework tailored for nuclear power plant actuators. The framework integrates unified dynamic modeling, NN-based prediction design, multi-step error compensation, MPC constraint formulation, and nuclear-grade safety mechanisms. The objective is to achieve high-accuracy, predictable, and robust control for diverse actuators while satisfying the stringent verifiability, auditability, and reliability requirements of the nuclear industry. Ultimately, the proposed methodology provides a practical and standardized technical route for the digitalization and intelligent modernization of nuclear power plant control systems.

2 UNIFIED MODELING FRAMEWORK FOR ACTUATORS IN NUCLEAR POWER PLANT DRIVE SYSTEMS

Actuators in nuclear power plants—including fans, pumps, and valves—exhibit differences in their operational objectives and working conditions. Nevertheless, from the perspective of control architecture and dynamic characteristics, they can all be abstracted into a generalized “motor–actuator–fluid loop” dynamic system. Fans regulate air flow and heat exchange by driving impellers; pumps deliver hydraulic power and boost pressure through impeller rotation; valves modulate flow distribution and hydraulic resistance by adjusting their opening via mechanical transmission. The dynamic behavior of these devices is jointly influenced by input commands, internal mechanical characteristics, and coupling with external fluid systems. As a result, they typically exhibit nonlinear, time-varying, strongly coupled dynamics with multiple operational constraints [3-4].

When designing advanced control strategies for nuclear power plants, establishing a unified dynamic modeling framework for these three types of actuators brings significant advantages. Such a unified representation not only reduces modeling complexity but also greatly enhances engineering consistency in deployment, auditing, verification, and maintenance of control algorithms [5].

From a control-system perspective, although fans, pumps, and valves produce different physical outputs, their actuation mechanisms share common features: each is driven by a controllable motor input (e.g., motor frequency, motor current, valve opening command) and generates measurable dynamic responses such as pressure, flow rate, temperature, or environmental state variables. This commonality makes it feasible to describe all three types of actuators using a unified input–state–output structure, thereby providing a consistent interface for neural network (NN) prediction models and model predictive control (MPC). In safety-critical nuclear applications, unified modeling is not merely a methodological convenience, but an engineering necessity. Nuclear-grade control systems demand strict auditability, verifiability, consistency, and traceability. By adopting a unified modeling framework, the number of models in the

control system lifecycle can be greatly reduced, simplifying parameter review and version management and improving the overall engineering reliability of the control architecture [6].

2.1 Design Principles of the Unified Modeling Framework

The first step in establishing a unified model is to identify the common structural and control characteristics of fans, pumps, and valves. Although their physical mechanisms differ—for instance, rotational inertia dominates in fans and pumps, whereas valve dynamics are governed by mechanical transmission and the nonlinear relationship between opening and hydraulic resistance—all these systems can be abstracted as discrete-time nonlinear dynamics:

$$x(k+1)=f\big(x(k),u(k),d(k)\big),\ \ y(k)=h(x(k)) \tag{1}$$

where x(k) denotes internal system states such as motor speed, valve stem position, or other actuator-specific internal variables; u(k) represents control inputs (frequency command, motor current, valve opening); d(k) captures external disturbances and fluid-loop coupling, including hydraulic resistance, system load, or fluid temperature; and y(k) denotes system outputs such as flow rate, pressure difference, temperature, or environment-related variables.
The actuator must operate within strict physical and safety constraints:

$$u_{min}\ \le u(k)\le u_{max},\ \ y_{min}\ \le y(k)\le y_{max},\ \ |\Delta u(k)|\le \Delta u_{max} \tag{2}$$

In nuclear-grade systems, each constraint has explicit safety implications. Violations may trigger protective actions or cause equipment degradation. Thus, the unified modeling framework must embed these safety constraints in the model structure to enable their explicit treatment within MPC.
To achieve unified modeling, this study introduces a standardized input encoding combined with a mask mechanism. The mask ensures that each actuator uses only the input channels relevant to its physical characteristics, while irrelevant channels are suppressed. This enables the unified model to maintain a consistent structure while preserving the flexibility needed to represent different devices. As a result, fans, pumps, and valves share model parameters and the same learning architecture, significantly reducing deployment complexity in nuclear applications.

2.2 Construction of the Unified Input – State – Output Structure

To ensure that the unified model captures both internal and external factors affecting actuator dynamics, the input vector is designed to include actuator states, command inputs, fluid variables, thermal parameters, and context information. Specifically, the input vector is defined as

$$s(k)=[x(k),u(k),\Delta p(k),q(k),T(k),z(k)] \tag{3}$$

where:
x(k) denotes internal actuator states (motor speed, impeller inertia, valve stem displacement, etc.);
u(k)is the control input;
Δp(k) and q(k) represent pressure difference and flow rate, the two fundamental fluid variables in nuclear hydraulic systems;
T(k) denotes temperature or thermohydraulic indicators;
z(k) includes contextual variables such as operation mode, actuator type code, or pipeline resistance parameters.
All actuators share the same structure of $s(k)$, but irrelevant channels are masked via:

$$\tilde{s}(k)=m\odot s(k) \tag{4}$$

where m is a device-specific 0–1 mask vector.
The mask mechanism achieves both generalization and structural consistency, ensuring that the unified neural model remains easy to train while being capable of representing significantly different actuator behaviors. Moreover, the unified structure enables effective cross-device feature sharing. Many dynamic properties—such as fluid inertia, mechanical inertia, or delayed response—share similar mathematical patterns across actuators. A unified input encoding thus improves model generalization and reduces the required amount of training data [8].

2.3 Normalization and Data Availability in Nuclear Applications

Because variables exhibit different scales—and because flow, pressure, and temperature distributions vary widely across operating conditions—normalization is essential to ensure numerical stability. In nuclear applications, fixed-range linear normalization or rated-condition-based normalization is preferred over adaptive normalization schemes that change with training data. Fixed normalization ensures auditability, traceability, and reproducibility, all of which are required in nuclear safety reviews [9].
Data availability is another practical constraint. Nuclear power plants typically provide only limited datasets, mainly including normal operation data, load-varying conditions, cold-state commissioning data, and historical fault logs. The unified modeling framework mitigates this issue by enabling multi-device training, thereby expanding the effective dataset and improving model coverage and generalization capability.

2.4 Nonlinear Dynamics of Nuclear-Grade Actuators

Due to their fluid-mechanical interactions, fans, pumps, and valves all exhibit strongly nonlinear dynamic properties. Examples include:
Pressure–flow characteristics of fans vary significantly with pipeline resistance.
Pump head curves show steep variations in low-flow regions.
Valve flow coefficient curves follow typical S-shaped profiles, and stem friction introduces hysteresis.
Such nonlinearities are difficult to capture using linear models, while high-fidelity mechanistic models are computationally intensive and unsuitable for real-time MPC applications.
To address this challenge, this study adopts neural networks as the core predictive component of the unified dynamic model. Neural networks provide excellent nonlinear approximation capability and can effectively represent actuator dynamics using limited industrial data [10]. Moreover, the unified input structure allows the neural network to handle heterogeneous devices within a single model.
Given the interpretability and verifiability requirements of nuclear applications, the study employs shallow architectures—such as small multilayer perceptrons (MLPs) or lightweight recurrent networks—to avoid the opacity and unpredictability of deep or highly complex neural structures [11].

2.5 Schematic Diagram of the Unified Modeling Framework

The logical structure of the unified modeling framework is illustrated in Figure 1:

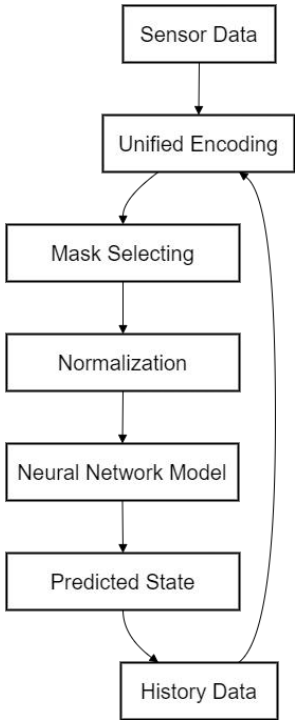


Figure 1 Unified Modeling Data Flow

Shows the information flow from actuators → unified input encoding → neural network predictor → MPC controller → actuator commands.
This diagram visualizes the complete data flow underlying the NN-MPC framework introduced in later chapters.

2.6 Requirements for Differentiable Predictive Models in MPC

Because the neural network serves as the predictive model within MPC, it must satisfy two critical requirements: differentiability and numerical stability. Differentiability is necessary for constructing Jacobians and gradients used in linearization and real-time optimization. Numerical stability ensures that predicted trajectories remain physically meaningful and do not generate nonphysical values such as negative flow or pressure [12].
To satisfy these requirements, the model employs smooth activation functions and avoids non-differentiable or excessively aggressive nonlinearities. A single-step prediction structure is adopted to reduce multi-step error accumulation, and multi-step robustness is further reinforced through the bias compensation mechanism introduced in subsequent chapters.

2.7 Engineering Value of the Unified Modeling Framework

The unified modeling framework provides substantial engineering benefits in nuclear applications. First, it enables a single predictive model to represent fans, pumps, and valves, thereby increasing consistency during system upgrades and reducing the burden of model auditing and nuclear-safety certification. Second, the unified model interfaces seamlessly with the NN-MPC controller, enabling consistent predictive control across heterogeneous actuators and improving plant-wide operational efficiency. Finally, the unified modeling approach enhances the manageability, controllability, and auditability of neural-network components in nuclear control systems, providing a clear technical pathway for digitalized and intelligent next-generation nuclear power plants (Figure 2).

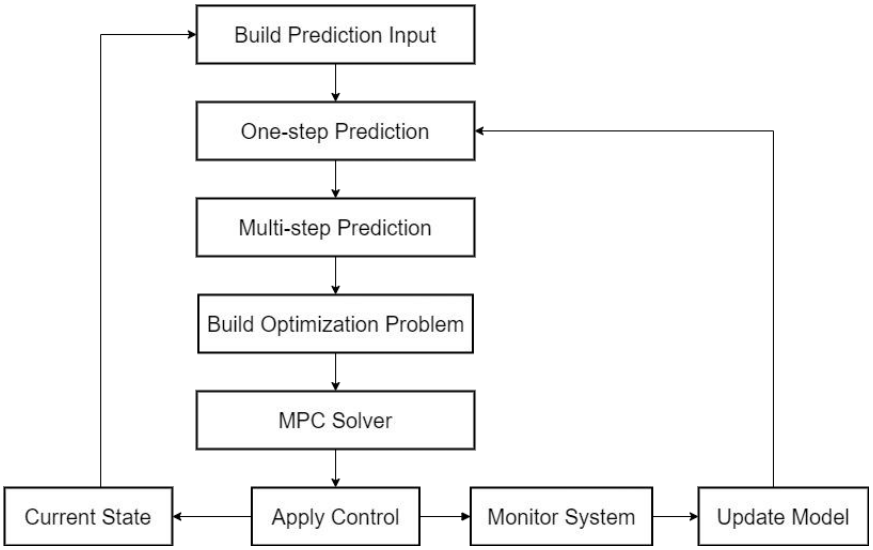


Figure 2 Closed-Loop NN-MPC Architecture

3 DESIGN OF THE NEURAL-NETWORK-ASSISTED MODEL PREDICTIVE CONTROL METHOD

The unified dynamic modeling framework developed for nuclear power plant actuation systems establishes a foundation for the construction of a Neural-Network-Assisted Model Predictive Control (NN-MPC) scheme. By exploiting the strong nonlinear function approximation capability of neural networks and the constraint-aware optimization structure of Model Predictive Control (MPC), the NN-MPC approach enables high-accuracy predictions and robust control of actuator systems characterized by nonlinear dynamics and multi-source disturbances. The combination of these two techniques effectively compensates for the limitations of conventional control strategies in fluid-coupled, highly constrained systems, making NN-MPC a promising direction for next-generation actuation control in nuclear power plants.

This chapter presents the complete NN-MPC methodology, including the design of the neural network prediction model, multi-step prediction construction, deviation compensation mechanisms, MPC optimization formulation, and the associated safety assurance strategies.

3.1 Structure of the Neural Network Prediction Model

Under the unified modeling framework, fans, pumps, and valves share the same neural-network architecture for dynamic prediction, enabling consistency and maintainability across heterogeneous devices. The neural network acts as a surrogate dynamic model that predicts the next-step state $\hat{x}(k+1)$ or output $\hat{y}(k+1)$ from the current state x(k), control input u(k), and operating condition variables. Considering the stringent real-time requirements of nuclear-class control loops, the prediction model must maintain a lightweight structure; hence 2–4-layer multilayer perceptrons (MLPs) or compact recurrent structures are typically adopted to ensure millisecond-level inference latency [10].

Using the unified input vector s(k), combined with the masking mechanism that filters out irrelevant channels, the model preserves device-specific characteristics while maintaining a common structure. The general neural network prediction model is formulated as

$$\hat{x}(k+1)=f_\theta\big(s(k)\big),\ \hat{y}(k+1)=\mathrm{h}\big(\hat{x}(k+1)\big) \tag{5}$$

where f_θ is a neural network parameterized by θ.

To satisfy the differentiability requirements of MPC solvers, smooth and fully differentiable activation functions such as ReLU or *tan*h are employed, enabling automatic differentiation to generate Jacobians and gradients for real-time optimization [11].

Given the safety-critical nature of nuclear applications, the prediction model must satisfy stability, interpretability, and traceability requirements. Therefore, fixed-length sliding windows are used instead of arbitrarily deep sequence structures to ensure verifiability. Furthermore, physical constraints must be embedded into the prediction outputs—for example, prohibiting negative flow rates or non-physical pressure predictions. To this end, post-processing layers, such

as positivity enforcement or sigmoid-based bounding, are incorporated to ensure physical plausibility at every prediction step.

3.2 Multi-Step Prediction and Receding-Horizon Optimization

MPC requires a future prediction horizon of multiple steps, while neural networks typically provide only single-step predictions. Multi-step prediction is therefore constructed through iterative rollout of the single-step model:

$$\hat{x}(k+i+1)=f_\theta(\hat{x}(k+i),u(k+i),\ldots) \tag{6}$$

In this recursive process, the prediction of the previous step becomes the input to the next. Although straightforward, recursive rollout is susceptible to error accumulation, especially under strong nonlinearities, potentially causing trajectory divergence. For nuclear-class applications, prediction stability is vital; thus, a deviation correction mechanism must be integrated to ensure consistency and robustness.

The receding-horizon framework forms the core of MPC: although a sequence of control inputs is optimized over the future horizon, only the first control action is applied:

$$u(k)=u^*(k|k) \tag{7}$$

and the optimization is reconstructed in the next sampling period. This mechanism enables the controller to continually adapt to system dynamics and disturbances.

3.3 Deviation Compensation and Robust Prediction

Neural network prediction accuracy may degrade due to limited training data, equipment aging, or gradual shifts in operating conditions. To ensure conformance with nuclear-class safety and reliability requirements, a two-level deviation compensation mechanism is adopted: (1) short-term deviation compensation and (2) long-term drift monitoring.

(1) Short-term deviation compensation

At each sampling instant, the most recent measured state $x(k)$ replaces the predicted state $\hat{x}(k|k\text{-}1)$, and a deviation term is computed:

$$\delta(k)=x(k)-\hat{x}(k|k\text{-}1) \tag{8}$$

This deviation is injected into future predictions:

$$\tilde{x}(k+i)=\hat{x}(k+i)+\delta(k) \tag{9}$$

effectively suppressing error propagation over the prediction horizon and enhancing stability.

(2) Long-term drift detection and model recalibration

If the deviation remains consistently large, indicating that the trained model no longer matches the real equipment dynamics, the system enters a semi-online recalibration mode. Selected neural network parameters or partial layers are updated using recent data to restore model fidelity [12].

Deviation compensation is essential for maintaining long-term stability of NN-MPC under component aging, gradual environmental changes, and operating-condition shifts. The deviation compensation process is illustrated in Figure 3.

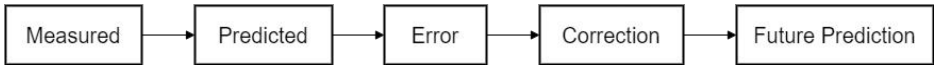


Figure 3 Flowchart of the Deviation Compensation Process

3.4 Formulation of the MPC Optimization Problem

The objective of MPC is to achieve accurate reference tracking while satisfying physical and safety constraints. The standard cost function is expressed as:

$$min \sum_{i=1}^{N_p} // y(k+i)\text{-}y_{\text{ref}}(k+i) //_Q^2 + \sum_{i=0}^{N_c\text{-}1} // \Delta u(k+i) //_R^2 \tag{10}$$

where N_p and N_c denote prediction and control horizons, respectively.

The control inputs and outputs of nuclear power actuators are subject to strict limits, which must be explicitly encoded as hard constraints:

$$\begin{gathered} u_{\min}\le u(k+i)\le u_{\max}, \\ \Delta u_{\min}\le \Delta u(k+i)\le \Delta u_{\max}, \\ y_{\min}\le y(k+i)\le y_{\max}. \end{gathered} \tag{11}$$

Because the neural network introduces nonlinearities, the resulting MPC problem becomes nonconvex and computationally demanding. To meet real-time requirements, fast optimization approaches such as real-time iteration Sequential Quadratic Programming (RTI-SQP) or ADMM-type accelerators are adopted. Automatic differentiation is employed to compute gradients and Jacobians efficiently, enabling millisecond-level solver execution [13].

3.5 Closed-Loop Architecture of the NN-MPC System

The complete closed-loop architecture of the NN-MPC system integrates sensing, neural-network prediction, deviation compensation, constrained optimization, and actuator execution. Its information flow is summarized in Figure 4.

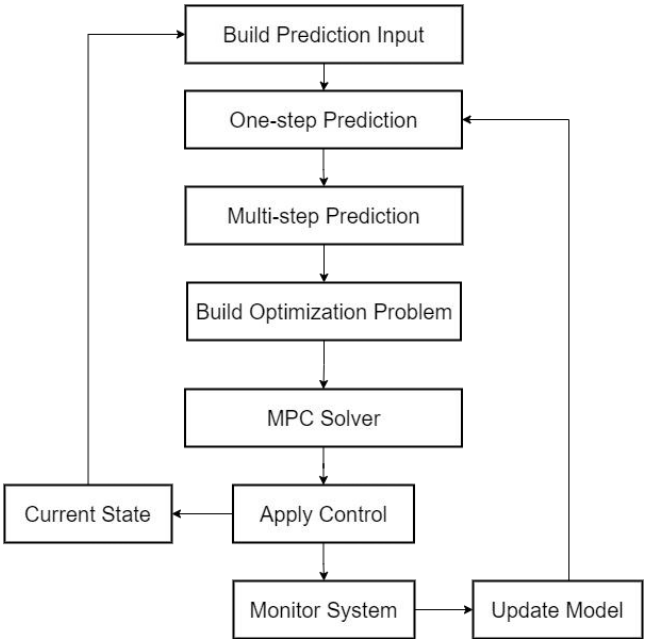


Figure 4 Closed-Loop Control Architecture of the NN-MPC System

This architecture highlights the tight interplay between data-driven prediction and optimization-based decision-making, forming a robust closed-loop suitable for safety-critical nuclear applications.

4 ENGINEERING DEPLOYMENT AND IMPLEMENTATION PATH OF NN-MPC IN NUCLEAR POWER PLANT ACTUATION SYSTEMS

The neural-network-assisted model predictive control (NN-MPC) framework offers high-accuracy nonlinear prediction and high-performance decision-making capabilities. However, deploying such an intelligent control framework in the actuation systems of nuclear power plants requires a rigorous engineering process. Challenges arise in control-system architecture adaptation, real-time computational assurance, hardware–software redundancy, model lifecycle management, version auditability, verification workflows, and coordination with safety-critical subsystems. Nuclear power plants represent mission-critical environments in which Digital Control Systems (DCS) must maintain exceptional reliability, traceability, maintainability, and fault isolation. Therefore, NN-MPC must be integrated in a manner that does not compromise established safety boundaries. Such integration extends far beyond algorithmic efficiency; it encompasses system-level capabilities for anomaly detection, graceful degradation, and safety isolation—hallmarks of nuclear engineering's uncompromising safety standards.

