

Journal of Advances in Engineering and Technology

Editor-in-Chief:

Prof. Dr. José Pereira

Universidade de Lisboa, Portugal

Copyright © 2025. ASIA PACIFIC SCIENCE PUBLICATIONS
COMPANY LIMITED. Complimentary Copy.

Journal of Advances in Engineering and Technology

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

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

About Publisher:

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

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

Publisher Headquarter

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

Fujian Province Office, China

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

Table of Contents

- 1 Study on Organic Molecular Damage Mechanism and Environmental Dependence of Silicone Rubber under the Synergistic Action of Complex Conditions**
Weixiao Li, Yibo Zhang, Shumin Yu, Yuxi Lin, Mingshen Xu
- 29 Design and Implementation of Time-Frequency Taming Algorithm of Rubidium Atomic Clock Based on FPGA**
Sun Songfei, Hou Yuanhang, Lv Yiheng, Wang Zhiming, Chen Qun
- 36 Design of Lightweight, High Precision, Multi Scene Module Based on Narrowband Internet of Things**
Sun Feiyu, Liu Xiangyu, Shangguan Lena, Ren Jingbo, Ma Qingsong, Chen Qun
- 43 Analysis of the Development Path of Artificial Intelligence Technology in the Big Data Era**
Yusen Huang
- 48 Research on APP Design for Albinism Family Psychological Therapy Based on User Needs**
Fan Wu, Jiaming Zeng, Congrong Xiao
- 56 Synthesis of Cobalt Doping Titanium Dioxide Fiber Supported by Reduced Graphene Oxide for Photocatalytic Degradation of Toluene in Air**
Guilan Gao, Xiran Zhang, Li Zhang, Shuai Chen, Pinnan Li
- 67 Study on the Extraction Process and Pharmacological Activity of Gentiopicroside from Gentiana**
Luo Dan
- 74 Synergistic Optimization and Risk Control: Integrated Price Forecasting Models**
Mingshen Xu, Tianxiang Hu, Boyi Zhang, Zhang Hanrui, Kexiao Wu, Teng Zhang, Xiaotian Li, Qianxve Mo, Xianliu Feng
- 91 Research on Financial Risks and Compliance Challenges in Cross-Border Mergers and Acquisitions**
Yixin Chen
- 101 Sentiment Analysis of IMDB Movie Reviews Based on LSTM**
Jiahao Lu, Hongji Fan, Yali Zhang

Study on Organic Molecular Damage Mechanism and Environmental Dependence of Silicone Rubber under the Synergistic Action of Complex Conditions

Weixiao Li¹, Yibo Zhang¹, Shumin Yu¹, Yuxi Lin², Mingshen Xu^{3*}

1.Department of Environmental Science and Engineering, North China Electric Power University, Baoding 071003, Hebei, China

2.Department of Electrical Engineering, North China Electric Power University, Baoding 071003, Hebei, China

3.Department of Mathematics and Physics, North China Electric Power University, Baoding 071003, Hebei, China

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

Copyright: 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: This study focused on the mechanism of corona discharge and nitric acid damage to silicone rubber organic molecules in order to solve the aging problem of composite external insulation materials in heavy pollution environment. The effects of corona strength, duration and nitric acid corrosion on the physical and chemical properties of silicone rubber were systematically investigated through the independent corona aging test system of composite materials (including 54 sets of pin-plate electrodes), combined with scanning electron microscopy (SEM), static contact Angle analysis, thermal stimulation current (TSC) and Fourier infrared spectroscopy (FTIR). The results show that: (1) During corona discharge, the high energy charged particles preferously destroy C-H bond and Si-C bond through mechanical collision, leading to methyl group loss and hydrophilic group formation; However, nitric acid corrodes Si-O main chain through hydrolysis reaction and forms $\text{Al}(\text{NO}_3)_3$. (2) Hydrophobicity loss showed phased characteristics: After a short time corona (<4h), the hydrophobicity quickly recovered due to the migration of small molecule siloxane; After long exposure (>7h), the inorganic crystal layer obstructs the migration and the recovery rate drops sharply. (3) The variation of trap characteristics shows that the charge-trapping ability increases with aging (TSC peak current increases from 26 pA to 165 pA), and nitric acid corrosion introduces shallow traps to form a double-peak current curve. (4) Under low pressure and high humidity environment, electric field distortion significantly promotes the formation and diffusion of nitric acid, and accelerates the crack propagation on the surface of the material; Under high pressure, charged particle bombardment is the dominant damage mode. The research results reveal the competitive mechanism of molecular chain breaking-oxidation-condensation reaction of silicone rubber under multi-factor coupling, and provide a theoretical basis for the molecular design of weather-resistant composite insulating materials.

Keywords: Corona Discharge; Organic Molecular Damage Mechanism; Hydrophobicity Recovery; Surface Topography Analysis; Environmental Dependence

Published: Apr 15, 2025

DOI: <https://doi.org/10.62177/jaet.v2i2.216>

1.Introduction

The aging behavior of silicone rubber (HTV-SR) in high voltage environment is significantly affected by the synergistic

effect of corona discharge and environmental factors. The research shows that humidity and ultraviolet can synergistically aggravate the damage of corona discharge to the molecular structure of materials. Bi et al.^[1] found that the corona discharge energy was significantly improved in high humidity environment, and accelerated the formation of hydroxyl (-OH) and carboxyl (-COOH) on the surface, leading to an increase in hydrophobic loss rate. Hakami et al.^[2] further confirmed that the coupling of ultraviolet light and high humidity doubled the aging rate of silicone rubber, but did not consider the complex interaction between temperature difference and air pressure in actual working conditions. Corona discharge directly destroys the molecular chain of silicone rubber by high-energy particle bombardment. Wang et al.^[3] research shows that the contact angle of the material decreases obviously after AC corona exposure, and the decrease of Si-O-Si and Si-CH₃ bond content is the main reason for the performance degradation. Haji et al.^[4] and Zhu et al.^[5] reveal the chemical mechanism of CH₃ group being replaced by hydrophilic OH group and the law of surface crack propagation through surface analysis. In the aspect of interface discharge behavior, Guo et al.^[6] confirmed that AC corona treatment can prolong the breakdown time of polypropylene/silicone rubber interface by three times, while Sarathi et al.^[7] found that corona aging materials were damaged more seriously in water droplet induced discharge, and the damage degree was significantly improved under negative DC voltage. The recovery of hydrophobicity depends on the migration of low molecular weight polydimethylsiloxane (LMW PDMS). Zhu et al.^[8-9] pointed out that the inhibition of pollution layer on the migration of LMW PDMS could delay the recovery time of hydrophobicity, while Zhang et al.^[10] found that the synergy of electric field and nitric acid increased the hardness and tensile strength of silicone rubber. However, most of the existing studies are based on idealized experimental conditions, and lack of in-depth analysis of the coupling mechanism between long-term aging data and multiple factors. Therefore, this paper studies the influence mechanism of the synergistic effect of corona discharge and nitric acid corrosion on the aging and degradation of silicone rubber composite external insulation materials in heavily polluted environment. Through multi-parameter regulation (corona intensity, nitric acid concentration, pressure and humidity), the internal relationship between the dynamics of organic molecular bond fracture, the cross-scale evolution of surface structure, the characteristics of electron traps and the dynamic response of hydrophobicity is revealed, which provides theoretical support for the design and aging evaluation of high-reliability external insulation materials.

2. Methods and results

Scholars both domestically and internationally have conducted extensive research on the aging characteristics of silicone rubber materials under corona effects, analyzing the mechanisms by which high-energy particles from corona and corona-generated substances—ozone and ultraviolet light—break down organic molecular bonds in silicone rubber materials. However, nitric acid is also a major product of corona discharge, and its mechanism for damaging organic molecular bonds in silicone rubber materials requires further investigation. Additionally, the sudden loss and recovery of hydrophobic properties during operation, as well as the reduction in flashover voltage of silicone rubber materials after aging, have not been adequately explained. Therefore, this section focuses on studying the mechanisms by which corona damages organic molecules in silicone rubber materials and provides an in-depth discussion of their impact based on various characterization methods, to supplement existing theories of corona-induced aging.

2.1 Study on the mechanism of corona damage to organic molecules of silicone rubber materials

Scholars both domestically and internationally have conducted extensive research on the aging characteristics of silicone rubber materials under corona effects, analyzing the mechanisms by which high-energy particles from corona and corona-generated substances—ozone and ultraviolet light—break down organic molecular bonds in silicone rubber materials. However, nitric acid is also a major product of corona discharge, and its mechanism for damaging organic molecular bonds in silicone rubber materials requires further investigation. Additionally, the sudden loss and recovery of hydrophobic properties during operation, as well as the reduction in flashover voltage of silicone rubber materials after aging, have not been adequately explained. Therefore, this section focuses on studying the mechanisms by which corona damages organic molecules in silicone rubber materials and provides an in-depth discussion of their impact based on various characterization methods, to supplement existing theories of corona-induced aging.

2.1.1 Corona aging test system construction

The composite material corona aging test system developed by the project team can simulate different corona aging conditions and carry out corona aging for multiple tests at the same time. It mainly includes four parts:

(1) Corona generation device

The corona generation device consists of two parts: the pressurization unit and the electrode system. The pressurization unit uses the YDT-5/50 type corona-free test transformer produced by Jiangdu Huayu High Voltage Electrical Co., Ltd. High voltage is transmitted through relays and conductive paper, ultimately reaching the stainless steel electrodes in the test chamber via cables. The electrodes adopt a traditional needle-plate configuration to simulate corona discharge in real-world conditions. The needle electrode serves as the upper electrode with a curvature radius of 0.2 mm, while the plate electrode acts as the lower electrode with a diameter of 30 mm. The entire electrode system comprises 54 sets of needle-plate electrodes, with the needle and plate electrodes arranged on upper and lower epoxy glass cloth plates connected by insulating rods, and each set of needle-plate electrodes facing each other; the distance between needles or plates on the insulating boards is 90 mm, and through electric field simulation, the influence of the electric field strength between each set of electrodes is minimal. Additionally, the upper and lower plates can be moved for easy sampling and adjustment of corona intensity. The details are shown in Figure 1-2.