4.1 Layered Control Architecture and the Integration Strategy of NN-MPC

Industrial nuclear control systems typically follow a multilayer hierarchical structure comprising field devices, actuation and drive units, process control layers, and supervisory and coordination layers. This structure, formalized in industry standards, ensures that sensing, actuation, control, and supervisory functions are separated according to their timing and safety requirements. Accordingly, NN-MPC must adhere to this layered architecture and is positioned between the process-control layer and the plant's actuation units. It shall not coexist parallel to the reactor protection system (RPS) nor interfere with any safety-grade interlocks.

Within this architecture, the NN-MPC module receives real-time sensor measurements—valve positions, pump rotational speed, fan differential pressure, flow rate, and other process variables. The prediction module uses these inputs to estimate future states, and the MPC optimizer subsequently computes the admissible control action under explicit constraints. Unlike PID controllers, NN-MPC relies on richer information and more sophisticated preprocessing to forecast system behavior and capture fluid-dynamic coupling.

The functional position of NN-MPC in this hierarchical structure is illustrated in Figure 5. It operates as the core algorithmic module in the process-control layer, bridging sensors and actuators while remaining isolated from all nuclear safety logics.

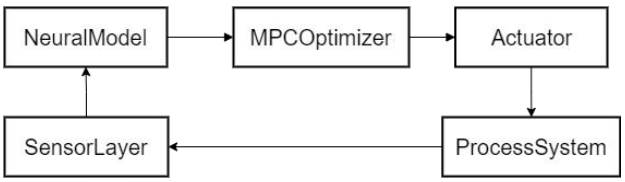


Figure 5 Integration of the NN-MPC Module within the Hierarchical Control Architecture of a Nuclear Power Plant

4.2 Engineering Assurance of Real-Time Operation and Computational Resources

Actuation systems in nuclear power plants—such as regulating valves, feedwater pumps, and cooling fans—typically operate with sampling periods on the order of tens to hundreds of milliseconds. Consequently, NN-MPC must complete data acquisition, model inference, gradient generation, constrained optimization, and command dispatch within this limited time window. Achieving this requirement demands optimizations across several engineering dimensions.

First, neural-network models deployed in the field are compressed through parameter pruning, reduced-precision floating-point formats, and activation-efficient architectures (e.g., ReLU) so that inference can run on industrial-grade controllers within milliseconds. Second, the MPC layer employs real-time iteration sequential quadratic programming (RTI-SQP), which performs only one linearization and one optimization step per sampling period. Automatic differentiation (AD) is used to derive Jacobians and gradients to maintain accuracy without manual linearization—an essential requirement for safety-critical nuclear applications.

NN-MPC must also conform to the DCS communication infrastructure, including redundant field buses, synchronized timing networks, and dual-channel communication pathways. If the NN-MPC module fails to produce an output within the scheduled control window, the system must automatically switch to a conservative fallback strategy, such as PID control or a predefined safe setpoint. This guarantees fail-safe behavior and ensures that NN-MPC is never a single point of failure [6].

The computational timing constraint can be formalized as:

$$TNN+TAD+TMPC+TIO \leq Ts \tag{12}$$

where:

- Ts: sampling period
- TNNT: neural-network forward-inference time
- TAD: automatic-differentiation time
- TMPCT: optimization solver time
- TIOT: input/output latency (sensor readout + actuator command)

This inequality is a core criterion in real-time verification reviews for nuclear-grade control software.

Deployment of Real-Time Iteration Solvers

Because the underlying system is nonlinear and constrained, NN-MPC must solve a nonlinear MPC problem at each step. In the RTI-SQP framework, each control cycle solves a single linearized subproblem:

$$H(k)\Delta U = -\nabla J(k) \tag{13}$$

where:

- $H(k)$: Gauss–Newton approximation of the Hessian
- $\nabla J(k)$: gradient of the cost function
- ΔU: control update

The fully differentiable neural network provides the required Jacobians directly via AD, enabling predictable computation and compliance with DCS certification procedures.

4.3 Model Version Management and Auditability

As data-driven components, neural networks require rigorous lifecycle management to ensure that every deployed model is transparent, traceable, and auditable. Each NN model version must undergo independent verification, including data-source auditing, training-process documentation, parameter integrity inspection, and behavioral boundary testing. Nuclear software standards mandate complete traceability; NN-MPC must therefore provide verifiable evidence for any prediction deviation or anomalous system response.

To achieve this, an engineering "model repository" is established to maintain version metadata, training datasets, training hyperparameters, and verification reports. Before field deployment, every model version must pass multistage validation—simulation testing, hardware-in-the-loop (HIL) experiments, and formal compliance reviews. Any model update triggers DCS consistency checks to prevent behavioral discontinuities in the actuation systems.

Model consistency across versions is validated using:

$$M_{v_i}(s) \approx M_{v_j}(s), \forall s \in D_{\text{valid}} \tag{14}$$

ensuring that a new version does not introduce unexpected behavioral deviations over the validated operational domain.

Digital signatures, version locks, and strict repository policies ensure that the prediction model, optimizer, and gradient interfaces remain synchronized, thus meeting the requirements of nuclear software configuration management (SCM).

4.4 Anomaly Detection and Safety-Fallback Mechanisms

Given the potential failure modes of data-driven models, NN-MPC must incorporate robust anomaly detection and multi-layer safety fallback mechanisms. Anomalies are identified through prediction deviation, drift detection, and out-of-distribution (OOD) analysis. During extreme or unforeseen operating conditions—fan blockage, pump cavitation, or valve friction escalation—prediction errors increase sharply, triggering safety logic.
Fallback strategies typically include:
1. Switching to validated PID or linear MPC
2. Safely clamping actuators to predefined safe operating points
a) e.g., valves to a predefined safe opening
b) pumps to low-speed mode
c) fans to fixed airflow mode
These mechanisms ensure that even if NN-MPC becomes unavailable, the actuation system maintains thermal-hydraulic stability.
Prediction-error-based health monitoring is defined as:

$$e(k)=// x(k)-\hat{x}(k) //\tag{15}$$

If:

$$e(k)>\epsilon_{safe}\tag{16}$$

the system triggers fault-handling logic to switch to fallback control or reload a stable model version.
All anomaly events are logged for subsequent engineering review and provide critical evidence during nuclear-safety audits (defense-in-depth compliance).

4.5 HIL Verification and Multi-Stage Testing Framework

Before deployment on real equipment, NN-MPC must pass simulation-based verification and hardware-in-the-loop (HIL) testing. This process is essential not only in the nuclear industry but also for certifying data-driven predictive models in safety-critical control environments.
First, unified simulation models of fans, pumps, and valves are used to evaluate NN-MPC performance across a wide operating envelope. Subsequently, the NN-MPC controller is deployed on a real-time industrial controller connected to a HIL plant emulator. Extreme operating scenarios—motor faults, sensor noise spikes, sudden hydraulic-resistance changes—are simulated to verify stability, constraint satisfaction, and correct triggering of fallback logic.
Only after all verifications pass can the system proceed to field deployment.

4.6 Engineering Commissioning and On-Site Deployment Workflow

Deploying NN-MPC in a nuclear power plant requires coordination across plant operators, control engineers, commissioning teams, software-verification units, and third-party reviewers. The deployment process comprises three major phases:
a) Offline preparation:
b) model training, parameter tuning, simulation testing, HIL validation, DCS model instantiation
c) Semi-online commissioning:
d) NN-MPC receives real sensor inputs but operates in a “shadow mode,” not directly controlling actuators.
e) Its outputs are compared with operator decisions or baseline controllers to verify consistency.
f) Full deployment:
g) NN-MPC assumes full control after all safety audits and verification processes are completed.
After deployment, all predictions, control outputs, and anomalies are logged for long-term monitoring. If any abnormal behavior is detected, immediate fallback and isolation procedures guarantee safe operation.

4.7 Engineering Value and Future Prospects of NN-MPC

The successful deployment of NN-MPC brings significant benefits to nuclear-plant operation. By establishing a unified modeling and control framework, fans, pumps, and valves can achieve enhanced stability, improved tracking performance, and more efficient power regulation. The model-based predictive structure combined with neural-network nonlinear estimation offers an upgrade path toward more intelligent and adaptive control systems.
NN-MPC’s explainability, explicit constraint handling, and built-in safety mechanisms satisfy nuclear industry requirements for transparency and reliability, demonstrating that data-driven components can be safely integrated into nuclear-grade control systems.
Future nuclear power plants—including advanced reactors and next-generation digital control systems—will benefit from NN-MPC in more critical components such as reactor coolant pumps, main feedwater pumps, and steam-generator

secondary-side flow-regulation valves. The integration of NN-MPC with digital-twin platforms will further enable online model adaptation, structural optimization, and autonomous performance enhancement, marking a significant step toward intelligent, self-optimizing nuclear control systems.

5 CONCLUSION AND FUTURE PERSPECTIVES

This study addresses the nonlinear, strongly coupled, multi-constrained, and high-safety requirements inherent in the control of actuation systems within nuclear power plants. To meet these challenges, we propose a novel control framework that integrates unified dynamic modeling with neural-network-assisted model predictive control (NN-MPC). The work provides a comprehensive discussion spanning theoretical modeling, algorithmic design, engineering deployment, and safety verification. Actuation systems—such as fans, pumps, and valves—play essential roles in reactor cooling, steam-generator feedwater supply, ventilation and heat exchange, and the stable operation of auxiliary subsystems. Their performance directly affects thermal-hydraulic balance, energy-conversion efficiency, and safety margins of nuclear generating units. Consequently, developing an advanced control methodology that is verifiable, robust, and predictive is of substantial significance for the nuclear control community.

The first contribution of this work is the formulation of a unified modeling framework for nuclear actuation systems. Fans, pumps, and valves are abstracted into a consistent input–state–output structure, with a Mask-based mechanism that filters device-irrelevant features. This allows cross-device model sharing and consistent input encoding. Compared with conventional physics-based modeling—which often becomes prohibitive in complex flow-coupling scenarios due to high dimensionality, difficulty of generalization, and expensive calibration—this unified approach reduces model quantity, lowers maintenance cost, and resolves traceability challenges that hinder engineering validation. The framework provides both a data foundation and a structural foundation for NN-MPC, while ensuring model auditability and facilitating systematic lifecycle management and certification in nuclear-grade environments [4].

In terms of control-method design, this study develops an NN-MPC architecture based on single-step prediction, bias compensation, and rolling optimization. Neural networks leverage the unified input representation to approximate nonlinear actuation dynamics with high fidelity, while MPC provides explicit handling of input and output constraints together with interpretable optimization-based decision making. The proposed bias-correction mechanism effectively mitigates error accumulation during multi-step prediction, enabling the controller to remain stable under equipment aging, operating-condition variations, and dynamic changes in pipeline resistance. Moreover, the integration of a real-time iteration solver (RTI-SQP) substantially reduces computational overhead, ensuring that NN-MPC satisfies the strict real-time requirements of nuclear actuation loops.

A complete engineering deployment scheme tailored for nuclear power plants is also proposed. It includes real-time optimization strategies, model version management, anomaly detection, safety fallback mechanisms, and multi-stage testing workflows. A model repository with strict version control enables traceable auditing of neural-network models and MPC configurations. Out-of-distribution (OOD) detection and drift monitoring allow the system to promptly recognize model degradation. Robust fallback strategies—switching to PID or conservative MPC—guarantee that actuation behavior always remains within safety boundaries. Additionally, extensive simulation and hardware-in-the-loop (HIL) validation ensure that NN-MPC performs safely even under extreme operating conditions, providing engineering evidence compatible with nuclear-grade safety requirements.

From an industry perspective, the proposed NN-MPC framework provides essential methodological support for the digitalization and intelligent upgrade of nuclear power plants. As nuclear systems become more complex and as expectations for extended operational lifetimes and enhanced load-following capabilities increase, traditional control solutions are no longer sufficient to meet advanced performance and lifecycle management requirements. By combining data-driven modeling with constraint-aware optimization, NN-MPC offers superior generalization, adaptability, and situational responsiveness, thereby offering an efficient and robust control solution for next-generation nuclear energy systems.

Future research directions may be explored along three lines.

First, with ongoing advances in machine-learning architectures, new lightweight, interpretable, or safety-oriented neural networks may be integrated to further enhance the suitability of NN-MPC for nuclear actuation systems.

Second, digital-twin technology can be incorporated into the control loop to enable online simulation and model calibration, providing NN-MPC with more accurate and timely dynamic information to improve responsiveness to abnormal or fast-evolving events.

Third, distributed MPC or partitioned control architectures may be considered for NN-MPC deployment in scenarios with stronger subsystem coupling—such as coordinated control of main and auxiliary feedwater systems or reactor coolant pumps and regulating valves—to support more advanced strategies for economic operation and load-following.

Overall, the proposed NN-MPC framework establishes a complete theoretical and practical foundation encompassing unified modeling, predictive modeling, optimization-based control, bias compensation, safety mechanisms, and engineering deployment. The framework not only advances the unification and intelligence of actuation-system control in nuclear power plants but also ensures verifiability, safety, and deployability at the engineering level. As intelligent control technologies continue to penetrate the nuclear industry, NN-MPC holds strong potential to become a core control strategy for future nuclear power plants, providing key technical support for safe, efficient, and intelligent operation of nuclear energy systems.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work is supported by the Talent Fund Project No. 25816 of China Institute of Atomic Energy (CIAE).

REFERENCES

[1] IAEA. Instrumentation and Control Systems Important to Safety in Nuclear Power Plants. IAEA Safety Guide NS-G-1.3, International Atomic Energy Agency, 2021.
[2] Cho N Z. Dynamic characteristics of nuclear power plant systems. Nuclear Engineering and Design, 2019, 350: 72–84.
[3] Isermann R. Actuators for process control: modeling, design and applications. IFAC Annual Reviews in Control, 2018, 46: 1–12.
[4] Li Y, Wang X, Wang Z. Unified modeling of fluid actuators under complex boundary conditions. Journal of Process Control, 2022, 108: 19–33.
[5] Rawlings J B, Mayne D Q, Diehl M. Model Predictive Control: Theory, Computation, and Design. Nob Hill Publishing, 2017.
[6] Camacho E F, Bordons C. Model Predictive Control. Springer, 2013.
[7] Chen Y, Yin S, Gao H. Data-driven predictive control for nonlinear industrial processes: methods and applications. IEEE Transactions on Industrial Electronics, 2021, 68(7): 5881–5891.
[8] Parisini T, Zoppoli R. Neural approximations for nonlinear optimal control problems. Automatica, 1995, 31(10): 1443–1455.
[9] Raissi M, Perdikaris P, Karniadakis G. Physics-informed neural networks: A deep learning framework for solving nonlinear differential equations. Journal of Computational Physics, 2019, 378: 686–707.
[10] Diehl M, Ferreau H J, Houska B. Efficient numerical methods for nonlinear MPC and moving horizon estimation. In: Nonlinear Model Predictive Control. Springer, 2009.
[11] Lataire R, Butcher M, Nandola N. Real-time nonlinear MPC using automatic differentiation and neural surrogate models. Journal of Process Control, 2020, 92: 31–45.
[12] Zhang X, Jiang B, Shi Z. Fault-tolerant control of safety-critical systems: methods and challenges. Annual Reviews in Control, 2020, 49: 88–102.
[13] Wang W, Chen G, Hou Z. Safety-aware data-driven predictive control for critical energy systems. IEEE Transactions on Control Systems Technology, 2022, 30(6): 2312–2324.
[14] Park HJ, Kim D. Digital-twin-driven control optimization for nuclear power plant thermal-hydraulic systems. Annals of Nuclear Energy, 2023, 182: 109740.

Eurasia Journal of Science and Technology
Print ISSN: 2663-1024
Online ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3123

AN INTEGRATED MULTI-SOURCE INFORMATION FUSION APPROACH FOR ENTERPRISE USER PORTRAIT CONSTRUCTION IN TECHNOLOGICAL DEMAND IDENTIFICATION

HongYu Su
China National Institute of Standardization, Beijing 100191, China.

Abstract: In the context of the innovation-driven development paradigm, the accurate identification of enterprise technological demands has become a core theoretical and practical issue in the field of technology transfer. Traditional demand identification paradigms are constrained by single-dimensional text analysis, leading to the lack of a systematic understanding of technological demands. To address this gap, this study proposes an integrated multi-source information fusion approach for enterprise user portrait construction oriented to technological demand identification. This approach systematically clarifies the connotation of enterprise technological demand portraits, defines the multi-dimensional information sources of portrait construction based on methodological logic, establishes a feature extraction system based on the dual dimensions of explicit and potential demands, and explores the mechanism of technological demand reconstruction through multi-source information fusion. This study enriches the methodological system of user portrait and technological demand identification, and provides technical guidance for breaking the information asymmetry in technology transfer.
Keywords: Technological demand; Enterprise user portrait; Multi-source information fusion approach; Demand identification; Feature extraction

1 INTRODUCTION

1.1 Research Background

With the in-depth advancement of the innovation-driven development strategy, enterprises, as the main body of technological innovation, their technological demands have become the core link connecting technological supply and industrial application[1]. However, the current technology transfer field is plagued by information asymmetry, which essentially stems from the lack of a systematic and in-depth methodological support for technological demand identification[2]. Traditional demand identification paradigms rely excessively on explicit text expression, ignoring the potential demand connotation contained in multi-dimensional information, and failing to form a holistic and in-depth understanding of enterprise technological demands[3].

The fundamental limitation of traditional paradigms lies in the one-sidedness of information sources and the shallowness of demand mining[4]. From the methodological perspective, technological demand is a multi-dimensional concept integrating explicit expression, potential implication and constraint conditions, which cannot be fully captured by a single information source[5]. Multi-source information, including explicit demand expression, behavioral implication and attribute constraints, has inherent complementary and verifiable characteristics, which provides a methodological possibility for breaking through the limitations of traditional paradigms[6]. Therefore, developing an integrated multi-source information fusion approach for enterprise user portrait construction oriented to technological demand identification has become an important technical path to solve the problem of information asymmetry in technology transfer.

1.2 Research Significance

The theoretical significance of this study lies in three aspects: first, it clarifies the connotation and multi-dimensional structure of enterprise technological demand portraits, which helps to enrich the methodological system of user portrait research in the specific field of technology transfer; second, it proposes an integrated multi-source information fusion approach for demand identification, breaking through the methodological limitations of single-dimensional text-driven demand identification, and providing a new technical perspective for in-depth demand mining; third, it explores the mechanism of technological demand reconstruction based on multi-source information fusion, which helps to clarify the logical relationship between multi-source information integration and accurate demand identification, and lay a methodological foundation for the improvement of technology transfer practice.

From the perspective of practical application, this study provides systematic technical guidance for the practice of enterprise technological demand identification. The proposed integrated approach can help technology transfer institutions establish a scientific demand cognition system, avoid the one-sidedness of demand understanding caused by single information sources; at the same time, it can guide enterprises to clarify their own technological demand

connotation, and promote the accurate matching between technological supply and demand at the operational level, which has important guiding significance for optimizing the efficiency of technology transfer.

1.3 Research Framework

This study follows the methodological research logic of "concept definition - problem analysis - approach development - mechanism exploration". Firstly, it clarifies the core connotation and boundary of enterprise technological demand portraits; secondly, it analyzes the methodological limitations of traditional demand identification paradigms and the necessity of multi-source information integration; thirdly, it develops the overall integrated multi-source information fusion approach for portrait construction, including information source dimension definition, feature extraction system and demand reconstruction mechanism; finally, it discusses the theoretical contribution and application extension of the approach. The full text focuses on methodological derivation and logical construction, forming a complete technical research chain.

2 CONNOTATION AND RESEARCH OBJECTIVES

2.1 Connotation of Enterprise Technological Demand Portraits

Enterprise technological demand portrait, in the methodological sense, refers to a systematic and structured abstraction of enterprise technological demand connotation, which is constructed through the integration and induction of multi-source information. Its core characteristics are as follows: first, systematicness, which integrates the multi-dimensional connotation of technological demand, including explicit demand, potential demand and constraint conditions, and avoids one-sided understanding of demand; second, operability, which provides clear technical indicators for demand identification rather than abstract theoretical description; third, guiding nature, which provides technical guidance for the identification and matching of technological demands, and is the core bridge connecting technological supply and demand.

From the theoretical structure, enterprise technological demand portrait is composed of three core dimensions: core demand dimension, potential demand dimension and constraint dimension. The core demand dimension reflects the explicit technological needs of enterprises, which is the direct embodiment of enterprise technology upgrading and development needs; the potential demand dimension reflects the implicit technological needs contained in enterprise behavior and development strategy, which has the characteristics of concealment and derivation; the constraint dimension reflects the objective conditions that restrict the satisfaction of technological demands, which is an important basis for the feasibility of technological demand matching.

2.2 Research Objectives

The core objective of this study is to develop a systematic and operable integrated multi-source information fusion approach for enterprise user portrait construction oriented to technological demand identification. The specific objectives include: first, clarifying the connotation, boundary and structural system of enterprise technological demand portraits, laying a solid foundation for subsequent methodological development; second, defining the multi-source information source system for portrait construction based on methodological logic, clarifying the basis and scope of each information dimension; third, establishing a feature extraction system based on the dual dimensions of explicit and potential demands, specifying the technical path and logical process of feature extraction; fourth, exploring the mechanism of technological demand reconstruction based on multi-source information fusion, and clarifying the logical relationship between multi-source information integration and accurate demand identification.

3 THEORETICAL BASIS AND LITERATURE REVIEW

3.1 Theoretical Basis

3.1.1 User portrait theory

User portrait theory is the core theoretical basis of this study. It originated from the field of marketing, and its core idea is to abstract the characteristics of users through the integration of multi-dimensional information, so as to realize the accurate understanding of users. In the field of technology transfer, the application of user portrait theory needs to be combined with the characteristics of technological demand[7]. Different from the consumer demand in the marketing field, enterprise technological demand has the characteristics of professionalism, complexity and long-term, which requires the portrait construction to focus on the professional connotation of technology and the strategic development needs of enterprises[8]. The application of this theory in this study focuses on the extension of user portrait theory to the specific field of technological demand, and develops a portrait system that conforms to the characteristics of enterprise technological demand and has strong operability.

3.1.2 Multi-source information fusion theory

Multi-source information fusion theory provides the core methodological support for this study. The core idea of this theory is to integrate information from different sources through specific technical methods to realize the complementary advantages of information and improve the accuracy and comprehensiveness of information

understanding[9]. In the process of enterprise technological demand portrait construction, multi-source information has the characteristics of heterogeneity and complementarity. Multi-source information fusion theory provides a methodological basis for solving the problems of information redundancy and conflict in the integration process, and is the key technical support for realizing in-depth demand mining[10].

3.1.3 Technological demand identification theory

Technological demand identification theory clarifies the core connotation and identification logic of technological demand. This theory holds that technological demand is the product of the interaction between enterprise development strategy and external environment, and its identification needs to take into account both explicit demand expression and implicit demand implication[11]. The theoretical basis of this study focuses on the integration of technological demand identification theory and multi-source information fusion theory, and develops a portrait-based demand identification approach, which helps to improve the systematicness and accuracy of demand identification.

3.2 Literature Review

Existing research on enterprise technological demand identification mainly focuses on two directions: single-dimensional text analysis and multi-source data-driven analysis. In the field of single-dimensional text analysis, scholars have carried out a lot of research on demand extraction based on text mining technology, and developed a variety of text-driven demand identification methods. However, these studies are limited by the one-sidedness of information sources, and it is difficult to fully capture the potential connotation of technological demand[12].

In the field of multi-source data-driven analysis, some scholars have begun to explore the integration of multi-dimensional data to identify technological demands, and initially developed a multi-source data integration framework. However, existing research still has obvious methodological limitations: first, the connotation and structural system of technological demand portraits are not clearly defined, and the methodological basis of portrait construction is insufficient; second, the selection of multi-source information sources lacks systematic methodological guidance, and the logical relationship between information sources and demand dimensions is not clear; third, the mechanism of multi-source information fusion and demand reconstruction is not deeply explored, and the technical logic of demand identification is not complete[13].

In the field of user portrait research, existing studies are mostly concentrated in the field of marketing and e-commerce, and there are few studies on user portrait construction in the specific field of technology transfer. The existing related research lacks the integration with technological demand characteristics, and cannot meet the methodological needs of enterprise technological demand identification. Therefore, there is an urgent need to develop a systematic integrated multi-source information fusion approach for enterprise user portrait construction oriented to technological demand identification, which is also the core methodological gap that this study aims to fill[14].

4 INTEGRATED MULTI-SOURCE INFORMATION FUSION APPROACH FOR ENTERPRISE USER PORTRAIT CONSTRUCTION

4.1 Overall Architecture of the Integrated Approach

Based on the theoretical basis of user portrait theory, multi-source information fusion theory and technological demand identification theory, this study develops an overall architecture of the integrated multi-source information fusion approach for enterprise user portrait construction, which is composed of four core modules: multi-source information source definition, demand feature extraction, multi-source information fusion and technological demand reconstruction. The logical relationship between each module is as follows: multi-source information source definition is the foundation of the approach, which clarifies the scope and connotation of information required for portrait construction; demand feature extraction is the core module of the approach, which abstracts demand features from multi-source information through specific technical methods; multi-source information fusion is the key technical means to realize in-depth demand understanding, which integrates multi-dimensional feature information to avoid information redundancy and conflict; technological demand reconstruction is the ultimate goal of the approach, which forms a systematic and structured demand understanding through the integration of multi-dimensional features.

The core technical logic of the approach is: taking the accurate identification of enterprise technological demand as the core goal, taking multi-source information integration as the technical means, through the systematic extraction and fusion of multi-dimensional information features, to realize the structured abstraction of enterprise technological demand, and construct a systematic and operable user portrait, so as to provide technical support for breaking the information asymmetry in technology transfer.

4.2 Definition of Multi-source Information Source Dimension

Based on the connotation of enterprise technological demand and the operational requirements of portrait construction, this study divides the multi-source information sources of enterprise user portrait construction into four core dimensions from a methodological perspective: explicit demand expression information, behavioral implication information, attribute constraint information and strategic guidance information. Each dimension has its unique connotation and demand reflection function, and provides targeted data support for the subsequent feature extraction module.

4.2.1 Explicit demand expression information

Explicit demand expression information refers to the direct expression of enterprise technological demand, which is the most direct data basis for understanding explicit demand. From the methodological perspective, this type of information reflects the core technological needs of enterprises for technology upgrading and development, and its core connotation includes the direction of technological demand, the core technical problems to be solved, and the expected goals of technology application. The technical significance of this dimension lies in providing a direct data basis for the extraction of core demand features.

4.2.2 Behavioral implication information

Behavioral implication information refers to the implicit information related to technological demand contained in enterprise behavior. From the methodological perspective, enterprise behavior is the external manifestation of its internal demand, and the behavioral information related to technology has inherent implication for potential technological demand. The core connotation of this dimension includes the attention behavior to technological information, the search behavior for technical solutions, and the interactive behavior with technology suppliers. The technical significance of this dimension lies in supplementing the limitations of explicit demand expression and providing a data basis for the extraction of potential demand features.

4.2.3 Attribute constraint information

Attribute constraint information refers to the objective attribute information of enterprises that restricts the satisfaction of technological demand. From the methodological perspective, the satisfaction of technological demand cannot be separated from the objective conditions of enterprises. Attribute constraint information determines the feasibility of technological demand matching. Its core connotation includes enterprise industry attributes, resource capacity attributes, and operation status attributes. The technical significance of this dimension lies in providing a data basis for the extraction of demand constraint features and ensuring the feasibility of demand identification.

4.2.4 Strategic guidance information

Strategic guidance information refers to the information related to technological development strategy contained in enterprise development planning. From the methodological perspective, enterprise technological demand is closely related to its long-term development strategy, and strategic guidance information can reflect the long-term and directional characteristics of technological demand. Its core connotation includes enterprise R&D planning, industrial upgrading strategy, and market expansion strategy. The technical significance of this dimension lies in improving the depth and long-term of demand understanding, and providing a data basis for the extraction of long-term demand features.

4.3 Construction of Demand Feature Extraction Module

Based on the multi-source information source dimension and the structural system of enterprise technological demand portraits, this study constructs a demand feature extraction module based on the dual dimensions of explicit and potential demands, which includes three core feature categories: core demand features, potential demand features and constraint features. The extraction of each category of features follows the technical logic of "information dimension - feature connotation - extraction method".

4.3.1 Extraction of core demand features

Core demand features are extracted based on explicit demand expression information and strategic guidance information, reflecting the explicit and directional connotation of enterprise technological demand. From the methodological perspective, the extraction of core demand features follows the technical path of "demand expression abstraction - strategic alignment - feature definition", with detailed technical details as follows:

First, in the stage of demand expression abstraction, a two-step text preprocessing process is adopted for explicit demand expression information (such as technical demand declarations and R&D planning texts). First, noise reduction is implemented through stop-word removal (using domain-specific stop-word lists, e.g., common functional words in technical documents) and part-of-speech filtering (retaining nouns, verbs, and adjectives related to technical concepts). Then, keyword extraction is performed using the TF-IDF algorithm combined with domain thesaurus weighting—specifically, the weight of terms included in the industry technological thesaurus is increased by 1.5 times to enhance the recognition accuracy of professional technical terms. The top 30 terms with the highest composite weights are selected as preliminary demand feature candidates.

Second, in the strategic alignment stage, a semantic similarity matching model based on BERT is constructed to align the preliminary feature candidates with the enterprise's strategic guidance information. The model is pre-trained on a large-scale corpus of enterprise technological strategy texts, and fine-tuned using domain-specific data to optimize the semantic representation of technical concepts. The alignment threshold is set to 0.7 (based on cosine similarity), and only candidates with similarity scores exceeding the threshold are retained to ensure that the extracted core demand features are consistent with the enterprise's long-term development strategy.

Finally, in the feature definition stage, semantic clustering is implemented using the K-means algorithm to eliminate redundant features. The number of clusters K is determined by the elbow method, and each cluster is labeled with a core technical concept (e.g., "intelligent manufacturing equipment" "green energy-saving technology") to form structured core demand features, including technological field features (clustered core concepts), technical goal features (derived from semantic analysis of demand goals, e.g., "efficiency improvement" "cost reduction"), and development stage matching features (aligned with strategic planning cycles, e.g., "short-term R&D" "long-term industrialization").

4.3.2 Extraction of potential demand features

Potential demand features are extracted based on behavioral implication information and strategic guidance information, reflecting the implicit and derivative connotation of enterprise technological demand. The technical path of "behavioral implication analysis - demand derivation - feature abstraction" is refined with specific technical details as follows:

In the behavioral implication analysis stage, behavioral sequence analysis is implemented using the Hidden Markov Model (HMM). Enterprise behavioral data (browsing, search, interaction) is converted into a sequence of behavioral states (e.g., "browsing technical achievements" "searching for technical solutions" "consulting experts"), and the transition probability between states is calculated. States with transition probabilities exceeding 0.6 are identified as high-correlation behavioral chains, which are considered to reflect potential demand tendencies. For example, the chain "searching for photovoltaic technology - browsing photovoltaic module test cases - consulting photovoltaic energy storage experts" indicates a potential demand for photovoltaic energy storage technology.

In the demand derivation stage, a preference inference model based on collaborative filtering is constructed. The model takes the high-correlation behavioral chains as input, and infers the enterprise's preference intensity for different technical directions by comparing the behavioral similarity with benchmark enterprises (with clear technological demand orientations). The preference intensity is quantified using a 5-point scale, where scores ≥4 indicate high-potential demand directions. Meanwhile, combined with strategic guidance information, the long-term derivation of potential demands is realized—for example, if the enterprise's strategic planning emphasizes "carbon neutrality", the potential demand for photovoltaic technology is further derived as "low-cost photovoltaic energy storage integration technology".

In the feature abstraction stage, the inferred potential demand directions are encoded into structured features using one-hot encoding combined with weight assignment. The weight of each potential demand feature is determined by the product of behavioral chain correlation coefficient and strategic alignment score, forming features such as technological attention preference features (e.g., "photovoltaic energy storage: weight 0.85"), demand derivative features (e.g., "low-cost integration: weight 0.72"), and long-term development potential demand features (e.g., "carbon-neutral oriented photovoltaic technology: weight 0.91").

4.3.3 Extraction of constraint features

Constraint features are extracted based on attribute constraint information, reflecting the objective constraint conditions of enterprise technological demand satisfaction. The technical path of "attribute analysis - constraint identification - feature definition" is supplemented with detailed technical operations as follows:

In the attribute analysis stage, attribute dimension reduction is implemented using the Principal Component Analysis (PCA) algorithm to eliminate redundant attribute information. The input attributes include enterprise industry category, registered capital, R&D investment ratio, number of R&D personnel, qualification certification level, etc. The PCA algorithm retains the principal components with cumulative variance contribution rate ≥85%, which are identified as core attribute dimensions affecting technological demand satisfaction.

In the constraint identification stage, a fuzzy clustering-based constraint classification method is adopted. The core attribute dimensions are clustered using the FCM (Fuzzy C-Means) algorithm, and each cluster is labeled with a constraint type based on domain knowledge. For example, clusters with low R&D investment ratio (≤3%) and few R&D personnel (≤10) are labeled as "resource capacity constraints"; clusters in traditional industries with no high-tech certification are labeled as "industry attribute constraints".

In the feature definition stage, feature coding is performed for each constraint type. For quantitative constraint indicators (e.g., R&D investment ratio), a 3-level coding is adopted (low: 0, medium: 1, high: 2) based on industry average values; for qualitative constraint indicators (e.g., qualification certification), binary coding is adopted (yes: 1, no: 0). Finally, structured constraint features are formed, including resource capacity constraint features (e.g., "R&D investment: low (0)"), industry attribute constraint features (e.g., "traditional industry (1)"), and operation status constraint features (e.g., "stable operation (1)"). These features provide an important technical basis for the feasibility judgment of subsequent technological demand matching.

4.4 Multi-source Information Fusion and Technological Demand Reconstruction Mechanism

Multi-source information fusion and technological demand reconstruction are the core technical links to realize the accurate identification of enterprise technological demand. This study develops a multi-source information fusion mechanism based on "layered fusion - semantic alignment - conflict resolution", and realizes the reconstruction of enterprise technological demand through this mechanism.

4.4.1 Layered fusion mechanism

Layered fusion mechanism refers to the step-by-step fusion of multi-source information according to the level of information abstraction, which includes three core layers with detailed technical implementation details:

First, information level fusion (data preprocessing layer). This layer adopts a unified data standardization method for multi-source original information: for text-type information (explicit demand expression, strategic guidance), it is converted into a 768-dimensional vector using the BERT pre-trained model; for behavioral sequence information, it is converted into a state transition matrix (dimension: number of behavioral types × number of behavioral types); for attribute constraint information, it is normalized to [0,1] using the min-max scaling method. After standardization, the multi-source information is integrated into a unified data pool, and duplicate data is removed using the SimHash algorithm (similarity threshold: 0.95) to complete preliminary data cleaning.

Second, feature level fusion (core fusion layer). This layer adopts a weighted feature concatenation method to integrate

core demand features, potential demand features, and constraint features. The weight of each feature type is determined by the information gain ratio: core demand features (information gain ratio ≥0.6) are assigned a weight of 0.5, potential demand features (information gain ratio 0.3-0.6) are assigned a weight of 0.3, and constraint features (information gain ratio ≥0.4) are assigned a weight of 0.2. The weighted features are concatenated to form a multi-dimensional feature vector (dimension: sum of dimensions of each feature type), and the L2 normalization is performed to eliminate the influence of feature scale differences.

Third, decision level fusion (result integration layer). This layer adopts the TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) multi-criteria decision-making method. The ideal solution and negative ideal solution are constructed based on the multi-dimensional feature vector, and the Euclidean distance between each feature vector and the ideal/negative ideal solution is calculated. The relative closeness degree of each feature vector to the ideal solution is used as the demand confidence score (range: [0,1]). Feature vectors with confidence scores ≥0.7 are selected to form a valid demand feature set, which provides a basis for subsequent technological demand reconstruction.

4.4.2 Semantic alignment mechanism

Semantic alignment mechanism is the key to solving the heterogeneity of multi-source information, with the following detailed technical logic:

First, a domain-specific semantic framework for technological demand is constructed. The framework includes three core semantic dictionaries: technical concept dictionary (collecting 5000+ professional technical terms in the field of technology transfer), demand type dictionary (classifying demands into "product upgrading", "process optimization", "core technology breakthrough", etc.), and constraint condition dictionary (defining 20+ common constraint terms). Each term in the dictionary is assigned a unique semantic ID and hierarchical category (e.g., "photovoltaic energy storage" belongs to "new energy technology" category, semantic ID: T00123).

Second, semantic mapping of multi-source information is implemented. For text-type information, semantic annotation is performed using the CRF (Conditional Random Field) model trained on domain corpus, mapping text terms to the semantic framework; for behavioral information, semantic inference is performed based on the correlation between behavioral states and technical concepts (e.g., "browsing photovoltaic module test cases" is mapped to "photovoltaic energy storage" (T00123)); for attribute information, semantic matching is performed based on the constraint condition dictionary.

Finally, semantic consistency verification is carried out. The cosine similarity between the semantic vectors of different source information is calculated (semantic vectors are derived from the semantic framework). If the similarity is <0.6, the semantic annotation is corrected by referring to the domain thesaurus; if the similarity is <0.4 after correction, the information is marked as "to be verified" and excluded from the current fusion process. Through semantic alignment, the redundancy and ambiguity of multi-source information are eliminated, and the accuracy of information fusion is improved.

4.4.3 Conflict resolution mechanism

Conflict resolution mechanism is used to solve the information conflict existing in multi-source information fusion, with detailed technical implementation as follows:

First, an information credibility evaluation index system is established, including three first-level indicators: source reliability (weight 0.4), logical consistency (weight 0.3), and timeliness (weight 0.3). Source reliability is graded according to the information source type (explicit demand declaration: 0.9, strategic planning: 0.8, platform behavior: 0.7, attribute information: 0.85); logical consistency is evaluated by the semantic similarity between conflicting information and other related information (similarity ≥0.7: 0.9, 0.5-0.7: 0.6, <0.5: 0.3); timeliness is scored based on the information update time (within 6 months: 0.9, 6-12 months: 0.6, more than 12 months: 0.3).

Second, the credibility score of conflicting information is calculated using the weighted sum method. For two conflicting pieces of information, if the credibility score difference is ≥0.2, the information with the higher score is retained; if the score difference is <0.2, a logical reasoning verification is performed using the domain rule base (e.g., "short-term R&D demand" is logically consistent with "small-budget investment", inconsistent with "large-scale industrialization investment").

Finally, if the conflict cannot be resolved through the above steps, the conflicting information is recorded in the "conflict information repository", and the demand reconstruction result is marked with "uncertainty" to remind subsequent users to verify through manual intervention. This mechanism ensures that the fused information is accurate and logically consistent.

4.4.4 Technological demand reconstruction logic

Based on the above multi-source information fusion mechanism, the technical logic of technological demand reconstruction is as follows: first, through layered fusion, the multi-dimensional information is integrated into a systematic feature set; then, through semantic alignment, the consistency of feature semantic connotation is ensured; finally, through conflict resolution, the accuracy of feature information is ensured. On this basis, the systematic integration of core demand features, potential demand features and constraint features is carried out to form a structured and systematic abstraction of enterprise technological demand, that is, the enterprise user portrait oriented to technological demand identification. This portrait realizes the comprehensive and in-depth understanding of enterprise technological demand, and provides a technical basis for the accurate matching of technological supply and demand.