Figure 1 Overall device diagram

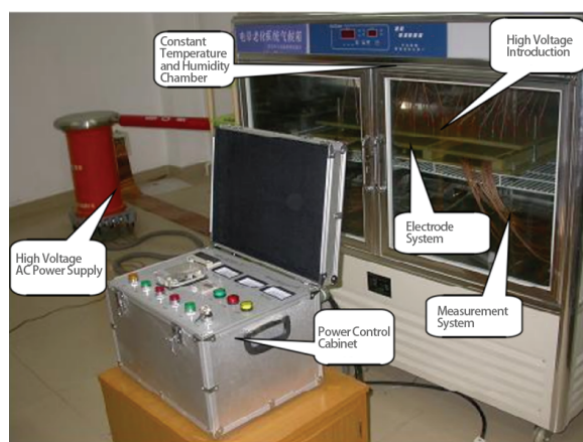


Figure 2 Electrode system



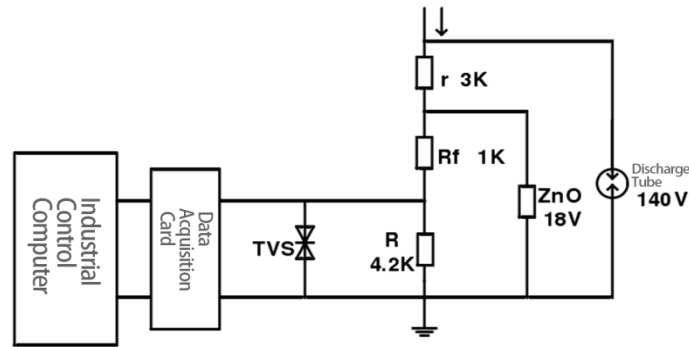
(2) Measurement system

(A) Hardware design and implementation

The corona current signal is introduced into the measurement system composed of protection circuit, measuring resistor and acquisition card through double shielded low noise cable (all interfaces are connected by BNC). The sampling resistor R is a high precision non-inductive resistor of $5k\Omega$, R is the protection resistor. The selection of the acquisition card mainly considers several aspects such as data acquisition channels, sampling rate, bit depth of the acquisition card, digital I/O, and price. The corona aging device designed in this paper has a total of 54 current loops. When all are running, an acquisition card with at least 55 channels is required (one channel is used to collect corona voltage signals). Using an oscilloscope to measure the corona discharge current in real-time at different sampling rates, it was found that when the sampling rate is set

to 10 kS/s, a clear corona current signal can be observed. Therefore, the sampling rate for each channel should be at least 10 kS/s. Considering that the accuracy requirement for measuring corona current is not high, a 12-bit acquisition card can be chosen. The selection of digital I/O ports should be based on the ports to be controlled. For this experiment, the DAQ2204 data acquisition card from Advantech was selected. This card has a PCI bus structure, with specific performance as follows: maximum sampling rate of 3 MS/s; 64 input/output channels, when using 55 channels simultaneously, the average sampling rate per channel is 55 kS/s; bit depth of 12 bits; voltage range for input signals set to $\pm 5V$; resolution of the acquisition card is 2 mV; and it has 24 digital I/O ports. The details are shown in Figure 3.

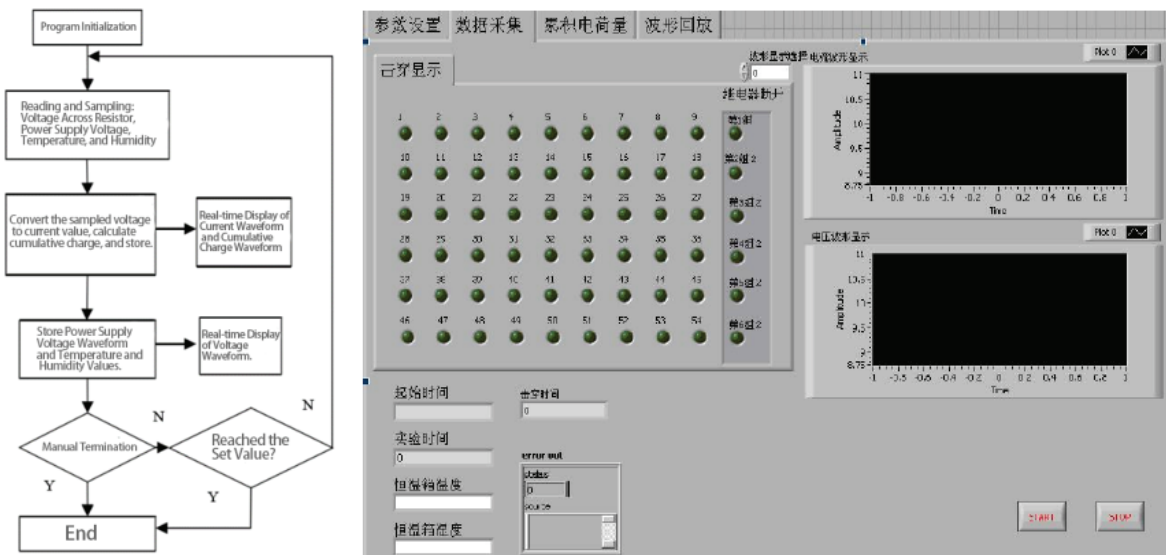
Figure 3 Schematic diagram of the measurement system



(b) Design and implementation of software system

Using LabView software and its DSC module, a software system for corona aging testing has been developed. This system can achieve the following functions: it can simultaneously collect and display measurement results of multiple corona current signals; it can display corona voltage waveforms; it can calculate and save the cumulative charge and total charge of both positive and negative half-cycles of corona discharge, as well as discharge power and energy; it can display charts showing the charge quantities and discharge energy trends over time for both positive and negative half-cycles of corona discharge current. The details are shown in Figure 4.

Figure 4 Main program flow chart Figure 42 Program front panel diagram



(3) Protection system

(A) High voltage side protection

The protection of the high-voltage side is tested using conductive paper connected between the high-voltage relay and the upper electrode. When a large current flows through a circuit, it causes the conductive paper to heat up and burn out within a certain period, thus breaking the test circuit. This ensures that the high-voltage source protection does not activate, allowing other circuits to be tested normally. Tests have shown that under this connection method, when 10mA of current flows

through the conductive paper, it will quickly burn out within 1 second, providing effective protection.

(b) Low voltage side protection

The low-voltage side protection mainly refers to the protection of the acquisition card in the measurement system, which adopts a three-level protection system composed of transient voltage suppression diodes, varistors and discharge tubes.

(4) Environmental control and monitoring

The whole experiment is carried out in a constant temperature and humidity chamber. The test chamber can realize the setting and control of temperature and humidity, and has the function of power failure memory and automatic compensation of power failure time, using air duct ventilation. At the same time, two external control fans are installed on the back side of the test chamber.

2.1.2 Study on the mechanism of damage of organic molecules in silicone rubber materials by corona discharge particles and nitric acid

(1) The influence of corona discharge time and intensity on the properties of silicone rubber materials

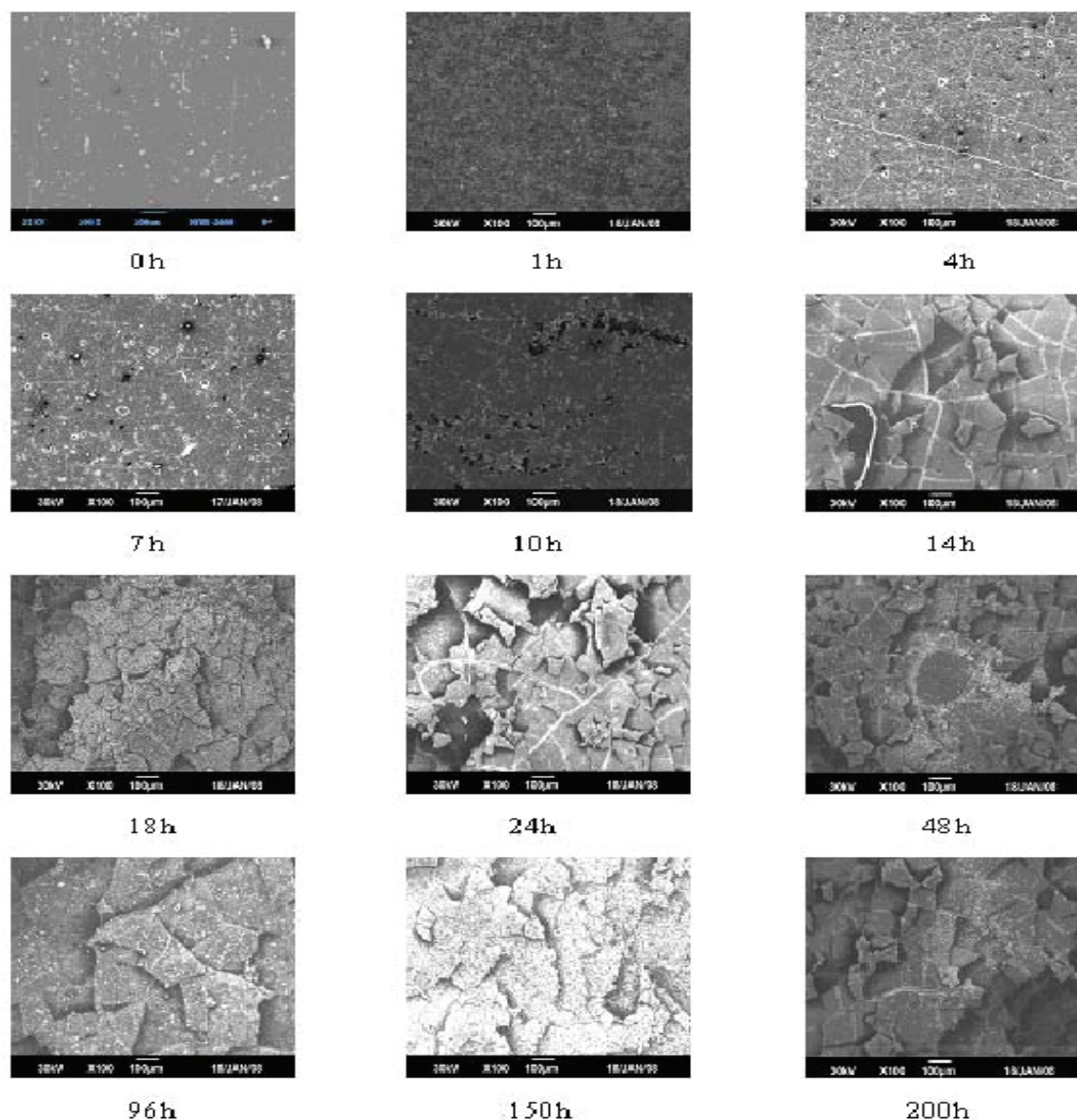
(A) The influence of corona time on the properties of silicone rubber materials

① Influence of voltage action time on surface morphology

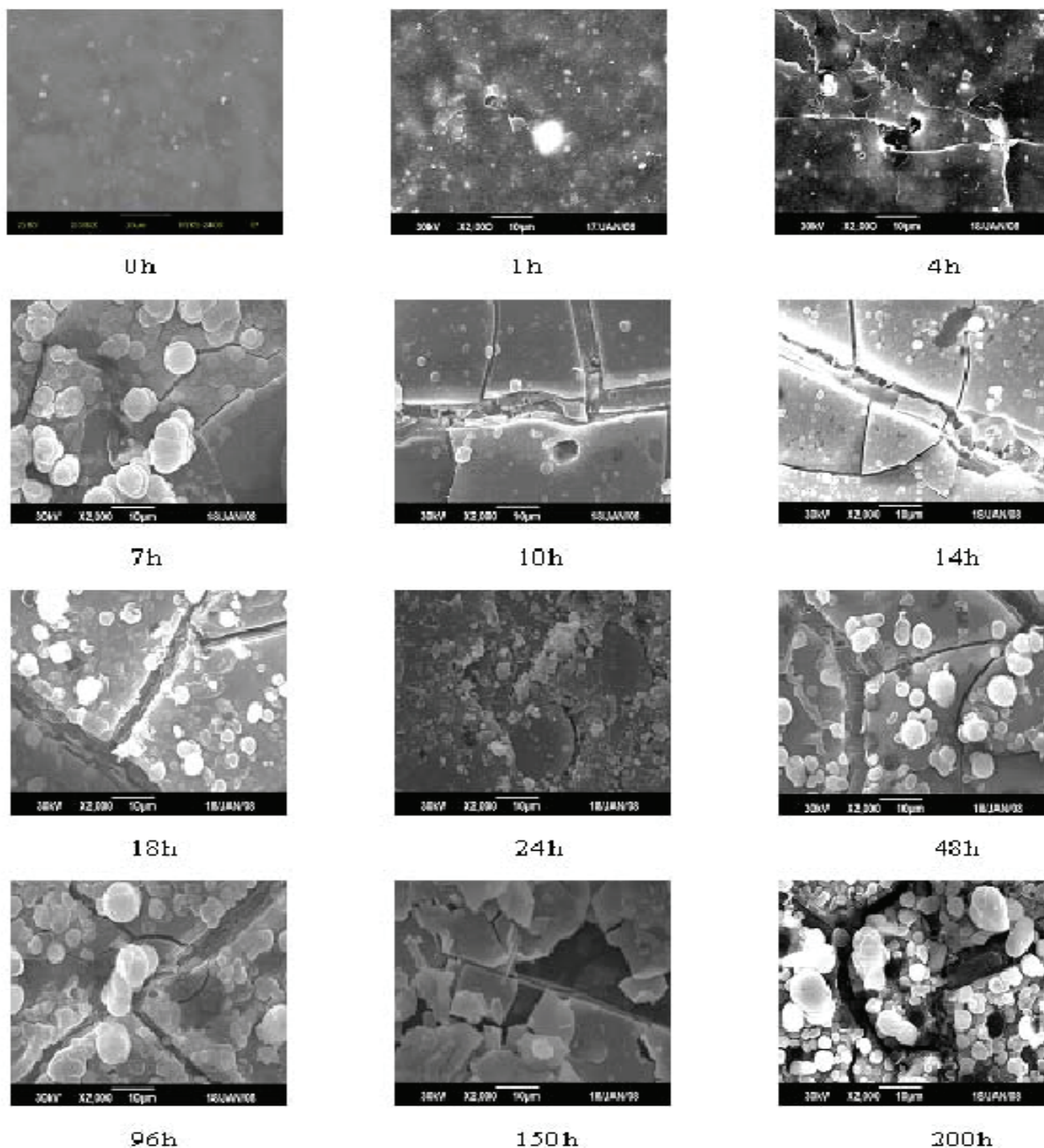
The SEM results of samples after corona at different times under 5kV voltage are shown in Figure 5.

Figure 5 SEM comparison

(A) The magnification is 100



(b) The magnification is 2000

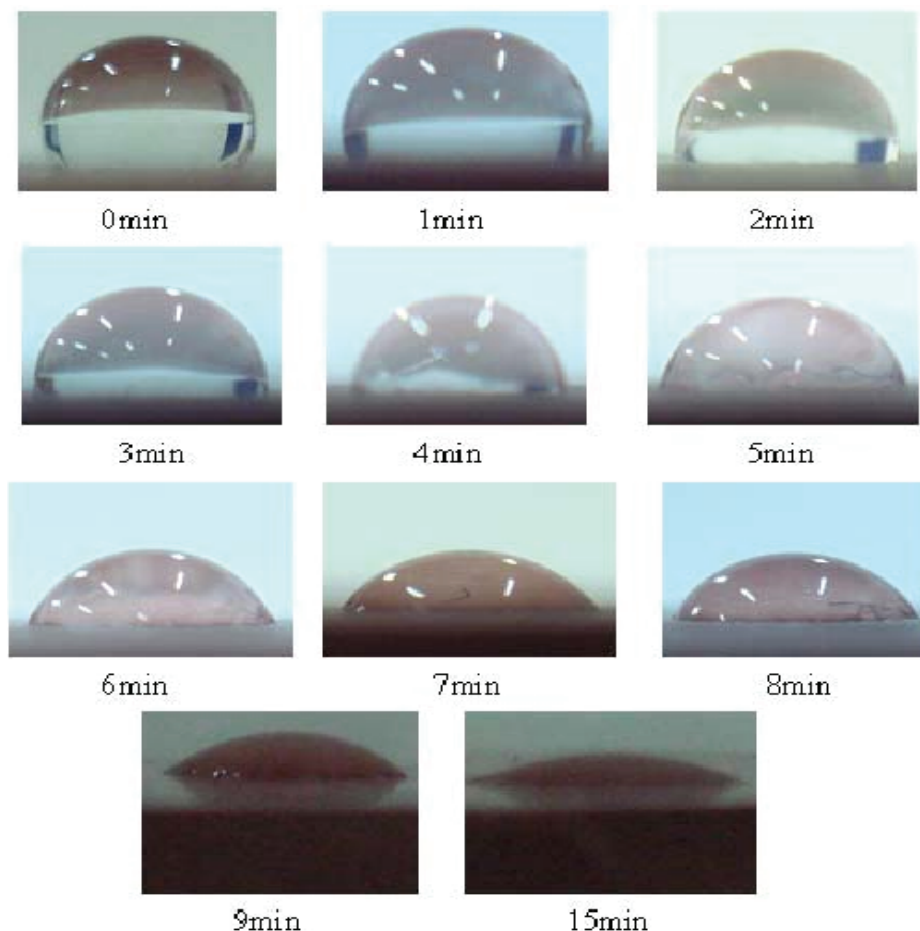


As can be seen from Figure (b), the sample surface becomes dirty after 1h of corona treatment; at 4 h, the surface pores are clearly visible; at 7 h, a thin layer of crystals can be observed on the sample surface; at 10 h, the dense crystal layer breaks, and faint cracks become visible, which gradually deepen; at 24 h, the sample surface becomes fluffy, with faint cracks visible underneath; at 48 h, the cracks completely penetrate the fluffy layer; at 96 h, long and deep cracks appear on the sample surface; at 150 h, large and deep cavities can still be seen, which worsen until the corona discharge ends.

② The effect of corona time on hydrophobicity

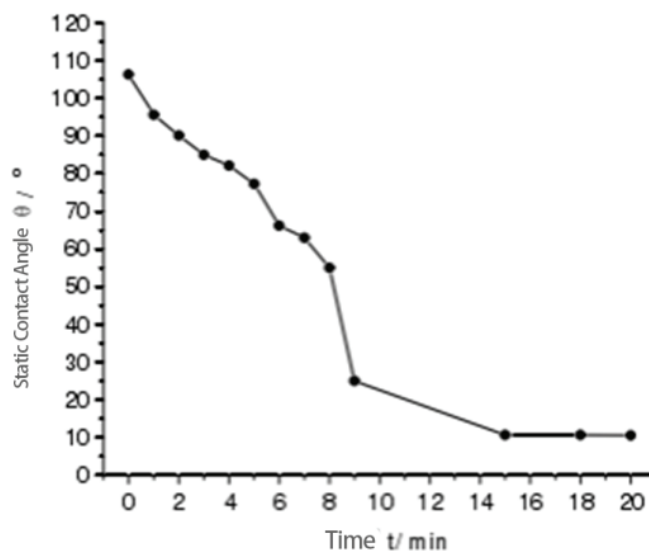
After corona at 5kV for 1h to 200h, the static contact angle of the sample was about 10°, and it temporarily lost its hydrophobic properties. In order to restore the whole process of loss of hydrophobic properties, the hydrophobic properties of the sample after short-time corona at 5kV were observed, as shown in Figure 6.

FIG. 6 Variation of hydrophobicity during corona testing



From the above figure, one can observe the entire process of hydrophobicity loss on the sample surface. At the same time, it is evident that under corona conditions, the hydrophobicity loss on the sample surface is a gradual process. The damage to the hydrophobic properties of silicone rubber material by corona discharge occurs relatively quickly; before pressure application, the static contact angle of the sample was 106.3° , which decreased to 10.5° over 15 minutes and remained around 10° until the end of the test. The details are shown in Figure 7.

Figure7 Static contact Angle of the sample at 5kV voltage with respect to the time of pressure application



At the same time, the hydrophobic recovery performance at different stages of the test was tested, as shown in the figure below. The details are shown in Figure 8-9.