5 CONTRIBUTION AND APPLICATION PROSPECT

5.1 Methodological Contribution

This study has three core methodological contributions: first, it enriches the application of user portrait theory in the specific field of technology transfer. By clarifying the connotation and structural system of enterprise technological demand portraits with operability, it realizes the extension of user portrait theory to the professional field of technology, and expands the methodological boundary of user portrait research.
Second, it develops an integrated multi-source information fusion approach for demand identification, breaking through the methodological limitations of traditional single-dimensional text-driven demand identification. The approach clarifies the technical basis and logical path of multi-source information integration, and enriches the methodological system of technological demand identification.
Third, it explores the multi-source information fusion and technological demand reconstruction mechanism, clarifying the logical relationship between multi-source information and demand identification. The constructed layered fusion, semantic alignment and conflict resolution mechanisms provide a systematic technical method for multi-source information integration in the field of technological demand identification, and improve the methodological depth of demand identification research.

5.2 Practical Application Prospect

From the perspective of practical application, the approach developed in this study has important guiding significance for the practice of technology transfer. First, it can guide technology transfer institutions to establish a scientific demand identification system, avoid the one-sidedness of demand understanding caused by single information sources, and improve the accuracy of technological supply and demand matching.
Second, it can provide technical guidance for enterprises to clarify their own technological demand connotation. By guiding enterprises to systematically sort out multi-dimensional information related to technological demand, they can realize the in-depth understanding of their own demand, which is conducive to improving the efficiency of enterprise technology introduction and R&D.
Third, it can provide a methodological basis for the construction of technology transfer information platforms. The multi-source information source dimension and feature extraction module defined in the approach can guide the design of information collection and processing modules of the platform, and promote the intelligent and efficient development of technology transfer platforms.

5.3 Research Limitations and Future Research Directions

This study is a methodological exploration, and there are certain limitations: first, the integrated approach developed needs to be further verified in practice. Although the approach is based on solid theoretical basis and logical derivation, its practical applicability needs to be tested in specific technology transfer scenarios.
Second, the research on the dynamic adjustment mechanism of portraits is insufficient. Enterprise technological demand is dynamic with the change of internal and external environment, and the current approach does not fully consider the dynamic characteristics of demand. Future research can focus on the following directions: first, carry out empirical research based on the integrated approach, verify and optimize the approach through practical data; second, explore the dynamic adjustment mechanism of enterprise technological demand portraits, and develop a dynamic approach that adapts to the changes of enterprise demand; third, expand the application scope of the approach, and explore its application in different types of enterprises and different technology fields.

6 CONCLUSION

This study develops an integrated multi-source information fusion approach for enterprise user portrait construction oriented to technological demand identification. The study clarifies the connotation and structural system of enterprise technological demand portraits, defines the multi-source information source dimension of portrait construction, constructs a demand feature extraction module based on the dual dimensions of explicit and potential demands, and explores the multi-source information fusion and technological demand reconstruction mechanism. The research breaks through the methodological limitations of traditional single-dimensional text-driven demand identification, enriches the methodological system of user portrait and technological demand identification, and provides technical support for breaking the information asymmetry in technology transfer.
The core conclusion of this study is that the accurate identification of enterprise technological demand can only be realized through the systematic integration of multi-source information. The multi-source information source dimension, demand feature extraction module and information fusion mechanism developed in the integrated approach form a complete technical logic chain, which provides systematic technical guidance for the practice of enterprise technological demand identification. Future research should focus on the empirical verification and dynamic optimization of the approach, so as to further improve the theoretical and practical value of the research.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work is supported by the State Administration for Market Regulation Science and Technology Plan Project, grant number 2023MK195.

REFERENCES

[1] Guo Haiyan. Construction and Empirical Research on Dynamic Optimization Mechanism of Express Cabinet Advertising Based on Terminal User Portraits. Management, 2025, 8(6).
[2] Xu Zhenyun, Hu Zhihua, Guan Yurong. Study on the Construction of User Portrait Model for Accurate Dissemination of Public Welfare Books in China. Journal of International Natural Science Studies, 2024, 1(4).
[3] Tang Xiao, Zhao Jianwen, Han Houkun, et al. Construction and application of power user portrait based on multifeature integration. SPIE, 2024, 13181: 131812G-131812G-10.
[4] Xu Yue, Hu Xuyan, Zhang Jiande. Construction of User Portrait System for Online Shopping of Cultural and Tourism Products Based on Generalized Regression Neural Network. 2024: 507-516.
[5] Zhou Ping, Duan Bowei, Tian Hang, et al. Research on the construction and application of mobile application user data portrait based on improved K-means algorithm. SPIE, 2023, 12941: 1294110-1294110-7.
[6] Leng Hongyong, Shao Jinxin, Zhang Zhe, et al. Research on dual-channel user portrait construction method based on DPCNN-BiGRU and attention mechanism. Journal of Intelligent & Fuzzy Systems, 2023, 45(2): 2579-2591.
[7] Peisen Song, Bengcheng Yu, Chen Chen. Fuzzy Neural Network Algorithm Application in User Behavior Portrait Construction. International Journal of Advanced Computer Science and Applications (IJACSA), 2023, 14(11).
[8] Wei Da, Zhu Sibo, Wang Jian, et al. Construction and application of automobile user portrait based on k-mean clustering model. Applied Mathematics and Nonlinear Sciences, 2022, 8(1): 909-918.
[9] Haowei Huang, Xueyin Liu, Jinyue Cao. Construction of "Internet+" Medical User Portrait. 2021 3rd International Conference on Biology, Chemistry and Medical Engineering, 2021, 15.
[10] Duan Tingting, Qin Ping, Jiang Ping. Study on the Construction of Digital Humanities Group Portraits. 7th Annual International Conference on Social Science and Contemporary Humanity Development, 2021, 610.
[11] Liu Chong, Du Junping. A User Portrait Construction Method for Financial Credit. Springer Singapore, 2020: 29-37.
[12] Technology. HEC Montreal Researchers Add New Study Findings to Research in Technology (An Enriched Customer Journey Map: How to Construct and Visualize a Global Portrait of Both Lived and Perceived Users' Experiences?). Journal of Technology, 2020: 1114.
[13] Ma Xiaodan, Wang Xiaofen. Construction of User Portrait Based on Alipay Big Data. Springer International Publishing, 2020: 467-472.
[14] Zixuan Cheng, Xiangxian Zhang. A novel intelligent construction method of individual portraits for WeChat users for future academic networks. Journal of Ambient Intelligence and Humanized Computing, 2020:1-12.

Eurasia Journal of Science and Technology
Print ISSN: 2663-1024
Online ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3125

MATHEMATICAL MODELING AND HIGH-ORDER FINITE DIFFERENCE METHODS FOR SPATIAL FRACTIONAL FITZHUGH-NAGUMO EQUATIONS

WenYe Jiang, QiZhi Zhang, Yu Li*
Department of Mathematics, College of Science, Northeast Forestry University, Harbin 150040, Heilongjiang, China.
**Corresponding Author: Yu Li*

Abstract: Mathematical modeling of the spatial fractional-order FitzHugh-Nagumo equations provides a critical framework for describing the propagation of electrical potentials in heterogeneous cardiac tissues, a problem that continues to attract significant research attention. Due to the geometrically irregular cross-sections of cardiac tissue, numerical solutions on conventional regular or approximately irregular domains remain limited, and analytical solutions are generally unavailable. In this study, we employ a high-order finite difference method in space coupled with a fourth-order Runge-Kutta scheme in time to solve this nonlinear fractional-order system. Numerical experiments are conducted to validate the high-order convergence and computational stability of the proposed numerical approach.
Keywords: Riesz fractional derivative; FitzHugh-Nagumo model; Finite difference method; The fourth-order explicit Runge-Kutta method

1 INTRODUCTION

Differential equations can describe phenomena ranging from simple linear growth or decay to complex dynamical systems, such as fluid flow, heat conduction, population dynamics, and electrical circuits. As fundamental tools for studying real-world processes, differential equations can be analyzed through analytical methods or numerical computations to reveal the evolution of systems over time[1]. Hodgkin and Huxley proposed a mathematical model describing the transmission of neuronal action potentials, which revealed the mechanism of neural signal conduction. Due to the complexity of the differential equations involved, FitzHugh and Nagumo subsequently proposed a simpler integer-order FitzHugh-Nagumo (FHN) model. Known for its concise mathematical form and accurate depiction of the core dynamics of excitable systems, it has become a fundamental model in neuroscience, cardiac electrophysiology, nonlinear dynamics, and engineering[2].

However, classical integer-order models have limitations in describing the inherent temporal memory dependence and spatial correlation complexity of biological systems. The introduction of fractional-order FHN models effectively addresses this issue, significantly enhancing the biophysical realism and adaptability of the models[3,4]. In recent years, research on fractional-order FHN models has progressed: In 2019, Prakash and Kaur used the homotopy perturbation transform technique to study fractional-order FHN equations for neural impulse transmission, validating the algorithm's high-order convergence and errors[5]; In 2021, Wang et al. used the finite difference method to study the fractional-order FHN monodomain model on moving irregular domains, addressing challenges in cardiac electrophysiological simulation[6]; In 2025, Kumar and Erturk used generalized Caputo fractional derivatives to construct a fractional-order modified FHN neuron model, employed the L1 predictor-corrector method for numerical solutions and conducted error and stability analysis, studying periodic spiking, chaos, and other neuronal electrical activities generated by the model[7].

This paper studies the following one-dimensional spatial fractional FHN equation with Riesz fractional derivative[4]

$$\frac{\partial u}{\partial \mathrm{t}} = -K_u(-\Delta)^{\frac{\alpha}{2}}u - u(1-u)(u-\mu) - v, \tag{1}$$

$$\frac{\partial v}{\partial t} = \varepsilon(\beta u - v), \tag{2}$$

$$u(x,0) = u_0(x), v(x,0) = v_0(x). \tag{3}$$

where $(x,t) \in \Omega \times [0,T]$, $\Omega = (a,b)$, $1 < \alpha \le 2$, ε, μ and β are constants, K_u is the diffusion coefficient. The one-dimensional Riesz fractional derivative is defined as[8]

$$-(-\Delta)^{\frac{\alpha}{2}}u = \frac{\partial^\alpha u}{\partial |x|^\alpha} = -\frac{1}{2\cos(\pi\alpha/2)}\left({}_xD_L^\alpha u + {}_xD_R^\alpha u\right),$$

where ${}_xD_L^\alpha u$ and ${}_xD_R^\alpha u$ are Riemann-Liouvillefractional operators, defined respectively as[8]:

$$ {}_{x}D_{L}^{\alpha}u\left(x,t\right)=\frac{1}{\Gamma(2-\alpha)}\frac{\mathrm{d}^2}{\mathrm{d}x^2}(\int_{a}^{x}(x-\tau)^{1-\alpha}u(\tau,t)\mathrm{d}\tau), $$

$$ {}_{x}D_{R}^{\alpha}u(x,t)=\frac{1}{\Gamma(2-\alpha)}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\int_{x}^{b}(\tau-x)^{1-\alpha}u(\tau,t)\mathrm{d}\tau\right). $$

The finite difference method, due to its high approximation accuracy for fractional derivatives and ease of implementing discrete schemes, has become an important numerical tool for solving spatial fractional differential equations. For the discretization of the Riesz fractional derivative, Huang et al. proposed a fourth-order accuracy finite difference scheme, reducing the truncation error of fractional operator discretization and providing a high-order numerical method for solving high-dimensional fractional reaction-diffusion equations[9]; Han et al. used this method to handle the Riesz fractional diffusion term, successfully solved the fractional Gray-Scott model, and validated its applicability in nonlinear systems[10]. The fourth-order explicit Runge-Kutta method, due to its high temporal accuracy and good stability, is widely used for solving semi-discrete ordinary differential equation systems. In the study of integer-order FHN equations, Chin used this method for temporal discretization, confirming that the fourth-order explicit Runge-Kutta method can accurately capture the system's dynamic behavior and has good stability[11]. This paper will apply the finite difference method combined with the fourth-order explicit Runge-Kutta method to discretize the one-dimensional spatial fractional FHN equation with Riesz fractional derivative and prove the stability, convergence, and error analysis of the numerical method.

The content arrangement of this paper is as follows: Section 1 discretizes the one-dimensional Riesz fractional derivative in the spatial fractional FHN equation using the finite difference method, transforming the original partial differential equation system into a semi-discrete ordinary differential equation system, then performs temporal discretization on the ODE system to obtain a fully discrete numerical scheme, and proves the stability and convergence of the numerical method. Section 2 validates the theoretical analysis results through several numerical experiments; Section 3 provides a brief conclusion.

2 NUMERICAL SCHEME AND ITS PROPERTIES

2.1 Numerical Scheme

2.1.1 Spatial discretization

This subsection uses finite difference approximation for the fractional Laplacian to obtain a semi-discrete ordinary differential system. Let the spatial step size $h=\left(b-a\right)/M$, where is a constant. The spatial grid points are defined as

$$ \Omega_h=\left\{x_i|\ x_i=a+ih, i=0,1,\ldots,M-1\right\}, $$

denote $L^{4+\alpha}(\mathbb{R}):=\{f\in L^{4+\alpha}(\mathbb{R})\,|\,\int_{-\infty}^{\infty}(1+|\xi|)^{4+\alpha}\,|\,\hat{f}(\xi)|\,\mathrm{d}\xi<\infty\}$, $\hat{f}\left(\xi\right)=\int_{-\infty}^{\infty}f(x)e^{i\xi x}\mathrm{d}x$, then the numerical scheme for approximating the Riesz fractional derivative $-(-\Delta)^{\alpha/2}u$ is given by the following lemma.

Lemma 2.1[4,9] Assuming $f(x)\in L^{4+\alpha}(\mathbb{R})$, then

$$ \frac{\partial^{\alpha}f}{\partial|x|^{\alpha}}=-\delta_h^{\alpha}f(x)+O(h^4), $$

where $\delta_h^{\alpha}f(x)=\frac{1}{h^{\alpha}}\sum_{k=-\infty}^{\infty}\omega_k^{\alpha}f(x-kh)$, and the weight coefficients ω_k^{α} are defined as

$$ \omega_k^{\alpha}=\begin{cases}\frac{4}{3}g_k^{\alpha}-\frac{1}{3\cdot 2^{\alpha}}g_{k/2}^{\alpha}, & k\ \ is\ \ even\\ \frac{4}{3}g_k^{\alpha}, & k\ \ is\ \ odd\end{cases}, $$

where $g_k^{\alpha}=\frac{(-1)^k\Gamma(\alpha+1)}{\Gamma(k+1)\Gamma(\alpha-k+1)}$.

Let $u_i=u\left(x_i,t\right)$, $v_i=v\left(x_i,t\right)$, performing finite difference approximation on equations (1)(2), we obtain

$$ \frac{\partial u_i}{\partial \mathrm{t}}=-K_u\delta_h^{\alpha}u_i-u_i(1-u_i)(u_i-\mu)-v_i\ , \tag{4} $$

$$ \frac{\partial v_i}{\partial \mathrm{t}}=\varepsilon(\beta u_i-v_i), \tag{5} $$

where

$$\delta_h^\alpha u_i = \frac{1}{h^\alpha}\sum_{k=0}^{i} g_k^\alpha u_{i-k},\ g_k^\alpha = (-1)^k C(\alpha,k),\ C(\alpha,k) = \alpha(\alpha-1)(\alpha-2)\cdots(\alpha-k+1)/k!.$$

Let $U=(u_0,u_1,\ldots,u_{N-1})^T$, $V=(v_0,v_1,\ldots,v_{N-1})^T$, equations (4)(5) transform into the following semi-discrete ordinary differential equation system

$$\begin{cases} \dfrac{dU}{dt} = -K_u\delta_h^\alpha U - U(1-U)(U-\mu) - V, \\ \dfrac{dV}{dt} = \varepsilon(\beta U - V), \end{cases} \tag{6}$$

further let $Q=[U,V]^T$, $F(Q)=[-K_u\delta_h^\alpha U - U(1-U)(U-\mu)-V, \varepsilon(\beta U - V)]^T$, then we have

$$\frac{dQ}{dt} = F(Q). \tag{7}$$

2.1.2 Temporal Discretization

Next, the fourth-order Runge-Kutta method is used to solve the ordinary differential equation (6). Let the time step size $\tau = T/N$, where N is a constant. Spatial and temporal grid points are formed by $\Omega_h^\tau = \Omega_h \times \Omega_\tau$, where $\Omega_\tau = \{t_n \mid t_n = n\tau, n=0,1,...,N\}$, applying the fourth-order explicit Runge-Kutta method to equation (7) yields:

$$\begin{cases} k_1 = F(Q^n), \\ k_2 = F(Q^n + 2\tau k_1), \\ k_3 = F(Q^n + 2\tau k_2), \\ k_4 = F(Q^n + \tau k_3), \\ Q^{n+1} = Q^n + \tau\Phi(Q^n,\tau), \end{cases} \tag{8}$$

where $Q^0=[U_0,V_0]^T$, $\Phi(Q^n,\tau) = 1/6(k_1+2k_2+2k_3+k_4)$.

2.2 Properties of the Numerical Scheme

2.2.1 Stability

Define $\|\bullet\|_2$ as the l_2 norm, the l_2 norm of Q^n is defined as:

$$\|Q^n\|_2 = \sqrt{\|U^n\|_2^2 + \|V^n\|_2^2} = \sqrt{\sum_{i=0}^{M-1}\sum_{j=0}^{M-1}\left((u_i^n)^2 + (v_i^n)^2\right)}.$$

Theorem 2.2[10,12] If $\Phi(Q,\tau)$ satisfies the Lipschitz condition in Q , then this numerical method is stable, i.e., there exists a positive constant τ 、 h independent of C , such that for any time step $n(0\le n\tau\le T)$, $\|Q^n\|_2 \le C\|Q^0\|_2$.

Proof By the premise of the theorem, $\Phi(Q,\tau)$ satisfies the Lipschitz condition in Q , i.e., there exists a Lipschitz constant $L_F>0$, for any two vectors Q_1、Q_2 , we have

$$\|\Phi(Q_1,\tau)-\Phi(Q_2,\tau)\|_2 \le L_F\| Q_1 - Q_2\|_2.$$

Let $Q_1=Q^n$, $Q_2=0$ (zero vector), then it can be deduced that

$$\|\Phi(Q^n,\tau)\|_2 \le L_F\| Q^n\|_2 + \|\Phi(0,\tau)\|_2,$$

however, since at the initial time Q^0 is obtained by discretizing the given initial conditions $u_0(x)$ 、 $v_0(x)$, its norm $\|Q^0\|$ is bounded. By Lemma 2.1 and its derivation, δ_h^α is a fourth-order accuracy approximation, and the norm of its corresponding matrix l_2 is bounded, i.e., $\|\delta_h^\alpha U\|_2 \le C_\alpha h^{-\alpha}\|U\|_2$, C_α is a constant only related to α . It is concluded that the linear term $F(Q)$ in (containing $\delta_h^\alpha U$) is bounded, and the nonlinear term $-U(1-U)(U-\mu)$ is locally

bounded within a bounded domain, so $\left\|\Phi\left(0,\tau\right)\right\|_2$ can be controlled by a constant independent of τ、h, temporarily denoted here as C_0, i.e., $\left\|\Phi\left(Q^n,\tau\right)\right\|_2 \le L_F\left\|Q^n\right\|_2 + C_0$.