Figure 8 Hydrophobic recovery performance of samples after corona 1h

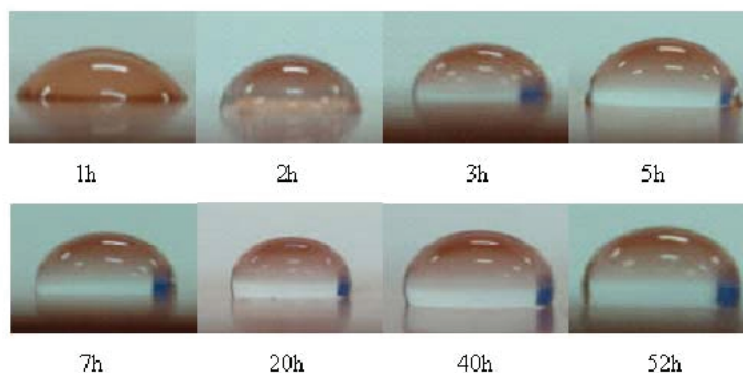
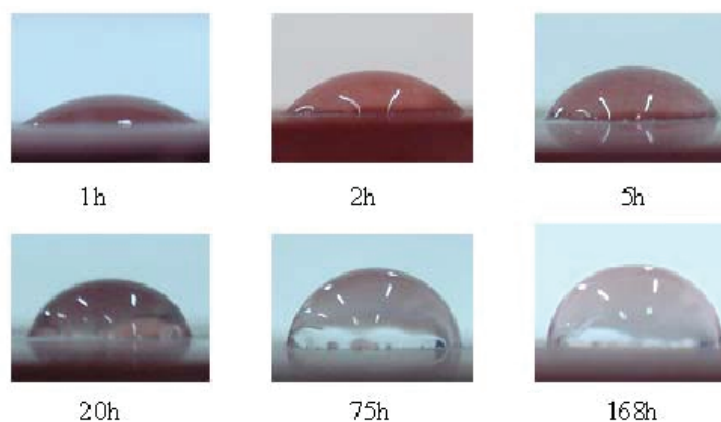
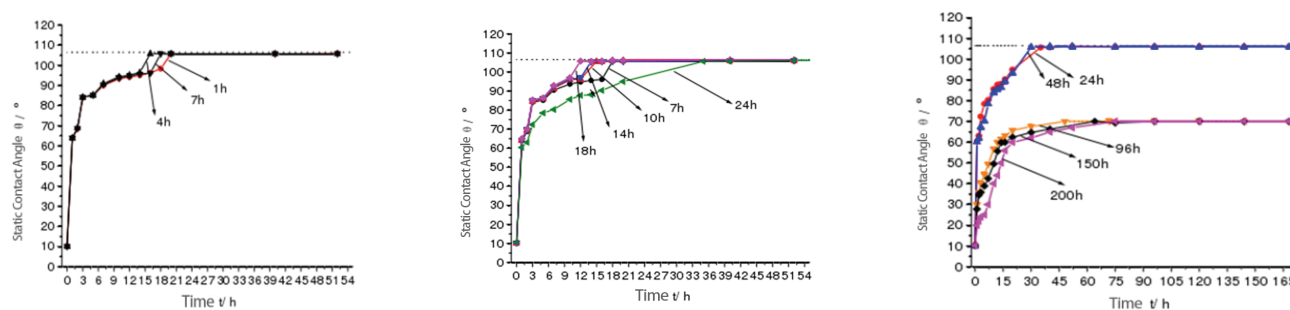


Figure 9 Hydrophobic recovery performance of samples after corona 200h



As can be seen from the figure, after 1h corona, when the sample is left standing for 20h, the hydrophobicity of the sample surface is basically completely restored; after 2000h corona, when the sample is left standing for 20h, its hydrophobicity does not return to the initial level, and then the recovery rate is slow until 75h, when the hydrophobicity no longer changes.

Figure 10 Recovery of static contact Angle of the sample



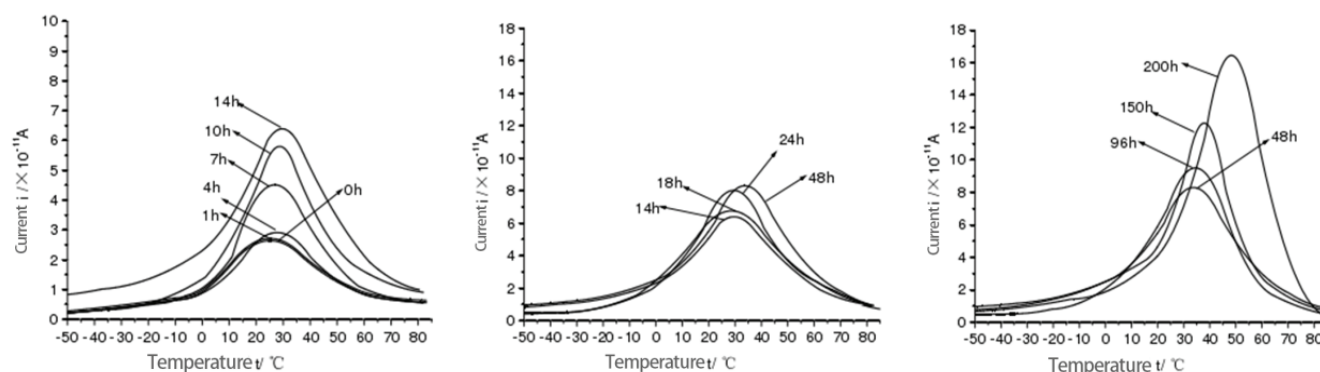
It can be observed that when the corona time is less than 4 hours, the recovery of the static contact angle of the sample accelerates with increasing corona time. When the corona time reaches 7 hours, the recovery of the static contact angle suddenly slows down but is faster than that of the sample after 1 hour of corona treatment; after the corona time exceeds 7 hours, the recovery of the static contact angle gradually accelerates with increasing corona time until it reaches 24 hours when the recovery rate slows down again, and by 35 hours, it returns to the initial level; compared to the sample after 24 hours of corona treatment, the sample after 48 hours of corona treatment recovers its static contact angle to the initial level earlier. After 96 hours of corona treatment, the static contact angle of the sample recovers to nearly 70° after being left undisturbed for 45 hours and remains at 70° thereafter; the static contact angles of the samples after 150 hours and 200 hours of corona treatment also remain at this level. Although short-term corona treatment damages the surface material of the silicone rubber sample and consumes hydrophobic substances within the material, according to previous research findings, the hydrophobic

substances in the inner layer of the material will accelerate their diffusion to the surface, restoring its hydrophobic properties. As the corona time increases, excessive consumption of hydrophobic substances prevents them from reaching the initial level of hydrophobicity.

③ The influence of corona time on trap characteristics

To more clearly demonstrate the changes in the sample during the corona aging process, the TSC curves of the sample after corona at different times under a voltage of 5kV are shown in the following figure. It can be seen that the TSC curves of the sample at different corona times all exhibit an independent TSC peak. The current peak in the TSC curve before corona is located around 26.5°C, with a peak value of about 26 pA; the TSC curve after 1h of corona is almost identical to that of the uncorroded sample, with only a slight increase in the current peak; after 4h of corona, the TSC peak shifts towards higher temperatures and becomes larger, followed by a slight increase in the current peak, further shifting the TSC peak towards higher temperatures. The temperature corresponding to 200h of corona is the highest, approximately 52°C, and the TSC peak also gradually increases with the duration of corona, reaching its maximum value at 200 h, about 165 pA. The details are shown in Figure 11.

Figure. 11 TSC curve of the sample with respect to corona time



According to the TSC curve of the sample and relevant calculation methods, the trap parameters of the sample after corona for different times at 5kV voltage were obtained, as shown in the table below. It can be observed that the trap energy level and trap charge of the sample do not change significantly after 1h of corona. However, as the corona time increases, the trap energy level gradually deepens, and the trap charge generally shows an increasing trend. The details are shown in Table 1.

Table 1 Trap parameters of samples after different times of corona

Charged voltage action time t / h	peak point current $i / \times 10^{-11} A$	The temperature corresponding to the peak of the current $T / ^\circ C$	trapping level E / eV	Trap charge Q / nC
0	2.61	26.52	0.434	87.43
1	2.65	26.52	0.434	90.36
4	2.92	27.94	0.451	105.34
7	4.51	28.03	0.472	124.68
10	5.80	29.05	0.533	157.86
14	6.39	29.52	0.538	217.92
18	6.71	31.04	0.540	222.45
24	7.82	33.04	0.547	232.76
48	8.33	34.01	0.553	268.49
96	9.52	35.03	0.680	253.56
150	12.0	40.07	0.745	294.40
200	16.5	51.08	1.044	393.10

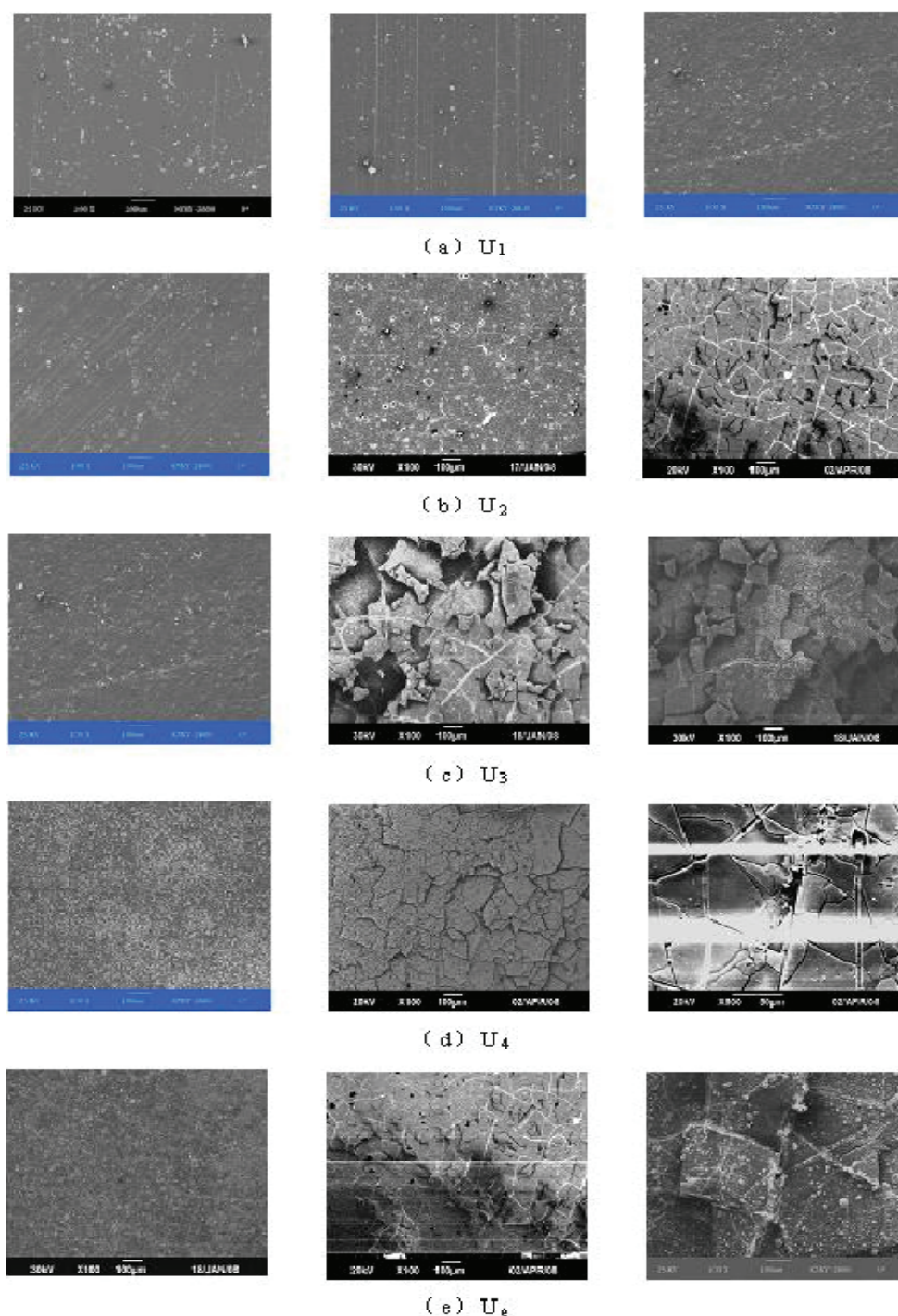
(b) The influence of corona intensity on the properties of silicone rubber materials

The corona is U_1 (2kV)、 U_2 (3kV)、 U_3 (5kV)、 U_4 (7kV)、 U_6 (Corona discharge was carried out on the sample at (10kV). After 1h,24h and 200h of corona discharge, the sample was taken out for relevant tests.

① The influence of voltage size on surface morphology

As shown in the figure below, after 1h of voltage action, the surface of the sample is still flat and the structure is relatively complete; after 24h of voltage action, U_2 Under the action of the sample surface is distributed with pinhole voids, U_3 The surface of the sample under the action of U_4 is covered with a fluffy layer, while the surface of the sample after the action of U_4 is full of cracks. After the action, large areas of missing were appeared, and cracks and cavities could be clearly seen; after 200h of voltage action, with the increase of the applied voltage, cracks appeared on the surface of the sample first, then evolved into cracks, until serious missing appeared in pieces. The details are shown in Figure 12.

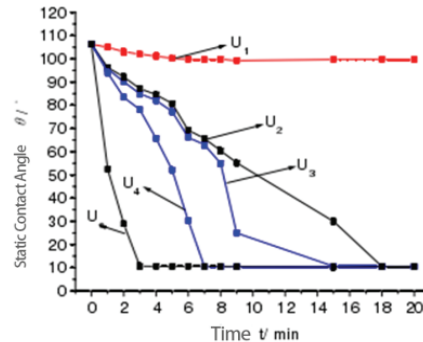
Figure 12. SEM comparison of samples under different voltages



② Influence of voltage on hydrophobicity

The hydrophobicity test results show that the sample is at corona $U_2 \sim U_c$. The static contact angle after 1h of corona treatment drops from about 106° before corona to around 10° ; however, even after 200h of corona treatment at voltage U_1 , the static contact angle remains above 90° . The hydrophobicity after short-term corona treatment at various voltage levels was tested, as shown in Figure 13.

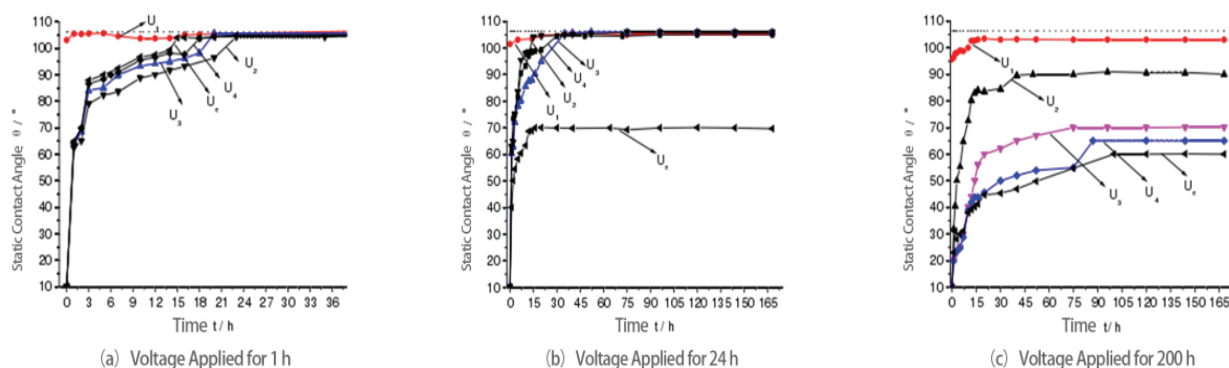
Figure 13. Static contact angle of samples at different voltages with respect to pressurization time



It can be seen that the voltage is U_1 . The static contact angle of the sample decreased slightly with the increase of corona action time, and remained above 100° even after 20min; from U_2 . At the beginning, with the increase of voltage, the static contact angle decreases faster and faster; when the voltage is U_2 . At this time, after about 18min, it drops to 10° ; when the corona is U_c . At that time, only 3min later, it was reduced to 10.3° .

The hydrophobicity recovery is shown in Figure 14. It can be observed that for samples tested after 1 hour of voltage application, as the applied voltage increases, the rate of static contact angle recovery gradually accelerates. After more than 20 hours, all samples recover to near their initial levels with little difference in recovery speed. For samples tested after 24 hours of voltage application, the higher the applied voltage, the slower the static contact angle recovery, but U_4 . The static contact angle after the action is restored to the initial level before the action of U_3 , U_c . After 24h of action, the static contact angle of the sample is difficult to recover to the initial level. When the sample is acted on by different voltages for 200h, the recovery rate of the static contact Angle decreases with the increase of the applied voltage, and when the voltage increases to U_3 . After that, the hydrophobicity decreased permanently, about 60° .

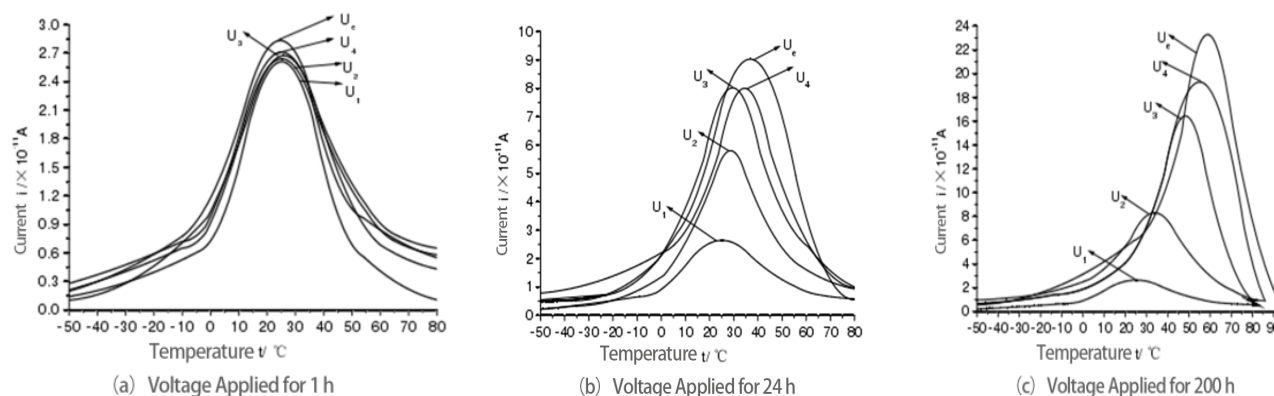
Figure 14. Static contact Angle recovery of samples under different voltages



③ The influence of voltage on trap characteristics

As shown in the figure below, after 1h of voltage action, the shape of TSC curve of the sample changes little, and only the current peak increases slightly with the increase of applied voltage. After 24h of voltage action, the higher the voltage, the greater the peak of TSC curve, and the temperature corresponding to the TSC peak increases until the voltage is increased to U_4 . After that, there is a more obvious increase. When the pressurization time reaches 200h, TSC peak and the corresponding temperature increase gradually with the increase of the applied voltage, and under the action of voltage U_c , TSC current peak is close to 240pA, and the corresponding temperature is located near 60°C .

Figure 15. TSC curve at different voltages



The trap parameters of the sample under different voltages are shown in the table below. It can be seen that after 1 hour of voltage application, as the applied voltage increases, the trap charge of the sample slightly increases, with little change in the trap energy levels; at 24 hours of voltage application, the trap charge of the sample increases with the rise in applied voltage, and the trap energy levels remain at the corona U_4 . After 200h of pressure, the trap charge and trap energy level of the sample increased significantly with the increase of the applied voltage. It can be seen that in different stages of aging, the trap parameters of the sample showed different characteristics with the change of the applied voltage. The details are shown in Table 2.

Table 2 Trap parameters of samples under different voltages

time t/h voltage V/kV	Trap energy level E/eV			The trap charge is Q / nC		
	1	24	200	1	24	200
U_1	0.434	0.434	0.434	87.43	87.44	87.44
U_2	0.433	0.483	0.553	88.92	157.86	268.49
U_3	0.434	0.547	1.044	90.36	232.76	393.10
U_4	0.434	0.651	1.056	91.20	236.63	528.64
U_e	0.435	0.678	1.101	94.12	277.21	625.45

(2) Influence of corona products-nitric acid on the properties of silicone rubber materials

Corona discharge in the air will form a certain amount of nitric acid, which is a strong oxidizer and a strong corrosive agent. Nitric acid may react with the surface material of silicone rubber, destroying its molecular structure, which is also an important factor causing corona aging.

(a) experimental method

① Determination of nitric acid concentration

The different corona discharge conditions can cause variations in nitric acid concentration. Therefore, to determine the test parameters, the project team used a corona aging test apparatus. By observing changes in pH paper and comparing it with a pH meter, they could obtain the concentration of nitric acid generated under these conditions. Through experimentation, it was found that the pH range for nitric acid should be set between 1.2 and 3.5.

② Determination of treatment parameters after nitric acid soaking

Before conducting the relevant aging property tests on samples soaked in nitric acid, it is necessary to remove the residual nitric acid from the surface of the samples, which requires post-treatment after nitric acid soaking. The method used is as follows: first, place the samples soaked in nitric acid in deionized water for a certain period to remove the nitric acid; then,

place the samples in a constant temperature chamber and dry them at the set temperature for a certain period to remove the influence of deionized water; finally, remove the samples for later use. Through experimental comparison, the final deionized water soaking time was determined to be 24 hours (based on Study 2.1.1, this duration results in minimal hydrolysis of silicone rubber molecules and does not significantly affect the test results), with deionized water being changed after soaking for 6 hours and 16 hours; the drying time in the oven is 14 hours.