Taking the l_2 norm of the fifth equation in the fully discrete scheme (9), by the triangle inequality

$$\left\|Q^{n+1}\right\|_2 = \left\|Q^n + \tau\Phi\left(Q^n,\tau\right)\right\|_2 \le \left\|Q^n\right\|_2 + \tau\left\|\Phi\left(Q^n,\tau\right)\right\|_2,$$

Substituting the above estimate for $\left\|\Phi\left(Q^n,\tau\right)\right\|_2$, we obtain

$$\left\|Q^{n+1}\right\|_2 \le \left\|Q^n\right\|_2 + \tau(L_F\left\|Q^n\right\|_2 + C_0) = (1+\tau L_F)\left\|Q^n\right\|_2 + \tau C_0,$$

Next, analyze the evolution trend of the norm through iteration: at initial time $\mathrm{n}=0$, $\left\|Q^0\right\|_2$ is bounded (denoted as $C_1 = \left\|Q^0\right\|_2$);

When $n=1$, $\left\|Q^1\right\|_2 \le (1+\tau L_F)C_1 + \tau C_0$;

When $n=2$, $\left\|Q^2\right\|_2 \le (1+\tau L_F)\left\|Q^1\right\|_2 + \tau C_0 \le (1+\tau L_F)^2 C_1 + \tau C_0(1+(1+\tau L_F))$;

And so on, for any n, we have: $\left\|Q^n\right\|_2 \le (1+\tau L_F)^n C_1 + \tau C_0\sum_{k=0}^{n-1}(1+\tau L_F)^k$,

using the geometric series sum formula $\sum_{k=0}^{n-1}(1+\tau L_F)^k = \frac{(1+\tau L_F)^n - 1}{\tau L_F}$, ($L_F \ne 0$, if $L_F = 0$ then $F\left(Q\right)$ is a constant vector, stability obviously holds), substituting yields:

$$\left\|Q^n\right\|_2 \le (1+\tau L_F)^n C_1 + \frac{C_0}{L_F}[(1+\tau L_F)^n - 1].$$

By the properties of the exponential function, for any $x>0$, positive integer m, have $(1+x)^m \le e^{mx}$, here $x=\tau L_F$、$m=n$, and $n\tau \le T$ (final time fixed), so $(1+\tau L_F)^n \le e^{n\tau L_F} \le e^{TL_F}$ (let $C_2 = e^{TL_F}$, independent of τ、h).

Since $C_2 = e^{TL_F}$ is always constant, and $\frac{C_0}{L_F}[(1+\tau L_F)^n - 1] \le \frac{C_0}{L_F}(C_2 - 1)$ (denoted as C_3, independent of τ、h).

Therefore: $\left\|Q^n\right\|_2 \le C_2 C_1 + C_3$.

Let $C = \max\{C_2C_1 + C_3, C_1\}$ (independent of τ、h), then for any n exist $\left\|Q^n\right\|_2 \le C$, and since $C_1 = \left\|Q^0\right\|_2$, C can be further adjusted such that $\left\|Q^n\right\|_2 \le C\left\|Q^0\right\|_2$ (if $\left\|Q^0\right\|_2 = 0$ then the numerical solution is always zero, if $\left\|Q^0\right\|_2 > 0$, then C can be taken as $\max\left\{C_2 + C_3/C_1, 1\right\}$), i.e., the l_2 norm of the numerical solution vector remains bounded during the evolution process, hence this numerical method is stable.

2.2.2 Convergence and error analysis

Theorem 2.3 Let the global error be $e_n = Q(t_n) - Q^n$, combining Lemma 2.1 and equation (8) obtained by the fourth-order Runge-Kutta method, if is satisfied $\tau/h^\alpha \le 1$, then there exist positive constants C_4, C_5, (independent of τ and h), such that $\left\|e_n\right\| \le C_4h^4 + C_5\tau^4$.

Proof Let the projection vector of the exact solution $u\left(x,t\right)$、$v\left(x,t\right)$ at grid point Ω_h be $Q(t) = [U(t)^T, V(t)^T]^T$, where $U(t), V(t)$ are composed of $u\left(x_i,t\right)$ and $\mathrm{v}\left(x_i,t\right)$ respectively.

By Lemma 2.1, the spatial discretization truncation error is $O(h^4)$, i.e.

$$\left\|-\delta_h^\alpha u(x) - \frac{\partial^\alpha u}{\partial|x|^\alpha}\right\| \le K_1h^4,$$

where K_1 is a constant, $u \in L^{4+\alpha}(\mathbb{R})$. The exact solution $Q_h\left(t\right)$ of the semi-discrete system (8) satisfies

$$\| Q(t) - Q_h(t) \| \le C_s h^4, \forall t \in [0, T],$$

where $C_s > 0$ is a constant independent of h, t.

The local truncation error of the fourth-order Runge-Kutta method is $O(\tau^5)$, i.e.

$$\| Q_h(t_{n+1}) - Q_h(t_n) - \tau\Phi(Q_h(t_n), \tau) \| \le K_2 \tau^5,$$

where $K_2 > 0$ is a constant independent of τ. Under the stability condition $\tau / h^{\alpha} \le 1$, by the Lipschitz condition we obtain $\| Q_h(t_n) - Q^n \| \le C_t \tau^4, \forall t_n \le T$, where $C_t > 0$ is a constant independent of τ, n.

The global error is decomposed as: $e_n = Q(t_n) - Q^n = \left[Q(t_n) - Q_h(t_n) \right] + \left[Q_h(t_n) - Q^n \right]$,

applying the triangle inequality: $\| e_n \| \le \| Q(t_n) - Q_h(t_n) \| + \| Q_h(t_n) - Q^n \| \le C_s h^4 + C_t \tau^4$, taking $C_4 = C_s$, $C_5 = C_t$, yields the conclusion.

3 NUMERICAL EXPERIMENTS

3.1 Experiment 1

Consider the following one-dimensional FHN equation[13,14]

$$\frac{\partial u}{\partial t} = u_{xx} + u(1-u)(u-\mu),$$

whose exact solution is $u(x,t) = \frac{1}{2} + \frac{1}{2}\tanh[k(x - ct)]$, where $k = \frac{1}{2\sqrt{2}}$, $c = \frac{2^{\mu} - 1}{\sqrt{2}}$. The initial condition is $u(x,0) = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{x}{2\sqrt{2}} \right)$.

Define the errors E_2 and E_{∞} as

$$E_2 = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left[u(x_j, t) - u^*(x_j, t) \right]^2}, \quad E_{\infty} = \max_{1 \le j \le N} \left| u(x_j, t) - u^*(x_j, t) \right|,$$

where $u(x_j, t)$ and $u^*(x_j, t)$ are the numerical and exact solutions respectively.

In Table 1, taking $\mu = 0.75$, $h = 0.1$, for different values of τ, calculate the E_2 and E_{∞} errors from time 0.01 to 50 respectively. From Table 1, it can be seen that the errors decrease over time and with decreasing time step size. In Table 2, for different values of h, taking $\mu = 0.75$, $\tau = 10^{-5}$, calculate the E_2 and E_{∞} error norms and from time 0.01 to 50 respectively. From Table 2, it can be seen that the error norms decrease over time and with decreasing spatial step size. The solution obtained by this method is shown in Figure 1, where $h = 0.1$, $\tau = 5 \times 10^{-3}$, μ are taken as 0.25, 0.50, 0.75, 0.10 respectively.

Table 1 Error Norms under Different for τ, $h = 0.1$, $\mu = 0.75$

t	$\tau = 10^{-3}$		$\tau = 10^{-4}$		$\tau = 10^{-5}$	
	E_2	E_{∞}	E_2	E_{∞}	E_2	E_{∞}
0.01	2.55e-05	5.53e-05	2.56e-06	5.53e-06	3.31e-07	5.53e-07
1	4.21e-05	5.93e-05	3.52e-06	5.93e-06	3.72e-06	4.26e-06
10	1.67e-05	3.06e-05	1.70e-06	3.06e-06	4.24e-07	4.85e-07
50	1.23e-09	1.89e-09	4.23e-10	3.93e-10	3.58e-10	3.45e-10

Table 2 Error Norms under Different for h, $\tau = 10^{-5}$, $\mu = 0.75$

t	$h = 0.1$		$h = 0.05$		$h = 0.01$	
	E_2	E_{∞}	E_2	E_{∞}	E_2	E_{∞}
0.01	3.31e-07	5.53e-07	2.00e-07	5.53e-07	1.25e-07	5.53e-07
1	3.72e-06	4.26e-06	7.65e-07	8.73e-07	3.60e-07	5.93e-07

10	4.24e-07	4.85e-07	1.79e-07	3.06e-07	1.40e-07	3.06e-07
50	3.58e-10	3.45e-10	9.46e-11	9.03e-11	1.10e-11	1.89e-11

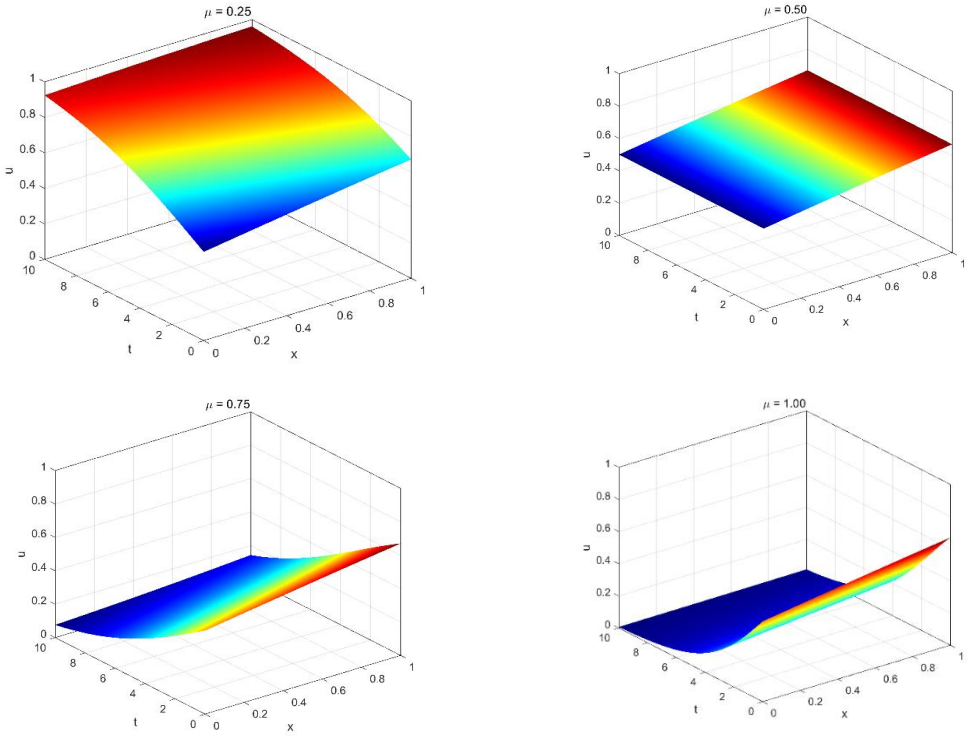


Figure 1 Numerical Solution under Different for μ, $h = 0.1$, $\tau = 5\times10^{-3}$

4 CONCLUSION

This paper proposes a numerical solution scheme combining high-order finite difference method and fourth-order explicit Runge-Kutta method for the one-dimensional spatial fractional FitzHugh-Nagumo model with Riesz fractional derivative. The paper provides rigorous theoretical analysis of the stability and convergence of the numerical method, proving that under certain conditions the scheme is stable and has fourth-order accuracy in both time and space.

Numerical experimental results show that this method exhibits good numerical performance under different fractional orders, time step sizes, and spatial step sizes. Errors decrease significantly with decreasing step sizes, validating the correctness of the theoretical analysis. The intuitive display of the evolution behavior of numerical solutions under different parameters further confirms the effectiveness and reliability of this method in simulating the dynamic processes of fractional FHN systems. This work also serves as an illustrative case in the curriculum reform project : Restructuring and Practical Research on an Interdisciplinary PBL Course System for Mathematical Modeling Driven by Generative AI, demonstrating how advanced computational methods can be integrated into modeling education.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

The research received funding by Undergraduate Education and Teaching Research Project of Northeast Forestry University: Research on the Reconstruction and Practice of an Interdisciplinary PBL Curriculum System for "Mathematical Modeling" Driven by Generative AI (Project No.: DGY2025-43).

REFERENCES

[1] Han C, Wang Y L, Li Z Y. Novel patterns in a class of fractional reaction–diffusion models with the Riesz fractional derivative. Mathematics and Computers in Simulation, 2022, 202, 149–163.

[2] Zhang L, Sun H W. Numerical solution for multi-dimensional Riesz fractional nonlinear reaction–diffusion equation by exponential Runge–Kutta method. Journal of Applied Mathematics and Computing, 2019, 60 (1-2): 357-378.

[3] Zeng F, Liu F, Li C P, et al. A Crank--Nicolson ADI spectral method for a two-dimensional Riesz space fractional nonlinear reaction-diffusion equation. SIAM Journal on Numerical Analysis, 2014, 52(6): 2599-2622.

[4] Li X, Han C, Wang Y. Novel Patterns in Fractional-in-Space Nonlinear Coupled FitzHugh-Nagumo Models with Riesz Fractional Derivative. Fractal and Fractional, 2022, 6(3), 136.
[5] Prakash A, Kaur H. A reliable numerical algorithm for a fractional model of Fitzhugh-Nagumo equation arising in the transmission of nerve impulses. Nonlinear Engineering, 2019, 8(1):719-727.
[6] Wang Y H, Feng X L, Li H. Ghost structure finite difference method for fractional FitzHugh-Nagumo monodomain model on moving irregular domains. Journal of Computational Physics, 2021, 424: 110862.
[7] Kumar P, Erturk V S. A fractional-order improved FitzHugh-Nagumo neuron model. Chinese Physics B, 2025, 34(1): 018704.
[8] Lee H G. A second-order operator splitting Fourier spectral method for fractional-in-space reaction–diffusion equations. Journal of Computational and Applied Mathematics, 2018, 333: 395-403.
[9] Huang Y Y, Qu W, Lei S L. On τ-preconditioner for a novel fourth-order difference scheme of two-dimensional Riesz space-fractional diffusion equations. Computers & Mathematics with Applications, 2023, 145: 124-140.
[10] Han C, Wang Y L, Li Z Y. A High-Precision Numerical Approach to Solving Space Fractional Gray-Scott Model. Applied Mathematics Letters, 2022, 125: 107759.
[11] Chin P W M. The effect on the solution of the FitzHugh-Nagumo equation by the external parameter using the Galerkin method. Journal of Applied Analysis & Computation, 2023, 13(4): 1983-2005.
[12] Han C, Wang Y L, Li Z Y. Numerical solutions of space fractional variable-coefficient KdV–modified KdV equation by Fourier spectral method. Fractals, 2021, 29(08): 2150246.
[13] Zhang J J, Chen Y M, Wang X M. High-precision compact difference method for the FitzHugh-Nagumo equation. Journal of China West Normal University (Natural Science Edition), 2021, 42(2): 126–132.
[14] Luo Y, Song L Y. Solve the generalized Fitzhugh-Nagumo equation with time-varying coefficients based on the Chebyshev spectral method. Journal of Jiamusi University (Natural Science Edition), 2025, 43(3): 166–170.

Eurasia Journal of Science and Technology
Print ISSN: 2663-1024
Online ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3124

A CONTROL-DRIVEN DATA ASSET CLASSIFICATION METHOD FROM A RIGHTS PERSPECTIVE

DaXing Chen, Kun Meng*, YuChen Zhao, QiYuan Wang
Beijing Information Science and Technology University, Beijing 100192, China.
**Corresponding Author: Kun Meng*

Abstract: Against the backdrop of the digital economy, the process of data assetization is accelerating. However, in operational links such as registration, utilization, and transaction, institutional data rights have failed to be translated into enforceable implementation mechanisms, leading to ambiguous rights, responsibilities, and inefficient circulation. From a rights management perspective, this paper proposes a tripartite classification framework for data assets: identification data resolves issues of attribution, raw material data regulates usage based on clarified ownership, and tool data further enables the distribution of benefits. This framework integrates the national policy of "Tripartite Rights Separation", analyzes the distinct rights characteristics of different data types, and designs differentiated governance pathways. It aims to provide institutional support for data assetization operations and foster the efficient functioning of the data factor market.
Keywords: Data assetization; Data classification; Tripartite rights separation; Differentiated governance

1 INTRODUCTION

With the rapid development of the global digital economy, data is transitioning from a traditional auxiliary resource into a core asset with quantifiable and tradable characteristics [1]. This transformation not only promotes industrial upgrading and business model innovation but also brings entirely new challenges for data governance. The core challenge in data asset governance lies in achieving efficient utilization and value release of data resources on the premise of ensuring data security and privacy [2], and the issue of data rights attribution is the root of this challenge [3]. For data assets, given that related subjects exist, and these subjects may not be a single entity but rather diverse, and are often separated from the producing entity and the value carrier. It is therefore difficult to clarify which subject the data belongs to, and it is also difficult to represent the rights relationships that continuously change during the circulation process. This inevitably leads to disputes over related rights and interests.
In view of this situation, the Opinions of the CPC Central Committee and the State Council on Building a Basic Data System to Better Harness the Role of Data as a Production Factor (hereinafter referred to as the “Data Twenty Measures”) innovatively proposes an institutional framework for the separation of data property rights. This framework establishes a tripartite structure consisting of data resource holding rights, data processing and usage rights, and data product operation rights. Building on this logic, this paper develops a differentiated governance framework by categorizing data assets into three groups—identifier-type, material-type, and tool-type. The aim is to formulate management strategies tailored to the specific characteristics of different data assets in ownership determination, usage authorization, and revenue allocation, thereby effectively addressing the current issues of unclear rights and responsibilities and inefficient circulation in data-asset operations.

2 OPERATION OF DATA ASSETIZATION

2.1 Overview on Data Assetization

"Assets" is an accounting concept [4], referring to resources arising from an enterprise's past transactions or events, owned or controlled by the enterprise, and expected to yield economic benefits. In contrast, resources place greater emphasis on the potential input of production factors and generally lack defined asset attributes in economic accounting. Asset attributes require clear rights and control over the resource. Data, as a form of electronically recorded information, possesses heterogeneity and value appreciation, making it an important production factor in modern economic systems. Unlike traditional resources, data exhibits non-exclusivity and replicability [5]; the same dataset can be simultaneously utilized by multiple entities, and its value is continuously enhanced through repeated use and in-depth analysis. These characteristics endow data with significant potential commercial value and provide impetus for its development toward assetization.
It is generally acknowledged that data resources represent the collective term for data with value potential, whereas data assets refer to data resources legally owned or controlled by a specific entity, measurable in monetary terms and capable of generating economic benefits. The transition of data from a potential resource to an explicit carrier of value constitutes a trend [6]. This shift not only signifies the realization of data's value but also reflects enterprises' strategic objective of achieving full lifecycle management and value appreciation through assetization operations. The core of this transformation lies in establishing a supportive ecosystem for data value discovery, rights confirmation, dynamic

governance, and efficient circulation, thereby endowing data with asset attributes and enabling systematic management and commercial transactions.

2.2 Implementation Path of Data Assetization

Data assetization must fulfill three key conditions: clear ownership, measurability, and tradability. Its implementation pathway can be summarized into three stages: data governance, data packaging, and data transaction (as shown in Figure 1). This pathway exhibits a progressive relationship: data governance serves as the foundation, achieving a usable and trustworthy state for data resources through systematic management; data packaging is the critical transformation stage, endowing data with legal and economic asset attributes through rights confirmation, valuation, and registration; data transaction is the value realization stage, enabling the market circulation and monetization of data assets based on established rules and platforms. These three stages advance progressively, collectively forming a complete closed loop from data resources to assets, ultimately realizing market value.

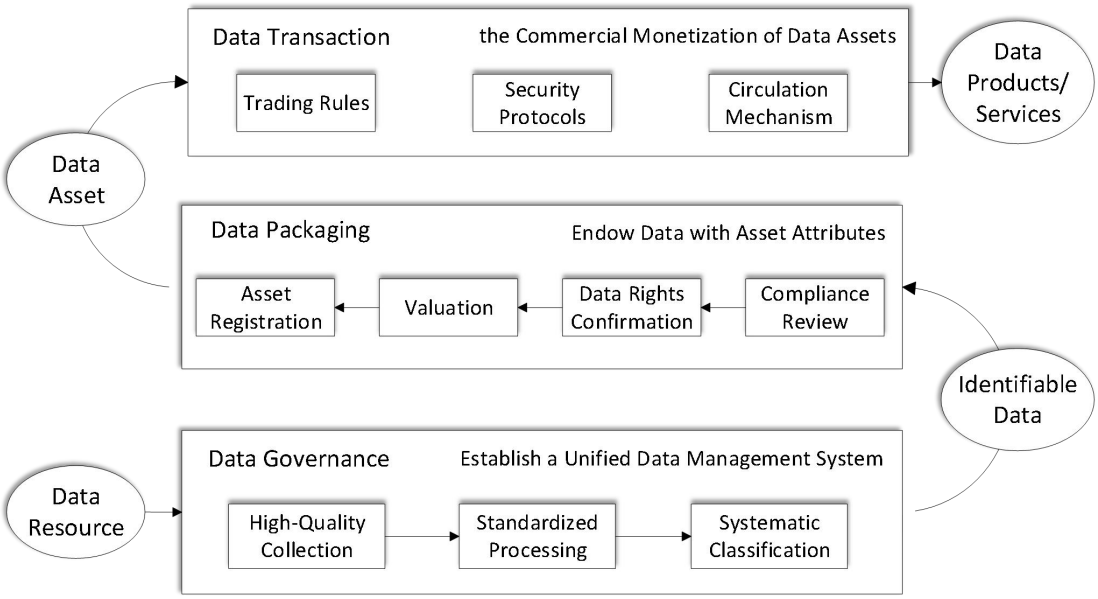


Figure 1 Schematic Diagram of the Data Assetization Implementation Path

2.3 Challenges in the Operationalization of Data Assetization

As stated in the Introduction, the core challenges in the operationalization of data assetization lie in the unclear definition of rights and inefficient circulation mechanisms. Specifically, at the operational level, these are mainly reflected in the following two issues:

Lack of actionable implementation standards for data rights confirmation: There is currently a lack of unified laws and regulations clarifying the nature, subjects, and content of data rights, leading to an overreliance on a singular ownership concept in practice, which fails to accommodate the characteristics of multi-party contribution and dynamic circulation of data.

Lack of differentiated rule design for data transactions: The failure to establish tiered and classified transaction standards and contractual frameworks based on the different rights attributes and application scenarios of data constrains circulation efficiency.

3 DATA CLASSIFICATION FROM A RIGHTS PERSPECTIVE

3.1 Basis for Data Asset Classification

Due to differences in data classification methods, data assets can be categorized differently [7]. They can be distinguished by rights attribution into public data, enterprise data, and personal data; by source into internal data and external data; and by purpose into data for self-use, commercialized use, and public welfare use. These classifications each have their own focus and are conducive to establishing differentiated standards at a macro level. However, they also possess significant limitations: Attribution-oriented classifications emphasize the subject scope but often fail to trace the chain of contributors, making it difficult to reflect the multi-party contributions and rights distribution in the process of data generation and circulation. Source or purpose-oriented classifications help identify the data origin and usage scenarios but are inadequate in providing granular governance rules for the rights and usage constraints involved in complex circulation scenarios. Therefore, relying solely on single-dimensional, generic classifications can hardly support the governance requirements of data assetization. To enhance the suitability of control measures, rules should

be further refined under the overall classification framework, using rights requirements as the starting point to formulate differentiated controls, thereby addressing the challenges in the operationalization of data assetization.

3.2 The Tripartite Rights Separation in Data Assets

Data property rights constitute the cornerstone of the fundamental data institutional system. Regulating data rights and interests based solely on “ownership” is not conducive to realizing the value of data assets or facilitating their circulation. The future returns from data assets depend not only on “who owns them” but, more importantly, on “who exploits and utilizes them.” Based on the data property rights separation framework outlined in the aforementioned “Data Twenty Articles,” this paper operationalizes the relevant rights as follows:

Right to Hold: This requires that the data subject matter clearly and uncontroversially identifies all contributor information and the authorized entity for that data. Contributors include the subject reflected in the data content, the collector or generator, and participants in its storage, transmission, processing, etc. The authorized entity for the data must comply with legal and regulatory requirements and may be a single entity, a collective, or a specific mechanism.

Right to Use: This is the right to perform operations on the data subject matter in accordance with the data authorization stipulations. These stipulations include viewing, copying, modifying, publishing, etc., and primarily pertain to operations on the data content itself.

Right to Operate: This is the right, given the necessary resource support, to generate greater value and obtain economic benefits from the data subject matter by providing services related to it. Services involve establishing service relationships and delivering service content, which includes providing storage, transmission, transaction, operation, deployment, etc. Economic benefits primarily stem from the value of the services provided and from any new valuable outcomes subsequently generated through these services. Confirmation of the service relationship typically involves the authorized entity of the data.

3.3 Data Asset Classification Based on the "Tripartite Separation of Rights"

Building upon the rights-related logic of the "Tripartite Separation of Rights," and centering on the perspective of the data holder, data assets are categorized into three types according to their control objectives: Identification, Material, and Tool categories (as shown in Figure 2). This classification adopts "Ownership Confirmation—Usage Authorization—Value Operation" as its governance thread, aiming to translate institutional rights and functionalities into executable control rules.

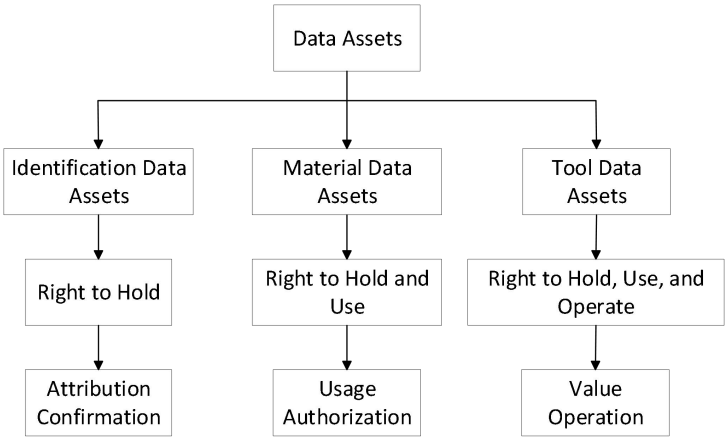


Figure 2 Data Asset Classification and Control Objectives

3.3.1 Identification data assets

Identification Data Assets primarily aim for open sharing and transparent attribution. Typical examples include government statistics and industry benchmark data. The emphasis for this category lies in ensuring reliable sources and accurate contributor identification. They are usually open for public use and do not serve as direct commercial products. Consequently, the control focus is on the clarity of the Right to Hold and the identification mechanism, with no restrictions required on the Right to Use and the Right to Operate (as shown in Table 1).

Table 1 Rights Control for Identification Data Assets

Subject Rights	Right to Hold	Right to Use	Right to Operate
Data Provider	Yes	No	No
Data Requester	No	No	No

3.3.2 Material data assets

Material Data Assets have "Usage Authorization" as their core objective. Serving as the raw material of data, such data assets include internally collected data and data obtained through legal channels. They function as fundamental inputs

within enterprises, with their commercial value often realized in subsequent deep processing and application. Control for this category requires the provider to ensure the Right to Hold while regulating "who can use" and "how to use" through explicit access authorization (as shown in Table 2). The Right to Operate, however, is not open to external parties; meaning the data itself is not independently commercially operated or involved in profit distribution, and does not directly participate in transactions as a commodity. The control emphasis for Material Data Assets lies in formulating detailed usage authorization rules and ensuring that the data, as "raw material," is applied within the defined scope after authorization, thereby enabling the secure transformation and reasonable sharing of data value.

Table 2 Rights Control for Material Data Assets

Subject Rights	Right to Hold	Right to Use	Right to Operate
Data Provider	Yes	Yes	No
Data Requester	No	Yes	No

3.3.3 Tool data assets

Tool Data Assets have "Value Operation" as their core objective. Their control objective is the right to obtain benefits based on attribution confirmation and usage authorization. These benefits primarily stem from two aspects: firstly, the Tool Data itself (such as algorithms, models, software, etc.) possesses direct service value; secondly, the derived data generated by processing, integrating, and analyzing data through these tools achieves "secondary value appreciation" in the market by providing data services, reflecting the additional commercial value of the new outcomes.

Table 3 Rights Control for Tool Data Assets

Subject Rights	Right to Hold	Right to Use	Right to Operate
Data Provider	Yes	Yes	Yes
Data Participant	No	Yes	Yes

Compared to the previous two categories of data assets, the governance of Tool Data Assets is more complex (as shown in Table 3), as it requires the simultaneous clarification of rights attribution and operational rules for both the tool itself and the derivative data generated from it. The core logic of this rights allocation scheme is as follows: The Data Provider, because they bear the development and maintenance costs of the original data, core tools, or platforms, and are the bearers of initial risk, therefore retains the Right to Hold over the original data and core tools and naturally enjoys the Right to Use them; simultaneously, as the initiator of the value ecosystem, they also retain the Right to Operate over the overall assets. Data Participants (including technical processors, service applicators, etc.), because they directly participate in the value-added process of the data and the creation of derivative value, are consequently granted the Right to Use to perform data processing and analysis; and based on their specific contributions in service provision, technical investment, and outcome production, they flexibly share the Right to Operate and its benefits through contracts or quantitative metrics, thereby achieving fair value return and incentive.

4 DATA ASSET CONTROL MODELS BASED ON CLASSIFICATION

The classified control of data assets is a critical pathway to unleashing the value of factors. Building upon the data asset classification framework and its corresponding rights configuration established in Chapter 2, this chapter will delve into the technological implementation paradigms for classified control. By analyzing the rights confirmation elements and technical implementation paths for each category of data, it aims to provide a verifiable institutional-technical dual-track solution for the rights confirmation and circulation of data assets.

4.1 Analysis of Data Rights Confirmation Elements

Data rights confirmation aims to address three fundamental issues: first, the nature of data rights, i.e., what type of rights protection should be granted to data; second, the subject of data rights, i.e., clarifying who should enjoy data rights; and third, the content of data rights, i.e., specifically defining what rights the data subject enjoys. To make the process of changes in data rights presentable and analyzable in a legal sense, it is necessary to adopt an internal perspective of rights analysis and complete the extraction of fundamental rights confirmation elements (as shown in Figure 3). In the process of building a data rights confirmation system, the bottom-up logic of "object → subject → behavior" is followed: first, the data "object" must be defined and identified to provide an anchor point for rights; on the basis of a clearly defined object, the identity and rights boundaries of the "subject" become meaningful; and only when both the subject and object are clear can the rules constraining the operational "behavior" of the subject be effectively defined and executed.

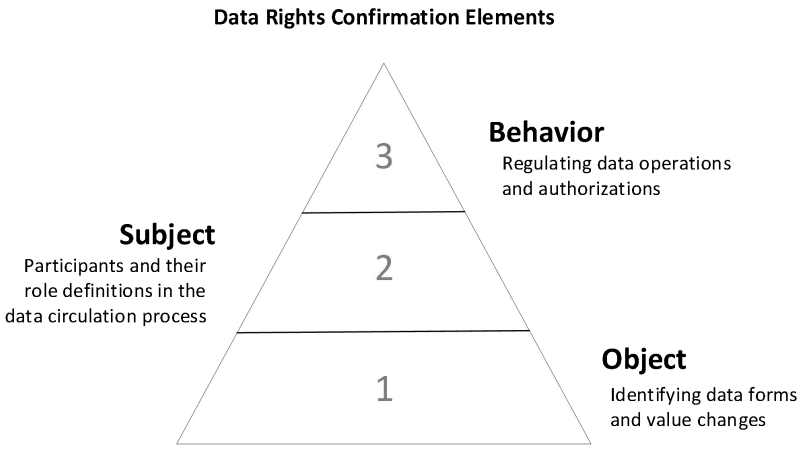


Figure 3 Schematic Diagram of Data Rights Confirmation Elements

4.2 Implementation Path for Data Rights Confirmation Based on the Classification Architecture

To translate the rights confirmation logic of "Object-Subject-Behavior" into an operational system, we have constructed a data asset circulation architecture (as shown in Figure 4). This architecture uses the blockchain network and cloud storage system as its underlying infrastructure [8], and through core components such as the identity authentication module, access control module, and value distribution module, provides differentiated technical implementation paths for the three categories of data.

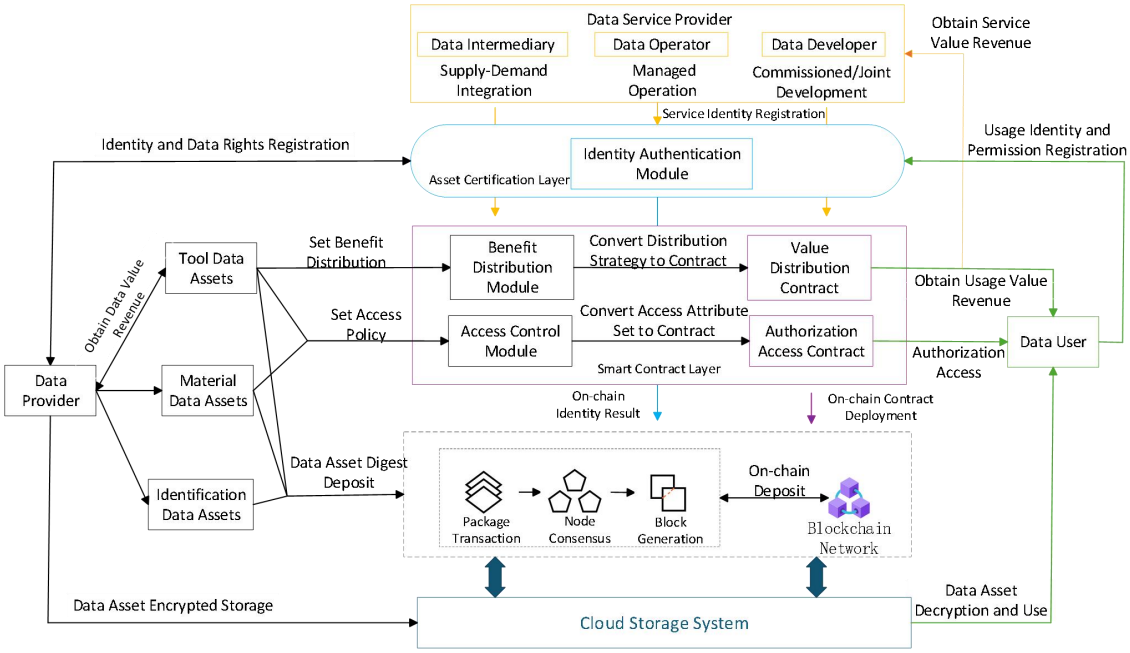


Figure 4 Architecture Diagram of Data Asset Circulation Based on This Classification

For Identification Data Assets, the rights confirmation path focuses on "Object" authentication and "Subject" binding to establish a credible ownership foundation. By depositing the data hash value as evidence on the blockchain network, the unique identification of the data "Object" is completed; simultaneously, by relying on the identity authentication module, the reliable association between the data provider's "Subject" identity and the data object is achieved. The technology combination centered on blockchain deposit and digital identity, through the clear labeling of contributor information and authorized subjects, provides a credible rights credential for the open sharing of data in a lightweight manner.

Building upon the rights confirmation of Identification Data Assets, the control of Material Data Assets extends further to the fine-grained control of usage "Behavior". Its implementation path, while ensuring "Object" confidentiality (through data encryption storage) and clarifying "Subject" permissions, transforms institutionalized usage rules into executable code in smart contracts. When a data user initiates an access behavior, the system automatically verifies their permissions and controls the data decryption scope through the contract, thereby achieving real-time authorization and post-event auditing for each usage behavior.

The rights confirmation for Tool Data Assets comprehensively covers the three elements, focusing on resolving the "Behavior" quantification and value distribution issues in multi-party collaboration. A version traceability system for data objects is constructed through trusted timestamps, multiple subject identities are confirmed relying on the identity

registration layer, and finally, the contribution behaviors of all parties are converted into automatically executable economic benefits through the value distribution contract. On the basis of credible objects and clear subjects, a sustainable value creation and distribution system is built through the precise quantification of operational behaviors.

Table 4 Analysis of Data Rights Confirmation Elements and Implementation Paths

Data Category	Core Rights	Rights Confirmation Elements	Implementation Path
Identification	Right to Hold	Object, Subject	Object hash deposit; Subject identity binding
Material	Right to Use	Object, Subject, Behavior	Object encryption; Subject authorization; Behavior controlled via smart contracts
Tool	Right to Operate	Object, Subject, Behavior	Object version tracing; Multi-subject registration; Behavior contribution quantification and profit-sharing

The rights confirmation solutions for the three data categories constitute a layered and progressively advancing technical system (as shown in Table 4): the Identification category employs low-overhead deposit, focusing on the object-subject attribution confirmation; the Material category relies on medium-overhead data encryption and policy execution, requiring the addition of constraints and auditability for usage behaviors on the basis of object-subject; the Tool category necessitates a high-overhead on-chain accounting system, constructing a verifiable value closed-loop supporting the Right to Operate distribution while comprehensively covering the three elements. This differentiated technical adaptation confirms the "scenario-specific rights and capabilities" control paradigm, systematically building a progressive control chain from attribution confirmation, through behavior constraints, to value distribution, providing a solution for the efficient circulation of the data factor market.

5 CONCLUSION

This research addresses the issues of rights attribution and circulation efficiency in the process of data assetization by innovatively constructing a classified control framework based on the "Tripartite Separation of Rights". By analyzing the rights and functionalities of the data assets' Right to Hold, Right to Use, and Right to Operate, a three-category system was established. Based on the analysis of rights confirmation elements, differentiated technical implementation paths were designed for each data category. The Identification category ensures open sharing through rights transparency, the Material category achieves secure circulation through dynamic constraints on the Right to Use, and the Tool category activates multi-party collaboration through revenue contribution quantification. This paper provides an institutional foundation for data registration, transaction, and compliance operations. In the future, better implementation can be achieved through iterations in technologies such as privacy-preserving computation and cryptography, and the application of this classification standard in more complex data circulation scenarios can be explored.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] International D. DAMA-DMBOK: Data management body of knowledge. Technics Publications, LLC, 2017.
[2] Hu C, Li Y, Zheng X. Data assets, information uses, and operational efficiency. Applied Economics, 2022, 54(60): 6887-6900.
[3] Wu T, Song X. Data Asset Analysis, Financial Reporting, and Recommendations. Academic Journal of Business & Management, 2025, 7(3): 32-36.
[4] Li Y. Concepts, Accounting Treatment and Pricing of Data Assets. Journal of Economic Insights, 2024, 1(1).
[5] Fan J I N. Research on Integrating Data Assets Into Accounting Discipline System. Journal of Modern Accounting and Auditing, 2023, 19(4): 101-112.
[6] Xu T, Shi H, Shi Y, et al. From data to data asset: conceptual evolution and strategic imperatives in the digital economy era. Asia Pacific Journal of Innovation and Entrepreneurship, 2024, 18(1): 2-20.
[7] Newhouse W, Souppaya M, Kent J, et al. Data Classification Concepts and Considerations for Improving Data Protection. National Institute of Standards and Technology, 2023.
[8] Zhang T, Jiang J, Wen B, et al. A Dual-certification Integrated Data Property Registration and Application Innovation Paradigm//2024 2nd International Conference on Management Innovation and Economy Development (MIED 2024). Atlantis Press, 2024: 256-263.

Eurasia Journal of Science and Technology
Print ISSN: 2663-1024
Online ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3126

LIFE ATTITUDE PROFILES AMONG CHINESE UNIVERSITY STUDENTS: A SURVEY STUDY OF POSITIVE, LYING-FLAT, BUDDHA-LIKE, AND NEGATIVE MINDSETS

Liu Yang
Organization and Publicity Department, Guangdong Agriculture Industry Business Polytechnic, Guangzhou 510507, Guangdong, China.