③ Nitric acid immersion test

Nitrated solutions with pH values of 1.2, 2.0, 2.5, 3.0 and 3.5 were selected for nitric acid immersion test of silicone rubber. The soaking time was set to 4h, 24h, 96h, 120h and 180h respectively. In order to facilitate subsequent analysis, the samples were numbered as shown in the following table. The new samples without soaking were V.

Table 3. Sample numbers for nitric acid immersion test

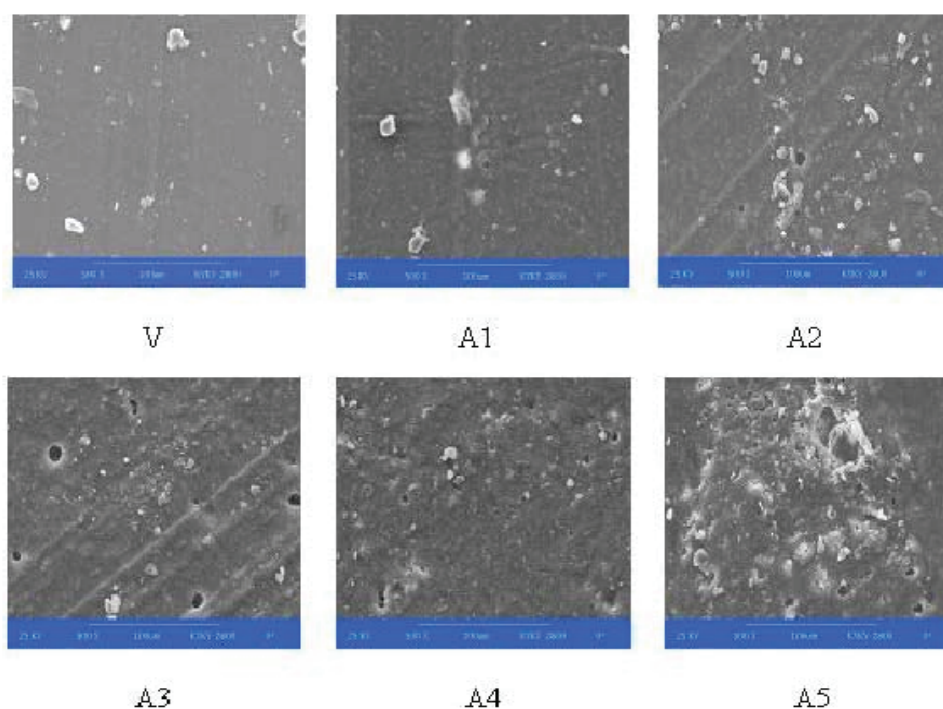
Soaking time t/h \ pH price	1.2	2.0	2.5	3.0	3.5
4	A1	B1	C1	D1	E1
24	A2	B2	C2	D2	E2
72	A3	B3	C3	D3	E3
120	A4	B4	C4	D4	E4
180	A5	B5	C5	D5	E5

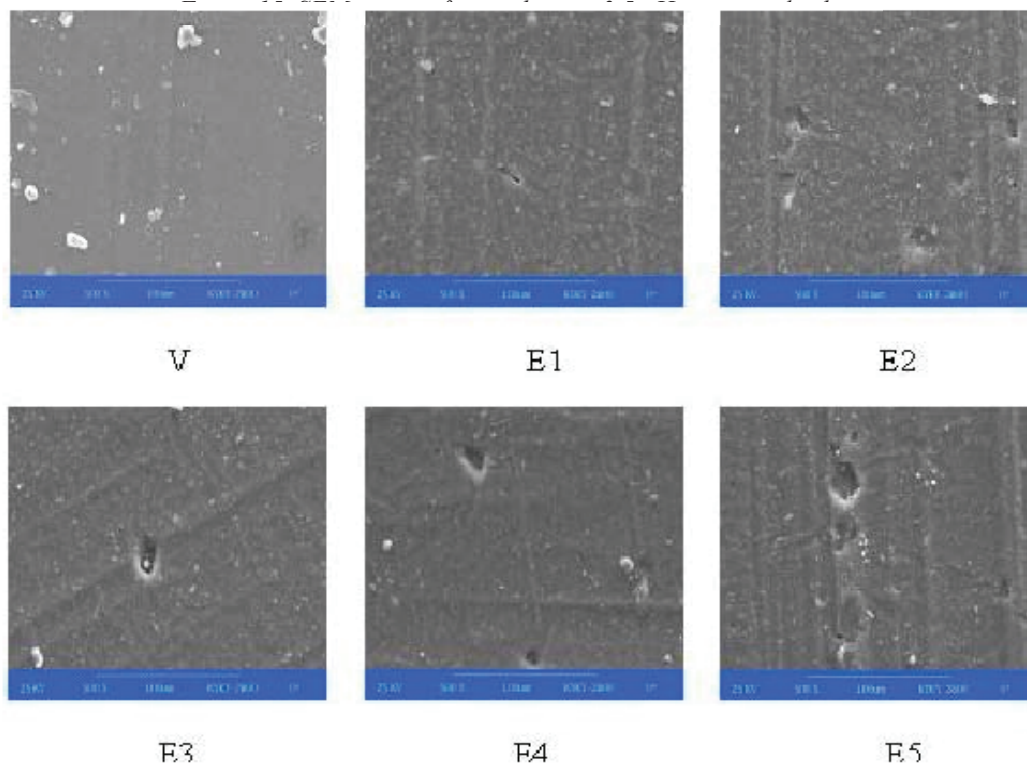
(A) Influence of nitric acid action time on the properties of silicone rubber materials

① SEM analyse

As shown in the figure below, after soaking in a nitric acid solution with pH=1.2, the integrity of the sample surface was compromised, and some holes appeared. Although their distribution was irregular, the density of these holes gradually increased with the duration of soaking. In particular, the surfaces of samples A4 and A5 were extremely dense with many small holes surrounding large ones, even forming clusters of interconnected holes. Additionally, as the soaking time increased, the depth of the holes also increased.

FIG. 54 SEM image of nitric acid soaked PH = 1.2





After soaking in nitric acid solution with pH=3.5, the structure of the sample surface showed different degrees of damage, and holes of different sizes appeared. With the increase of soaking time, the holes on the sample surface gradually became larger, but the number of holes changed little and the depth difference was not large, all of which were shallow holes.

At the same time, the surface state of samples measured at other pH values was basically consistent with the above results.

② FTIR analyse

As can be seen from the figure below, compared with the new sample V, several main infrared characteristic peaks of sample A5 after soaking in nitric acid solution with pH=1.2 for 180h all showed obvious changes, and the-OH content in silicone rubber molecules increased and-NO appeared₃group,CH₃The bending vibration was strengthened and the Si-O-Si structure was reduced. This indicated that after nitric acid immersion, the macromolecules of silicone rubber were damaged and oxidation and hydrolysis occurred. The changes of sample A series were most obvious, while those of sample E series were relatively gentle.

Figure 16. FTIR comparison before and after nitric acid immersion

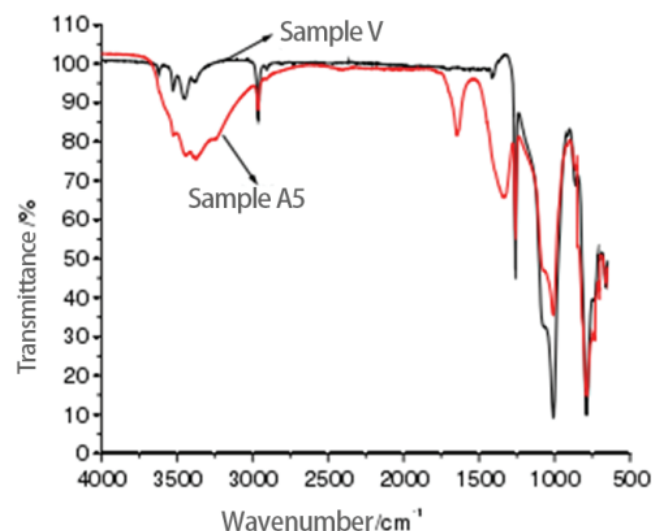
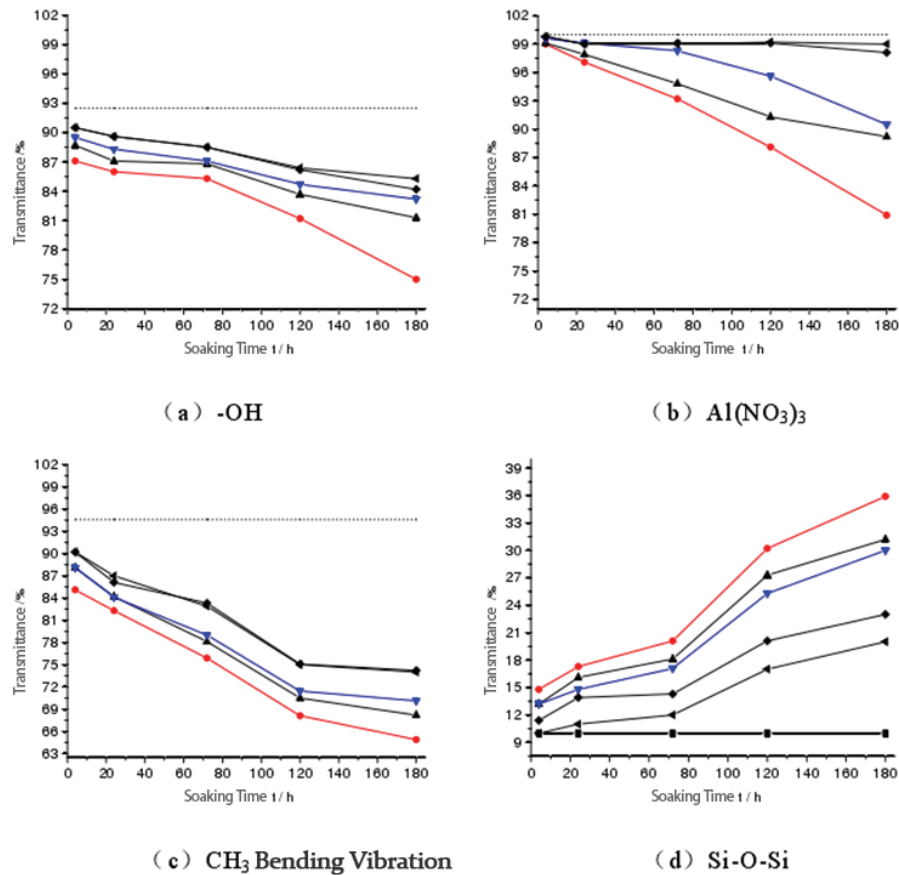
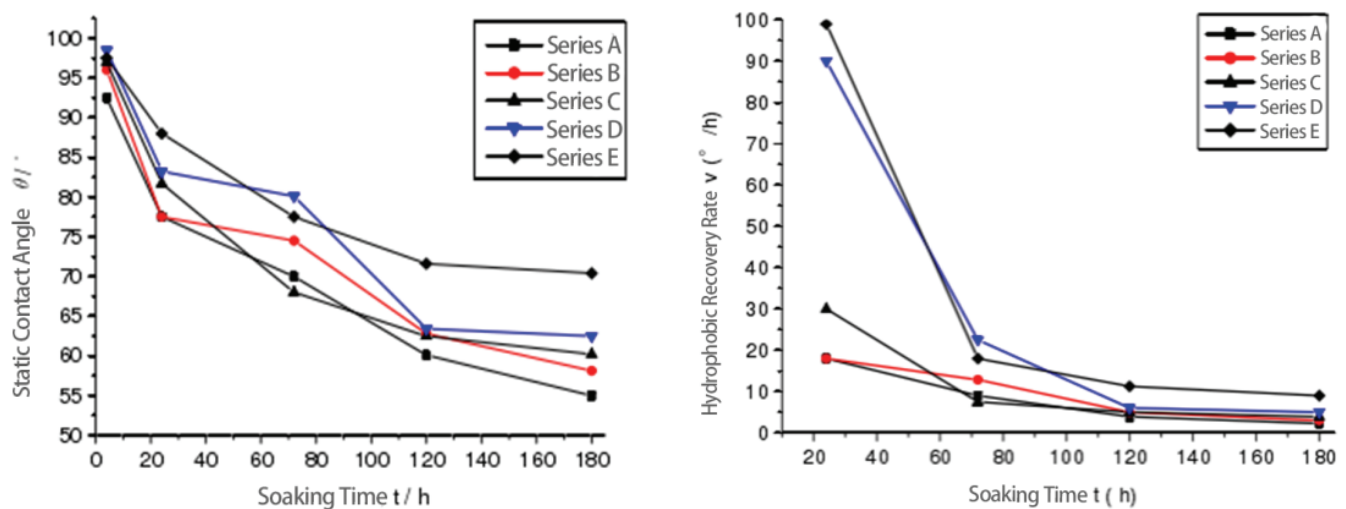


Figure 17. Variation of group content with soaking time



③ Hydrophobic performance analysis

Figure 18. Static contact Angle changes with immersion time FIG. 59 Water repellency recovery rate changes with immersion time



The static contact angle of various series of samples gradually decreases with increasing soaking time. When soaked for 4 hours, the static contact angle of all samples remains above 90° , indicating good hydrophobicity. After soaking for 120 hours, the contact angle stabilizes and changes more gently, reaching its minimum value ($55^\circ \sim 70^\circ$) at 180 hours, during which the hydrophobicity of the samples temporarily diminishes.

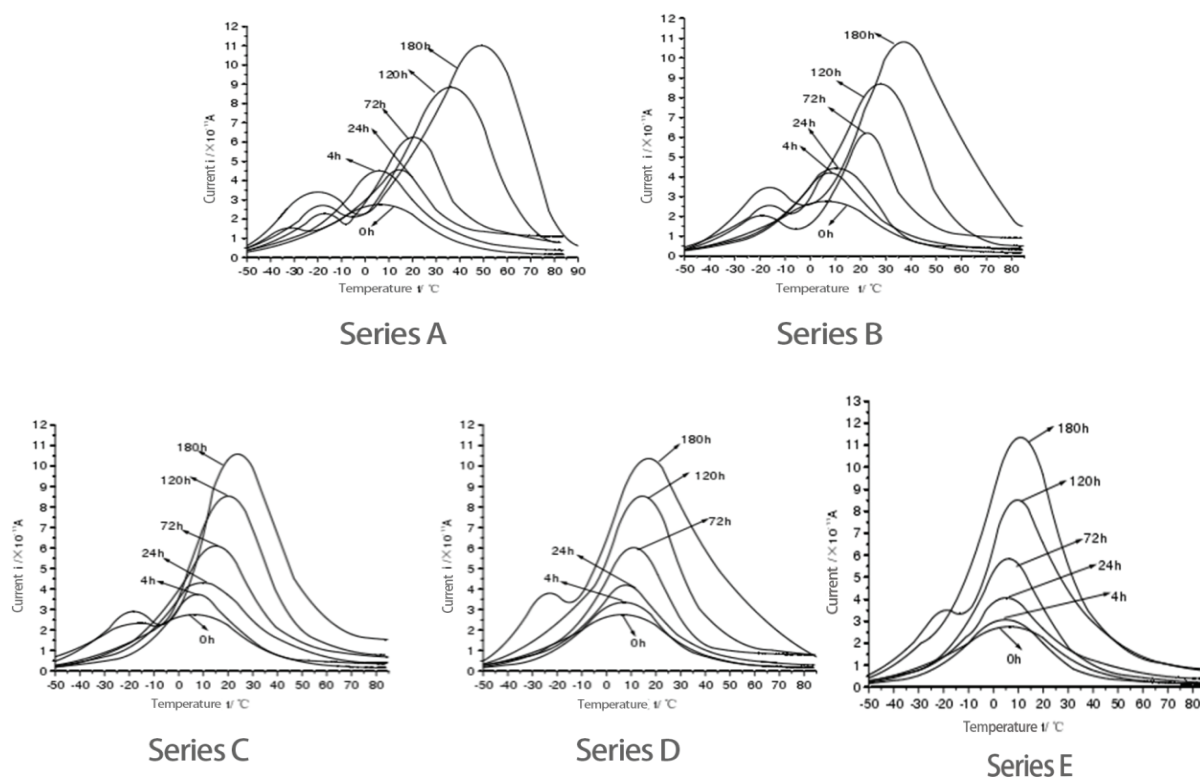
When the nitric acid immersion is stopped, the contact angle of all samples can be restored to more than 90° , and the recovery rate slows down with the increase of immersion time.

④ TSC analyse

As can be seen from the figure below, the TSC curves of samples in each series show changes with increasing soaking time:

- ① The shape of the curve changes, splitting from one current peak into two, one located in the low-temperature region (secondary current peak) and the other in the higher temperature region (primary current peak). At the same time, the appearance times of the secondary current peaks vary, distributed at 24h, 72h, 120h, 180h, and 120h;
- ② Both current peaks gradually increase;
- ③ The temperature corresponding to the primary current peak increases gradually with increasing soaking time, while the temperature corresponding to the secondary current peak remains relatively stable;
- ④ For samples in series E, the TSC curve shows little change in the temperature corresponding to the main current peak with increasing soaking time, whereas for sample A, the change is most pronounced.

Figure 19. TSC of the sample with respect to soaking time



(b) Influence of nitric acid concentration on aging characteristics of silicone rubber materials

The effect of nitric acid concentration on the aging characteristics of silicone rubber materials has been compared in the analysis of nitric acid solution action time, which will not be repeated here.

(3) Analysis of the mechanism of organic molecule damage to silicone rubber materials by corona discharge

Corona discharge produces high energy charged particles and nitric acid, which together cause the destruction of organic molecular bond energy of silicone rubber material.

(a) Analysis of charged particle action

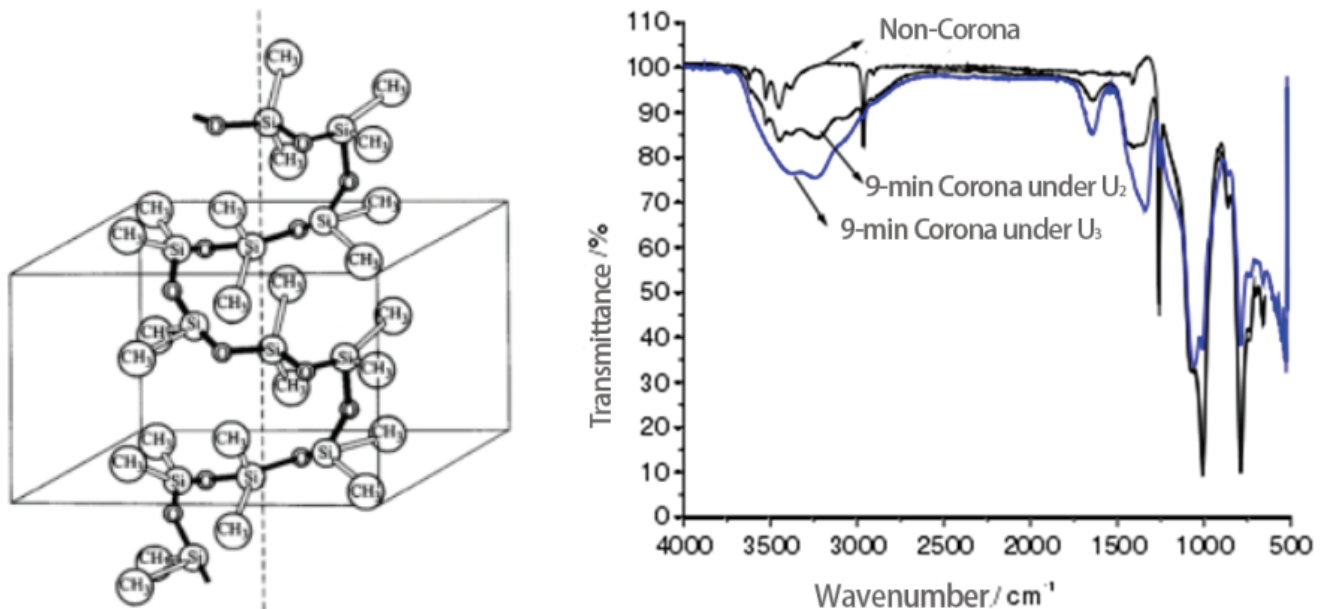
Corona discharge generates certain charged particles such as electrons and ions. Under the influence of an electric field formed by applied voltage, these particles, especially electrons, acquire kinetic energy and collide with the surface of insulating materials, causing mechanical damage to its surface. The fact that corona can drill holes in mica over several years indicates that corona is not merely a chemical corrosion process but also a mechanical impact-induced erosion, particularly at the edges of electrodes. The bond energy of Si-O bonds in the main chain of silicone rubber is approximately $461 \pm 33 \text{ kJ/mol}$, while the bond energy of Si-C bonds is 301 kJ/mol . The bond energy of C-H in the methyl groups of side chains is 413 kJ/mol . Estimating from the electric field strength during corona discharge and the average free path in air, the average energy of electrons is about 308 kJ/mol , with some reaching up to 963 kJ/mol . It is evident that particle bombardment in corona

discharge may cause all bonds in silicone rubber to break, thereby disrupting the structure of polymers within the material and leading to degradation.