Abstract: Introduction: Amid the massification of higher education and rapid social transformation in China, university students face intensified academic competition, uncertain mobility prospects, and heightened psychosocial stress. These conditions make "attitudes toward life" a consequential construct for understanding how students evaluate life circumstances and translate such evaluations into coping and action tendencies. However, existing studies often examine positive functioning and distress-related outcomes separately, with limited integration of co-existing orientations such as engagement, withdrawal, and goal attenuation. Objectives: This study aims to develop a context-sensitive Life Attitude Scale (LAS) and to map the overall pattern and subgroup differences of college students' life attitudes across four co-existing orientations: positive mindset, lying-flat mindset, Buddha-like mindset, and negative mindset. Methods: A cross-sectional, on-site paper-and-pencil survey was administered to students from nine colleges at a university in Guangzhou, yielding 287 valid responses (effective rate 89.69%). The LAS comprised 19 items across four subscales (5, 5, 5, and 4 items, respectively) rated on a 5-point Likert-type scale; survey data were analyzed in SPSS 27.0 with psychometric screening and nonparametric group comparisons, supplemented by interview materials for contextual interpretation. Results: Descriptive results showed that the positive mindset was highest (M = 4.093), while the lying-flat mindset (M = 2.419) and the negative mindset (M = 2.190) remained comparatively low; the Buddha-like mindset was near the midpoint (M = 2.987). Kruskal-Wallis tests indicated statistically significant differences across selected background variables, including discipline-related variation in positive mindset, as well as gender- and family-background-related variation in low-desire orientations and negative mindset. Discussion: The findings depict an overall positive life orientation among contemporary Chinese university students, alongside meaningful heterogeneity in withdrawal-style and low-desire coping patterns. The four-dimensional LAS provides a measurement basis for differentiated educational guidance and mental health promotion. Future research should extend validation to multi-site samples and longitudinal designs, and further test mechanism pathways linking structural stressors, relative deprivation, and life-attitude configurations.
Keywords: University students; Life attitudes; Current situation analysis; Cognitive differences

1 INTRODUCTION

Amid the massification of higher education and rapid social transformation in China, college students' psychological adaptation and value orientations have increasingly become public concerns. As a transitional cohort moving from "learners" to future workers and citizens, students must navigate layered demands involving academic competition, career planning, and social comparison. Their overarching evaluation of life circumstances and prospects, and the action tendencies derived from such evaluations, can be meaningfully captured by the construct of attitudes toward life. Recent synthesis evidence suggests that mental health burdens among Chinese university students remain substantial, underscoring the need to interpret students' attitudes and choices through an experience-near psychosocial lens [1,2].

Psychologically, attitudes toward life can be conceptualized as relatively stable evaluative orientations toward one's life conditions, perceived future possibilities, and the perceived value of action. Such orientations not only affect experiences but also concrete behaviors via goal setting, effort investment, and coping strategies. In this sense, life attitudes encompass judgments about whether life feels meaningful and assessments of whether continued investment is worthwhile. Empirically, variations in meaning-related orientations among Chinese college students have shown systematic links with life satisfaction as well as depression and anxiety [3].

Importantly, contemporary youths' life attitudes cannot be reduced to a simple positive–negative dichotomy; rather, they have become increasingly differentiated amid structural pressures and shifting opportunity structures. On the one hand, congested graduate labor markets and uncertain mobility prospects heighten sensitivity to the perceived return on effort. Research indicates that experiences such as personal relative deprivation can suppress self-improvement motivation, fostering withdrawal from competition and sustained effort [4]. On the other hand, vernacular repertoires such as tang ping ("lying flat"), bai lan ("let it rot"), and a "Buddha-like" stance of goal attenuation can function as interpretive frames and coping styles under high pressure and uncertainty. In Chinese college student samples, such attitudes have been incorporated into empirical models and linked with behavioral outcomes [5–8].

Prior research has substantially advanced our understanding of mental health and stress mechanisms among Chinese

university students. However, at the level of life attitudes, which is closer to young adults' everyday narratives and consequential choices, two gaps remain salient. First, many studies focus on single symptoms or isolated positive constructs, while less frequently integrating co-existing orientations such as active engagement and pressure-driven withdrawal/goal attenuation within one analytical framework. Second, systematic evidence on how life attitudes differ across gender, academic year, and disciplinary backgrounds remains limited in terms of both measurement comparability and contextual interpretability [1].

To address these gaps, this study investigates Chinese college students using a mixed-method design combining questionnaire surveys and semi-structured interviews. We develop and examine a four-dimensional structure of life attitudes: positive mindset, lying-flat mindset, Buddha-like mindset, and negative mindset. Quantitatively, we describe overall distributions and test demographic differences across key student subgroups; qualitatively, we use interview evidence to illuminate students' meaning-making and action logics under pressure. The study ultimately aims to provide actionable evidence for more targeted student development support and values-oriented guidance within universities.

2 LITERATURE REVIEW

College students' life attitudes can be conceptualized as a relatively stable orientation system for appraising and coping with salient developmental tasks, integrating value–motivational commitments (for example, meaning in life, purpose, and life goals) with cognitive, affective, and behavioral adaptation patterns [9]. Recent work in positive and developmental psychology increasingly treats meaning and purpose as pivotal resources during emerging adulthood, showing robust associations with lower psychological distress and better mental health [9]. Importantly, person-centered evidence suggests that meaning-related functioning is heterogeneous and clusters into distinct profiles rather than varying only along a single linear continuum [10]. Evidence from Chinese university populations further indicates that these profiles exhibit meaningful stability and transition over time, and that both the configuration and change of meaning-related profiles are systematically associated with mental health outcomes. [3].

Against the backdrop of social transformation and post-pandemic uncertainty, psychological distress and stress exposure among Chinese university students remain salient. A recent systematic review and meta-analysis on depression among university students in China underscores both the substantive burden and the multi-layered nature of associated factors, calling for integrative models that connect contextual stressors with individual mechanisms [1]. Complementarily, qualitative evidence in the post-pandemic era highlights sleep difficulties, anxiety, and stress as prominent challenges, with perceived academic pressure, peer influence, and social-acceptance pursuits as key explanatory themes [11]. In this context, academic involution—a form of escalating, inefficient competition under perceived scarcity—has moved from public discourse to psychometrically grounded research. Empirical studies show that an academic involution atmosphere predicts students' stress responses via relative deprivation and involution behavior in a chain-mediated pathway [12]; related models link involution atmosphere to mental exhaustion through deprivation and perceived academic pressure [13]; and recent scale-development efforts provide validated instruments for assessing involution tendencies and involutional behaviors among college students [14]. Closely related "low-desire" orientations, such as lying flat and Buddha-like coping, are also entering quantitative scholarship: social-comparison research explains lying-flat motivation via personal relative deprivation; a validated "lying flat tendency" scale has been developed for Chinese youth [15]; and emerging evidence suggests that Buddha-like characteristics may relate to social responsibility in a conditional manner depending on self-efficacy and social support [5].

Intervention-oriented literature indicates a shift from solely problem-focused counseling toward scalable promotion–prevention approaches, with digital delivery becoming increasingly prominent. A recent systematic review and meta-analysis of digital mental health interventions for university students reports medium effects on depression and anxiety symptom severity and suggests that intervention format (e.g., fully automated vs guided) may matter for outcomes [16]. In parallel, meaning-oriented coping appears protective for college students' mental health, operating partly through school connectedness—an actionable pathway for campus-based education and support design [17]. Nevertheless, prior studies rarely operationalize life attitudes as an integrated spectrum that simultaneously captures "positive," "lying-flat," "Buddha-like," and "negative" orientations, nor do they routinely test these orientations within a unified mechanism linking involution atmosphere, relative deprivation, and exhaustion-related processes. This motivates context-sensitive measurement and mechanism research to inform targeted, differentiated educational guidance and mental health promotion in higher education.

3 RESEARCH METHOD

3.1 Participants

Participants were recruited from a university in Guangzhou via an on-site paper-and-pencil survey. The sampling frame covered students from nine colleges within the university. A total of 320 questionnaires were collected through in-person completion. After data screening, 287 valid questionnaires were retained, resulting in an effective response rate of 89.69% (287/320). Eligibility criteria specified that respondents must be currently enrolled students from the nine participating colleges. Questionnaires were excluded if they exhibited signs of careless responding (e.g., random or inattentive response patterns) or contained aberrant/extreme values indicating invalid entries. The final sample consisted of 154 males (53.66%) and 133 females (46.34%). In terms of academic standing, master's students accounted for 28.92% (n = 83), followed by third-year undergraduates at 27.18% (n = 78). First-year undergraduates constituted

24.74% (n = 71), while second-year undergraduates and fourth-year undergraduates made up 8.01% (n = 23) and 5.92% (n = 17), respectively. Doctoral students represented the smallest proportion at 5.23% (n = 15).

3.2 Scale Development

To capture contemporary Chinese university students' life attitudes in a social-transition context, this study developed a context-sensitive Life Attitude Scale (LAS) consisting of four co-existing orientations: positive mindset, lying-flat mindset, Buddha-like mindset (foxi mentality), and negative mindset. Item generation followed established scale-development principles emphasizing conceptual definition, item pool expansion, iterative refinement, and psychometric evaluation [18]. In addition to synthesizing relevant work on attitude measurement and student psychosocial functioning, the wording and content were informed by recent empirical efforts to operationalize emergent youth cultural-psychological tendencies such as "lying flat", thereby improving contextual fidelity and content coverage [15].

The resulting LAS includes 19 items across four subscales: positive mindset (5 items), lying-flat mindset (5 items), Buddha-like mindset (5 items), and negative mindset (4 items). All items were administered in Chinese and rated on a 5-point Likert-type agreement scale (1 = strongly disagree, 5 = strongly agree) [19].

To strengthen content validity and response-process validity, we adopted a multi-stage item refinement procedure. First, an expert panel systematically evaluated each item's relevance, clarity, and representativeness, and items were revised, merged, or removed to align the scale content with the target construct definition [20,21]. At the decision-making level, CVI-type quantification was used to summarize expert judgments, providing a transparent and replicable basis for item retention and revision [22]. Second, cognitive interviewing and comprehension probing were conducted during pretesting to detect semantic ambiguity, culturally loaded wording, and unintended interpretations, thereby improving item interpretability and face validity [18,23]. This iterative process yielded the final item set used in the formal survey.

3.3 Data Processing

This study primarily relied on survey-based quantitative data, supplemented by interview materials, to profile the overall pattern of college students' life attitudes and examine between-group differences. All items were measured using 5-point Likert-type response formats and were aggregated according to the predefined scoring rules. After on-site paper-and-pencil collection, responses were digitized and analyzed using SPSS 27.0; data cleaning removed invalid cases showing clear indications of careless responding and aberrant values, consistent with recommended practices for improving survey data quality [19].

Psychometric screening focused on internal consistency and factorability diagnostics. Internal consistency was evaluated with Cronbach's alpha (overall scale $\alpha = 0.775$), and factorability was assessed using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity (overall KMO = 0.871, with Bartlett's test supporting non-identity of the correlation matrix) [24]. Because the study relied on single-source self-reports, we also assessed the potential risk of common method bias using procedural standardization and a post hoc Harman's single-factor test (unrotated EFA with all items). Four factors with eigenvalues greater than 1 were extracted, explaining 62.591% of total variance, and the first factor explained 33.433%, which did not indicate a dominant single method factor; however, given accumulated evidence that Harman-type diagnostics are often underpowered and design-dependent, the result was treated as a conservative check rather than definitive proof [25].

4 RESULTS

4.1 Overall Status of College Students' Life Attitudes

Based on the descriptive statistics reported in Table 1, this study outlines the overall profile of college students' life attitudes. The Life Attitude Scale comprises four dimensions: positive mindset, lying-flat mindset, Buddha-like mindset, and negative mindset. Overall, the mean score for the positive mindset was clearly above the scale midpoint, whereas the lying-flat mindset and negative mindset were below the midpoint. A Buddha-like mindset was close to the midpoint. This pattern indicates a predominantly positive profile, accompanied by a moderately neutral regulation orientation, while still suggesting meaningful individual differences in low-investment and negative tendencies.

For a positive mindset, the mean was 4.093, and the median was 4.200, with a standard deviation of 0.825, indicating a high level with relatively limited dispersion. The skewness was -1.098, suggesting that scores clustered toward the upper end of the scale, and the kurtosis was 1.692, implying a comparatively peaked distribution and substantial convergence in positive orientation. In contrast, the lying-flat mindset showed a mean of 2.419 and a median of 2.400, both below the midpoint, indicating generally low endorsement of low-investment tendencies. Its standard deviation was 0.885, and its skewness was 0.471, reflecting noticeable heterogeneity with a smaller subgroup reporting higher lying-flat tendencies. The kurtosis of -0.025 suggests a relatively flat distribution without pronounced concentration.

Buddha-like mindset had a mean of 2.987 and a median of 3.000, approximating the midpoint and reflecting an overall neutral tendency that may represent moderation-oriented psychological regulation under stress. The standard deviation was 0.933, with skewness of -0.075 and kurtosis of -0.462, indicating an approximately symmetric and relatively flat distribution. Negative mindset yielded a mean of 2.190 and a median of 2.000, suggesting low overall negative orientation; however, the standard deviation of 0.940 and skewness of 0.656 indicate that a non-negligible subset

reported elevated negative tendencies. Taken together, the results portray a positive dominant profile with differentiated variation in lying-flat and negative tendencies, warranting further subgroup-focused analyses in subsequent sections.

Table 1 Descriptive Statistics of College Students' Life Attitudes

Statistic	Positive mindset	Lying flat mindset	Buddha-like mindset	Negative mindset
Mean	4.093	2.419	2.987	2.190
Median	4.200	2.400	3.000	2.000
Standard deviation	0.825	0.885	0.933	0.940
Variance	0.681	0.783	0.87	0.883
Skewness	-1.098	0.471	-0.075	0.656
Std. error of skewness	0.144	0.144	0.144	0.144
Kurtosis	1.692	-0.025	-0.462	0.098
Std. error of kurtosis	0.287	0.287	0.287	0.287

4.2 Analysis of Positive Mindset Among College Students

At the overall level, the positive mindset dimension demonstrated consistently high scores (Table 2), indicating that participants tended to endorse constructive orientations regarding meaning in life, subjective well-being, self-acceptance, coping with setbacks, and the perceived value of effort. The item means ranged from 4.02 to 4.22, all above the scale midpoint, suggesting that positive orientation is broadly prevalent in the sample.

Item-level distributions further clarify this pattern. The statement "I believe life is beautiful and meaningful" showed the strongest endorsement, with 52.26 percent selecting "strongly agree" and a mean of 4.22, reflecting pronounced meaning affirmation. The item capturing happiness and energetic engagement in daily life yielded a mean of 4.02, with a high combined proportion of "agree" and "strongly agree," indicating generally sustained positive affect and engagement. Self-kindness and health concern also received strong support, with a mean of 4.08, suggesting relatively clear endorsement of self-acceptance and well-being awareness. The items on viewing setbacks as opportunities for learning and endorsing effort as a pathway to happiness had means of 4.05 and 4.09, respectively, indicating constructive appraisal and effort valuation in most respondents. Meanwhile, standard deviations were close to 1.00 across items, and a nontrivial share of "uncertain" responses appeared in several items, such as 14.98 percent for setback appraisal and 19.16 percent for happiness and engagement, implying meaningful heterogeneity and an identifiable subgroup with less consolidated positive orientation.

Regarding subgroup variation, the Kruskal-Wallis test indicated a statistically significant difference in positive mindset across major categories ($p = 0.03$), suggesting that positive orientation varies by disciplinary background. The difference by student leadership status was marginal ($p = 0.057$), implying a potential association that warrants further verification in subsequent analyses. Other grouping variables, including gender, grade level, and political status, did not show significant differences (p values greater than 0.05), indicating comparatively limited explanatory power for positive mindset in this dataset.

Table 2 Positive Mindset Among College Students

Item	Strongly disagree	Disagree	Uncertain	Agree	Strongly agree	Mean	SD
I believe life is beautiful and meaningful.	3.48%	3.83%	11.85%	28.57%	52.26%	4.22	1.03
I often feel happy and engage in daily life, work, and study with enthusiasm.	2.79%	4.53%	19.16%	34.49%	39.02%	4.02	1.01
I treat myself kindly, accept myself, and pay close attention to my physical and mental health.	2.09%	5.57%	16.72%	33.45%	42.16%	4.08	1.00
I view temporary setbacks or failures as opportunities for learning and growth.	2.44%	6.97%	14.98%	34.49%	41.11%	4.05	1.03
I believe happiness can be achieved through persistent effort and striving.	2.44%	7.67%	11.50%	35.19%	43.21%	4.09	1.03

4.3 Analysis of Lying-flat Mindset Among College Students

Using the item-level descriptive statistics reported in Table 3, this section examines the lying-flat mindset among college students. Overall, item means ranged from 1.99 to 2.63, all below the scale midpoint, indicating generally low endorsement and suggesting that lying-flat orientation is not dominant in the sample. Meanwhile, standard deviations ranged from 1.14 to 1.26, implying meaningful heterogeneity and indicating that a subgroup of students reported relatively higher endorsement on specific lying-flat indicators.

Item distributions provide additional nuance. The highest mean was observed for the fatalistic statement indicating limited personal intervention and control, with a mean of 2.63 and a standard deviation of 1.22, and 28.57 percent selecting "uncertain," suggesting ambivalence regarding personal agency. The item "I have no specific goals or pursuits"

yielded a mean of 2.45 and a standard deviation of 1.15, with a combined 54.71 percent selecting "strongly disagree" or "disagree," indicating that most students retained goal orientation. The statement about severe involution and the limited effect of personal effort showed a mean of 2.54 and a standard deviation of 1.26, reflecting differentiated appraisals of competitive pressure. The item capturing reluctance to undertake tasks requiring substantial effort had a mean of 2.48 and a standard deviation of 1.20, suggesting a tendency among some respondents to reduce investment under high-demand conditions. Importantly, the item describing habitual minimal effort and perfunctory completion had the lowest mean of 1.99 and a standard deviation of 1.14, with 44.25 percent selecting "strongly disagree," indicating that most students did not endorse a fully disengaged behavioral orientation.

Subgroup analyses further revealed structured variation in lying-flat orientation. The Kruskal-Wallis test indicated statistically significant differences in lying-flat mindset by gender (p less than 0.0001), political status (p equals 0.0096), and father's occupation (p equals 0.0376), suggesting that lying-flat tendencies are not randomly distributed and may relate to social roles and family background. By contrast, grade level, region, and mother's occupation did not show significant differences (p values greater than 0.05), indicating comparatively limited explanatory relevance in this dataset. Taken together, the findings suggest a generally low level of lying-flat endorsement alongside meaningful subgroup differentiation that warrants further mechanism-oriented interpretation.

Table 3 Lying-flat mindset among college students

Item	Strongly disagree	Disagree	Uncertain	Agree	Strongly agree	Mean	SD
I tend to accept fate and believe everything is predetermined, so I should not intervene or control too much.	20.56%	28.22%	28.57%	12.89%	9.76%	2.63	1.22
I have no specific goals or pursuits in life and study.	24.74%	29.97%	24.74%	16.38%	4.18%	2.45	1.15
I think academic involution is severe, and personal effort cannot change much.	25.44%	26.48%	26.13%	12.54%	9.41%	2.54	1.26
I am unwilling to do things that require great effort to accomplish.	25.09%	28.92%	26.48%	12.20%	7.32%	2.48	1.20
No matter what it is, I do not make efforts and only complete what I should do perfunctorily.	44.25%	28.57%	15.68%	6.62%	4.88%	1.99	1.14

4.4 Analysis of Buddha-like Mindset Among College Students

According to the item-level descriptive statistics reported in Table 4, this section examines the Buddha-like mindset among college students. Overall, item means ranged from 2.62 to 3.44, indicating a moderate level that clustered around the scale midpoint. In contrast to the high level observed for positive mindset, Buddha-like mindset appears to reflect a moderation-oriented regulation tendency under pressure and uncertainty rather than a clearly positive or negative orientation. Standard deviations ranged from 1.18 to 1.34, indicating substantial heterogeneity in this mindset within the sample.

Item distributions further illustrate this pattern. The highest mean was observed for the relational stance item, "In love and friendship, I do not initiate, I do not force, and I do not settle," with a mean of 3.44 and a standard deviation of 1.34. The combined proportion of "agree" and "strongly agree" was 52.26 percent, suggesting that many students endorse a restrained and non-compulsive approach to close relationships. The lifestyle preference item, emphasizing slowing down and relaxing as an ideal way of living, yielded a mean of 3.26 and a standard deviation of 1.23, with notable endorsement levels, indicating broad acceptance of a low-intensity lifestyle orientation. By comparison, the item stating that future development should be left to "go with the flow" showed a lower mean of 2.85 with a relatively high "uncertain" response rate of 27.18 percent, suggesting ambivalence regarding applying this stance to long-term planning. The items reflecting competition avoidance and achievement de-emphasis produced means of 2.77 and 2.62, respectively, and both showed uncertainty rates around one quarter, indicating that students were more divided and less uniform on performance- and competition-related issues.

Subgroup analyses revealed structured differentiation. The Kruskal-Wallis test indicated statistically significant differences in Buddha-like mindset by gender (p equals 0.0024) and father's occupation (p equals 0.0460), suggesting that moderation-oriented regulation tendencies vary across social roles and family background. In contrast, grade level, political status, and region did not show significant differences (p values greater than 0.05), indicating comparatively limited explanatory relevance for these variables in this dataset. Overall, the results suggest a moderate level of Buddha-like orientation alongside meaningful heterogeneity, warranting further mechanism-oriented interpretation with attention to gender and family background.

Table 4 Buddha-like Mindset Among College Students

Item	Strongly disagree	Disagree	Uncertain	Agree	Strongly agree	Mean	SD
For my future development, I think it is fine to go with the flow.	17.42%	24.04%	27.18%	19.16%	12.20%	2.85	1.26

Item	Strongly disagree	Disagree	Uncertain	Agree	Strongly agree	Mean	SD
There is no better way of living than slowing down, relaxing, and being natural and easygoing.	11.50%	14.29%	27.53%	29.97%	16.72%	3.26	1.23
I am satisfied with my current state and do not want to be drawn into too much competition to prove myself.	16.72%	26.48%	27.53%	21.60%	7.67%	2.77	1.19
In work or study, good achievements are nice, but having no achievements is also acceptable.	19.16%	31.36%	25.44%	16.72%	7.32%	2.62	1.18
In love and friendship, I do not initiate, I do not force, and I do not settle.	11.85%	12.89%	23.00%	24.04%	28.22%	3.44	1.34

4.5 Analysis of Negative Mindset Among College Students

On the basis of the item-level descriptive statistics reported in Table 5, this section examines the negative mindset among college students. Overall, item means ranged from 1.87 to 2.40, all below the scale midpoint, indicating generally low endorsement of negative orientation and suggesting that negative experiences were not dominant in the sample. However, standard deviations ranged from 1.11 to 1.29, implying meaningful heterogeneity and indicating that a subset of students reported elevated negativity in areas such as perceived meaninglessness, interpersonal distrust, low frustration tolerance, and future-oriented anxiety.

Item distributions provide additional detail. The statement “I often feel that everything in life seems meaningless” yielded a mean of 2.20, with 63.06 percent selecting “strongly disagree” or “disagree,” suggesting that most students retained a sense of meaning and motivation, while 23.69 percent selecting “uncertain” indicates a notable subgroup with less consolidated meaning orientation. The item on perceived lack of sincerity and kindness in others had a mean of 2.30, and 17.42 percent endorsed “agree” or “strongly agree,” pointing to reduced social trust among a minority of students. The lowest mean was observed for the frustration tolerance item, “I can hardly bear any setbacks or failures,” with a mean of 1.87, and 77.35 percent selecting “strongly disagree” or “disagree,” indicating relatively strong perceived coping capacity in most respondents; nevertheless, 11.15 percent endorsed “agree” or “strongly agree,” suggesting a vulnerable subgroup. The highest mean was found for “I often feel frustrated and am worried and anxious about the future” (mean equals 2.40), with 21.95 percent endorsing “agree” or “strongly agree,” indicating that future-related anxiety was comparatively more salient and more differentiated in this sample.

Subgroup analyses further revealed structured variation. The Kruskal-Wallis test indicated statistically significant differences in negative mindset by political status (p equals 0.045), mother’s occupation (p equals 0.046), and student leadership status (p equals 0.025), suggesting that negative orientation may be associated with social identity, family background, and role experiences. In contrast, gender, grade level, and region did not show significant differences (p values greater than 0.05), indicating comparatively limited explanatory relevance for these variables in this dataset. Overall, the results suggest low average negativity alongside meaningful differentiation, particularly regarding future anxiety, interpersonal trust, and frustration tolerance, warranting further mechanism-oriented interpretation in subsequent analyses.

Table 5 Negative Mindset Among College Students

Item	Strongly disagree	Disagree	Uncertain	Agree	Strongly agree	Mean	SD
I often feel that everything in life seems meaningless.	35.19%	27.87%	23.69%	8.36%	4.88%	2.20	1.15
I often feel that people around me lack sincerity and kindness.	31.36%	29.97%	21.25%	12.54%	4.88%	2.30	1.18
I can hardly bear any setbacks or failures.	50.52%	26.83%	11.50%	7.67%	3.48%	1.87	1.11
I often feel frustrated and am worried and anxious about the future.	34.15%	21.25%	22.65%	14.63%	7.32%	2.40	1.29

5 DISCUSSION

5.1 Interpreting the Overall Profile and Differentiated Patterns of College Students’ Life Attitudes

The present study depicts a clearly differentiated four-dimensional profile of college students’ attitudes toward life. At the aggregate level, a positive mindset was markedly above the scale midpoint, whereas the lying-flat mindset and negative mindset were both below the midpoint; the Buddha-like mindset clustered around the midpoint. This pattern suggests that, within this sample, a generally constructive orientation toward life coexists with two non-mainstream but consequential tendencies, namely low-investment withdrawal and negative meaning-related distress. Theoretically, these results support the argument that contemporary youth life attitudes are not adequately captured by a simple

positive–negative continuum; rather, they are better conceptualized as a configuration of co-existing orientations that reflect meaning affirmation, effort valuation, pressure-related goal attenuation, and vulnerability-linked negativity.

First, the dominance of a positive mindset is consistent with person-centered evidence indicating that meaning-related orientations among Chinese university students are heterogeneous but, for many individuals, remain stable and protective for mental health outcomes [3]. In a broader mental-health landscape where depression, anxiety, and sleep problems remain salient among Chinese university students [1,2,11], a high average level of positive mindset can be interpreted as a form of adaptive meaning maintenance and effort-based agency. Importantly, the item-level distributions also indicate meaningful heterogeneity, as a nontrivial share of respondents selected "uncertain" on core meaning and coping items. This nuance aligns with recent network-oriented findings showing that meaning in life and depressive symptoms may be intertwined with perceived social support in student populations, implying that meaning affirmation may coexist with localized vulnerabilities rather than functioning as a uniform protective shield [6].

Second, the moderate endorsement of a Buddha-like mindset appears to reflect a regulation-oriented stance under pressure and uncertainty, rather than a uniformly disengaged orientation. Empirically, "foxi" characteristics have begun to enter psychological scholarship as a culturally specific coping repertoire whose implications may vary by resource conditions. For instance, evidence suggests that Buddha-like characteristics may relate to social responsibility through conditional pathways involving self-efficacy and social support [5]. In this study, the relatively higher endorsement of relational restraint and low-intensity lifestyle preferences, alongside lower endorsement of "go with the flow" in long-term planning and achievement de-emphasis, suggests that a Buddha-like mindset may function as selective de-intensification: students may reduce relational and lifestyle pressures while still maintaining instrumental commitments in study and career domains. This interpretation is compatible with stress-process perspectives, emphasizing that young adults may adopt partial disengagement strategies to preserve psychological balance without fully abandoning achievement goals.

Third, the generally low but heterogeneous lying-flat mindset resonates with recent theoretical and measurement advances that locate "lying flat" in perceived scarcity and low return on effort. A validated "lying flat tendency" scale has been developed for Chinese youth [15], and integrative theory has framed lying flat as the outcome of multiple social-psychological pathways, including perceived unfairness, diminished efficacy of effort, and withdrawal from competitive escalation [6]. Complementary evidence indicates that personal relative deprivation can suppress self-improvement motivation, thereby legitimizing disengagement from competitive striving [4]. Within higher education, the rise of academically "involutional" climates provides a plausible contextual driver: involution atmosphere has been linked to students' stress responses and mental exhaustion through relative deprivation and perceived academic pressure [12–14]. Against this backdrop, the present finding that most students reject fully perfunctory effort, while a minority shows higher endorsement of fatalism and reduced willingness to invest in difficult tasks, may reflect a threshold process: a lying-flat mindset emerges more strongly when students perceive effort as unlikely to be rewarded and when competitive escalation becomes psychologically inefficient.

Finally, the pattern of subgroup differentiation provides further interpretive leverage. Positive mindset varied significantly by major category, suggesting that disciplinary cultures and opportunity structures may shape meaning affirmation and effort valuation, while leadership status showed a marginal association that may reflect role-based agency and structured engagement. By contrast, the lying-flat mindset differed by gender, political status, and father's occupation, and the Buddha-like mindset differed by gender and father's occupation, highlighting the relevance of social roles and family background in shaping pressure-coping repertoires. Negative mindset differed by political status, mother's occupation, and student leadership status, indicating that social identity, family resource contexts, and role experiences may be consequential for vulnerability-linked negativity. Taken together, these findings extend current literature by demonstrating that low-investment and negative tendencies are not randomly distributed: they may be socially patterned responses to perceived pressure, relative deprivation, and resource security. This differentiated structure is consistent with calls for integrative models that connect contextual stressors with individual meaning and coping mechanisms in Chinese university populations [1,11], and it strengthens the case for targeted, subgroup-sensitive student development and support strategies rather than one-size-fits-all guidance.

5.2 Research Significance

This study offers significance at theoretical, methodological, and practical levels by clarifying the structure of college students' life attitudes in contemporary China and by providing evidence on how such attitudes are socially patterned.

At the theoretical level, the study advances a multidimensional understanding of life attitudes. Rather than treating students' orientation toward life as a single positive–negative continuum, the findings support a four-dimensional configuration in which positive mindset, lying-flat mindset, Buddha-like mindset, and negative mindset can coexist within the same population and reflect distinct psychological functions. This framing helps reconcile seemingly contradictory public narratives that portray today's students as simultaneously aspiring and disengaged, and it provides a conceptually tractable basis for integrating meaning-related orientations, coping preferences, and withdrawal tendencies into a unified explanatory perspective.

At the methodological level, the study contributes a structured operationalization of life attitudes through a concise scale with acceptable psychometric performance and clear interpretability at the item and dimension levels. By combining descriptive profiling with nonparametric group comparisons, the analysis offers a replicable template for

mapping attitudinal distributions and identifying differentiated subgroups, which can inform subsequent theory testing and mechanism-oriented modeling in student development research.

At the practical level, the results provide evidence for more targeted, subgroup-sensitive educational and support strategies in universities. The overall dominance of a positive mindset suggests a solid foundation for developmental education, while the nontrivial presence of lying-flat and negative tendencies points to specific risk segments that may require differentiated interventions. Moreover, the observed differences across disciplinary background, political status, family occupational background, and leadership experience underscore that life attitudes are embedded in social and institutional contexts. Universities may therefore improve effectiveness by aligning ideological and psychological education with students' lived pressures and resource conditions, strengthening meaning construction and agency, and offering tailored support for those exhibiting elevated withdrawal or future-oriented anxiety.

5.3 Limitations and Future Research Directions

This study has several limitations. First, the sample was drawn from a single university in Guangzhou using on-site offline administration, which limits external validity across regions, institutional types, and student populations. Second, the evidence is largely cross-sectional and self-reported, constraining causal interpretation and leaving the temporal dynamics of life attitudes unresolved. Third, although the scale showed acceptable internal consistency and factorability, the current manuscript does not yet provide a full psychometric validation package, such as confirmatory factor analysis on an independent sample, measurement invariance tests across key subgroups, and convergent and discriminant validity against established constructs.

Future research should therefore broaden sampling and strengthen design and measurement rigor. Multi-site studies covering different university tiers and regions, combined with longitudinal or repeated-measures designs, would clarify whether a positive mindset is stable and protective over time, whether a lying-flat mindset emerges under effort–reward imbalance, and whether a Buddha-like mindset functions as short-term regulation or an enduring orientation. Methodologically, future work should test measurement invariance before interpreting subgroup mean differences and should extend validation through test–retest reliability and external validity checks (for example, associations with meaning in life, perceived stress, academic burnout, and social support). Mechanism-oriented models, such as mediation and moderation analyses incorporating perceived academic involvement, employability expectations, and institutional support, would further explain why subgroup differences arise and how universities can design more targeted interventions.

6 CONCLUSION

This study contributes to a clearer and more nuanced understanding of how contemporary Chinese college students orient themselves toward life in a period marked by intensified competition, accelerated social change, and heightened uncertainty. Rather than reducing students' life attitudes to a single positive–negative spectrum, the findings support a differentiated structure in which constructive agency, moderation-oriented regulation, low-investment withdrawal, and vulnerability-linked negativity coexist as distinct but interrelated orientations. Conceptually, this multidimensional framing helps reconcile seemingly contradictory public narratives about youth as both striving and disengaged, and it offers a more analytically precise language for discussing students' psychological adaptation under structural pressures.

Beyond describing attitudinal patterns, the study underscores that life attitudes are socially embedded and therefore educationally actionable. The observed differentiation across disciplinary contexts, role experiences, and family background implies that students' orientations are shaped by how they perceive effort–reward expectations, the availability of support resources, and the credibility of future pathways. This has broader implications for higher education governance and student development: effective educational responses should move beyond uniform exhortations toward "positivity" and instead strengthen meaning construction and agency while addressing the concrete conditions that foster disengagement and anxiety. By providing a structured lens and empirical grounding for identifying both protective orientations and vulnerable segments, this research supports more targeted, context-sensitive strategies in ideological education and student support, and it lays a foundation for future multi-site, longitudinal, and mechanism-focused work on youth adaptation in the evolving landscape of Chinese modernization.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This research was funded by the Guangdong Agriculture Industry Business Polytechnic, which studies the path of deep integration of Party building and business work of secondary Party organizations in higher vocational colleges, taking Guangdong universities as an example (Project No.: XJDJZD2507).

REFERENCES

[1] LIN Z Z, CAI H W, HUANG Y F, et al. Prevalence of depression among university students in China: a systematic review and meta-analysis. BMC Psychology, 2025, 13, 373. DOI: 10.1186/s40359-025-02688-y.
[2] SUN C, ZHU Z, ZHANG P, et al. Exploring the interconnections of anxiety, depression, sleep problems and health-promoting lifestyles among Chinese university students: a comprehensive network approach. Frontiers in Psychiatry, 2024, 15. DOI: 10.3389/fpsyt.2024.1402680.
[3] KWOK S Y C L, FANG S, HUANG B M, et al. Stability and changes in meaning in life profiles and their impact on mental health among chinese university students: a latent transition analysis. Frontiers in Psychology, 2025, 16. DOI: 10.3389/fpsyg.2025.1529851.
[4] ZHENG X, JING C, LIU Y, et al. Why are people "Lying Flat"? Personal relative deprivation suppresses self-improvement motivation. The British Journal of Social Psychology, 2023, 62(2): 932-948. DOI: 10.1111/bjso.12611.
[5] QI Z, HAN F. Research on the impact of college students' "Buddha-like" characteristics on social responsibility and the mediating role of social support. Frontiers in Psychology, 2025, 16. DOI: 10.3389/fpsyg.2025.1691896.
[6] ZHANG L, HUI B P H, KONG F, et al. Why People "Lie Flat"? An Integrative Framework of Social-Psychological Pathways in China. Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc, 2025, 29(4): 371-382. DOI: 10.1177/10888683251358516.
[7] ZHANG S sen, ZHANG W hua, YONG S hong, et al. Network analysis of meaning in life, perceived social support, and depressive symptoms among vocational undergraduate students. Frontiers in Psychiatry, 2025, 16. DOI: 10.3389/fpsyt.2025.1510255.
[8] ZHU S, WANG L, GAN Y. Lying flat to play on smartphone excessively: the role of self-esteem. Frontiers in Psychology, 2025, 16: 1516869. DOI: 10.3389/fpsyg.2025.1516869.
[9] HE X X, WANG X qiang, STEGER M F, et al. Meaning in life and psychological distress: A meta-analysis. Journal of Research in Personality, 2023, 104: 104381. DOI: 10.1016/j.jrp.2023.104381.
[10] SU H, YE T, WU X. Revealing the heterogeneity of meaning in life and its relationship with creativity among university students: A latent profile analysis study. Acta Psychologica, 2025, 254: 104852. DOI: 10.1016/j.actpsy.2025.104852.
[11] NING X, LUO X, GUO S. Researching into Chinese university students' mental health in the post-pandemic era–problems and causes. Frontiers in psychology, 2024, 15: 1393603. DOI: 10.3389/fpsyg.2024.1393603.
[12] LIU A, SHI Y, ZHAO Y, et al. Influence of academic involution atmosphere on college students' stress response: the chain mediating effect of relative deprivation and academic involution. BMC Public Health, 2024, 24(1): 870. DOI: 10.1186/s12889-024-18347-7.
[13] NI J, LIU A, SHI Y, et al. Impact of academic involution atmosphere on college students' mental exhaustion: A chain mediation model. International Journal of Educational Research, 2024, 128: 102475. DOI: 10.1016/j.ijer.2024.102475.
[14] HE Y, LIU X. Development and validation of academic involution scale for college students. Scientific Reports, 2025, 15(1): 20876. DOI: 10.1038/s41598-025-04850-3.
[15] LU H, HOU J, HUANG A, et al. Development and Validation of the "Lying Flat" Tendency Scale for the Youth. Behavioral Sciences, 2023, 13(11). DOI: 10.3390/bs13110915.
[16] MADRID-CAGIGAL A, KEALY C, POTTS C, et al. Digital Mental Health Interventions for University Students With Mental Health Difficulties: A Systematic Review and Meta-Analysis. Early Intervention in Psychiatry, 2025, 19(3): e70017. DOI: 10.1111/eip.70017.
[17] LU Q, CHEN Q, ZHANG Y, et al. How Does Meaning-Centered Coping Influence College Students' Mental Health? The Mediating Roles of Interdependent Self-Construal and School Connectedness. Behavioral Sciences, 2025, 15(7). DOI: 10.3390/bs15070955.
[18] BOATENG G O, NEILANDS T B, FRONGILLO E A, et al. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Frontiers in public health, 2018, 6: 149. DOI: 10.3389/fpubh.2018.00149.
[19] SULLIVAN G M, ARTINO A R. Analyzing and Interpreting Data From Likert-Type Scales. Journal of Graduate Medical Education, 2013, 5(4): 541-542. DOI: 10.4300/JGME-5-4-18.
[20] ALMANASREH E, MOLES R, CHEN T F. Evaluation of methods used for estimating content validity. Research in Social and Administrative Pharmacy, 2019, 15(2): 214-221. DOI:10.1016/j.sapharm.2018.03.066.
[21] TERWEE C B, PRINSEN C A C, CHIAROTTO A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Quality of Life Research, 2018, 27(5): 1159-1170. DOI: 10.1007/s11136-018-1829-0.
[22] POLIT D F, BECK C T, OWEN S V. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health, 2007, 30(4): 459-467. DOI: 10.1002/nur.20199.
[23] BALZA J S, CUSATIS R, MCDONNELL S M, et al. Effective questionnaire design: How to use cognitive interviews to refine questionnaire items. Journal of Neonatal-Perinatal Medicine, 2022, 15(2): 345-349. DOI: 10.3233/NPM-210848.
[24] CRONBACH L J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16(3): 297-334. DOI: 10.1007/BF02310555.
[25] PODSAKOFF P M, ORGAN D W. Self-reports in organizational research: Problems and prospects. Journal of management, 1986, 12(4): 531-544. DOI: 10.1177/014920638601200408.