① The mechanism of loss and recovery of hydrophobic properties

When corona discharge occurs, it generates a certain amount of charged particles. These particles collide with the sample surface under the influence of an electric field, causing mechanical impact. According to the spatial structure of the organic molecules in silicone rubber—polymethyl methacrylate siloxane—the charged particles first cause the Si-C bonds and C-H bonds in the organic molecules (containing hydrophobic small molecules) in the silicone rubber material to break and oxidize. The FTIR comparison before and after the test is shown in the figure below. After 9 minutes of corona treatment, several key characteristic peaks in the sample show significant changes: ① -OH groups are formed and increase with the rise in voltage; ② wavenumber at 1643cm^{-1} New absorption peaks appeared and increased with the increase of voltage, indicating that the higher the voltage, the more C-H bond was destroyed and the more carbonyl was generated; ③ wavenumber was 1412cm^{-1} The absorption peak becomes wider and stronger with the increase of voltage, indicating that the bending vibration of methyl group is strengthened; ④ wavenumber is 1007cm^{-1} The characteristic absorption peak weakened, indicating that the Si-O-Si structure decreased; ⑤ wavenumber was 787cm^{-1} The characteristic absorption peak decreased with the increase of voltage, indicating that $\text{Si}(\text{CH}_3)_2$ The structure gradually decreases.

Figure 20. Molecular space structure of silicone rubber Figure 62 FTIR comparison before and after corona of samples



It is evident that under the influence of charged particles, C-H bonds and Si-C bonds are broken, leading to oxidation reactions on the sample surface. This consumes methyl groups on the surface, weakening their shielding effect against the hydrophilic Si-O backbone. At the same time, chemical reactions generate silanol, silyl alcohol, carbonyl, and other hydrophilic groups, and the formation of oxygen-containing groups increases the surface energy of the sample. These factors collectively result in a reduction or temporary loss of hydrophobicity on the sample surface. As the applied voltage increases, the electric field strength on the sample surface also increases, enhancing the ability of charged particles to gain energy. This intensifies the chemical reactions on the sample surface, consuming more methyl groups and generating more hydrophilic groups. Consequently, as the applied voltage rises, the rate at which the hydrophobicity of the sample surface decreases accelerates, meaning the rate of decrease in the static contact angle increases. On the other hand, once corona discharge occurs, the longer the corona duration, the greater the probability that charged particles will continuously impact the sample surface, absorbing more energy and producing more hydrophilic by-products. Therefore, the hydrophobicity gradually diminishes with increasing corona duration until it is completely lost.

However, as the corona time increases or the applied voltage rises, more high-energy charged particles are generated. These charged particles break the Si-O main chain, producing more mobile siloxanes (i.e., small hydrophobic molecules), which

aids in restoring hydrophobicity. At the same time, due to the high reactivity of -OH groups and low activation energy during condensation, under discharge conditions, the -OH groups at the ends or on both sides of the main chain can easily condense to form H_2O . Condensation reactions, on one hand, reduce the concentration of -OH groups, which is beneficial for restoring hydrophobicity on the sample surface; on the other hand, similar to oxidation reactions, condensation reactions produce polysiloxanes with cyclic structures, double-chain structures, or side-chain structures in their main chains, making it difficult for siloxane molecules to migrate to the silicone rubber surface, thus hindering the restoration of hydrophobicity. Combining SEM results, the process of restoring hydrophobicity can be explained as follows: when corona time is less than 4 hours, as the corona time increases, charged particles collide with the sample surface, generating more small molecular siloxanes. The appearance of pinhole-like pores facilitates the diffusion or migration of siloxanes to the sample surface, thereby accelerating the restoration of hydrophobicity; however, after the corona time exceeds 4 hours, a thin layer of crystalline material gradually forms over the pores, indicating signs of hardening on the sample surface. White crystalline substances are observed at 7 hours of corona treatment, suggesting the formation of inorganic silicates. Although siloxanes continue to increase, the presence of inorganic silicates hinders the migration and condensation reactions of siloxanes, thus slowing down the restoration of hydrophobicity; when the corona time reaches 10 hours, apart from pores, faint micro-cracks are visible on the sample surface, indicating that continuous particle impact leads to an increase in dense crystalline layers, with cracks deepening and further accelerating the restoration of hydrophobicity; when the corona treatment reaches 24 hours, the sample surface appears to have become powdery, hindering the migration of siloxanes. Subsequently, the powder layer splits into large chunks (after corona at 48 h), thus the hydrophobicity recovery rate first slows down and then accelerates; as the corona time increases, pores on the sample surface gradually evolve into cavities, with cracks and grooves appearing (after corona at 96 h). This indicates that under the influence of charged particles, more severe oxidation and condensation reactions occur on the sample surface, producing polydimethylsiloxanes with main chains containing cyclic structures, double main chain structures, or side-chain structures. These structures make it difficult for siloxanes to migrate to the silicone rubber surface. Additionally, their structural flexibility is much poorer compared to monomeric double helix structured polydimethylsiloxanes, reducing the speed at which siloxanes migrate to the surface. As the molecular weight of the polydimethylsiloxane increases, intermolecular forces also increase, further slowing down the migration of siloxanes. Furthermore, after the corona aging process reaches a certain stage, the rate at which siloxanes form on the sample surface is lower than their consumption rate (including migration into the air), leading to a decrease in siloxane content. The release of free carbon also reduces the hydrophobicity of the surface. Therefore, the hydrophobicity recovery rate slows down again or even loses its hydrophobicity.

It can be seen that the hydrophobic recovery characteristics of silicone rubber show different characteristics at each stage, and do not accelerate or slow down with the increase of corona time. The recovery characteristics of hydrophobic properties under different corona voltages are similar to the above analysis, which will not be repeated here.

② Reasons for changes in trap characteristics

From the aforementioned measurement results, it can be observed that under a 5kV voltage, as the corona time increases, more traps appear in the sample, and the trap energy levels gradually deepen. In conjunction with SEM results, it is evident that under the continuous bombardment of charged particles generated by corona discharge, a series of new interface states have appeared on the sample surface. The specific analysis is as follows: before 1h of corona treatment, the sample surface was relatively smooth and flat, with a more uniform structure. It is visible that its structure did not undergo significant changes compared to the sample before corona treatment, so the trap charge and trap energy levels were very close. As the corona time increased, cracks gradually developed into fissures, eventually forming grooves. This indicates that the structure of the sample surface was progressively damaged under the continuous action of charged particles, leading to more defects. The formation of these defects requires the injected trap charges to absorb more energy to be excited, resulting in higher temperatures during excitation and deeper trap energy levels. This is reflected in the TSC curve as an increase in the peak-to-peak current and a corresponding rise in temperature.

Observe the TSC curve of the sample after a certain period under different voltages, combined with SEM results, it can be

seen that after 1h at each voltage level, the surface of the sample remains flat and smooth. This indicates that the short-term effect of charged particles has not caused significant changes in the surface structure of the sample, thus the trap charge and trap energy levels before and after voltage application are not significantly different; when the exposure time increases to 24h, the voltage U_1 The sample surface remains structurally uniform, so its trap parameters do not change much. However, as the voltage increases, the following changes gradually occur on the sample surface: micro-pores-powder layer cracks-fractures-cavities, cracks. The sample surface also undergoes these changes after being pressurized for 200 hours. It is evident that the higher the voltage, the greater the energy gained by charged particles, leading to more severe damage to the sample surface, thus causing more and more serious defects in the sample. The increase in defects results in an increased amount of trapped charge, and the charge captured by the defects must absorb more energy to escape. Therefore, the higher the temperature at which excitation occurs, the higher the temperature corresponding to the current peak on the TSC curve.

To verify the changes in silicon rubber materials ability to capture electrons and its impact on surface flashover after altering trap characteristics, a new set of experiments was redesigned. The new silicon rubber samples were aged at 10kV AC voltage for 1 00 h, 2 50 h, 4 50 h, 6 50 h, and 1 000 h, respectively. After aging, circular discs with a diameter of 20mm (unchanged thickness) were made at the center of the corona discharge position. A negative polarity voltage was applied between the two ends of the sample until flashover occurred, and the charge distribution was measured using an electrostatic probe. The results are shown in the figure below. It can be seen that after aging, multiple peaks of surface charge appear on the sample surface. As the aging time and test voltage increase, the surface charge on the sample increases and gradually concentrates at positions perpendicular to the electrodes. After flashover, the surface charge in the region of the path where flashover occurred rapidly decreases. According to band theory and surface state theory, combined with the results of trap characteristic tests, as the degree of aging increases, the materials ability to capture electrons strengthens, forming strong electric field centers of charge at different locations on the material, leading to a decrease in its flashover voltage.

Figure 21. Measurement range of electrostatic probe

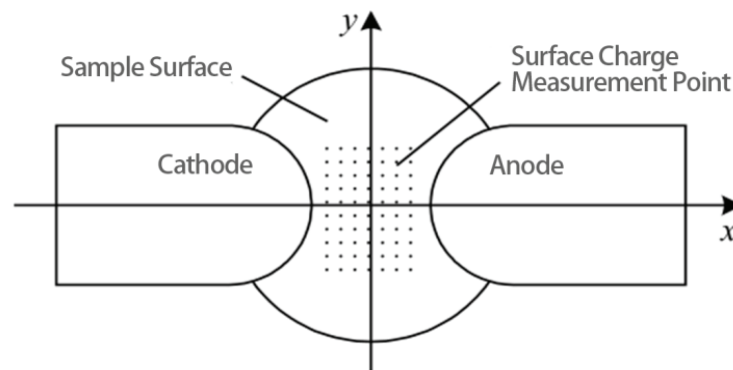
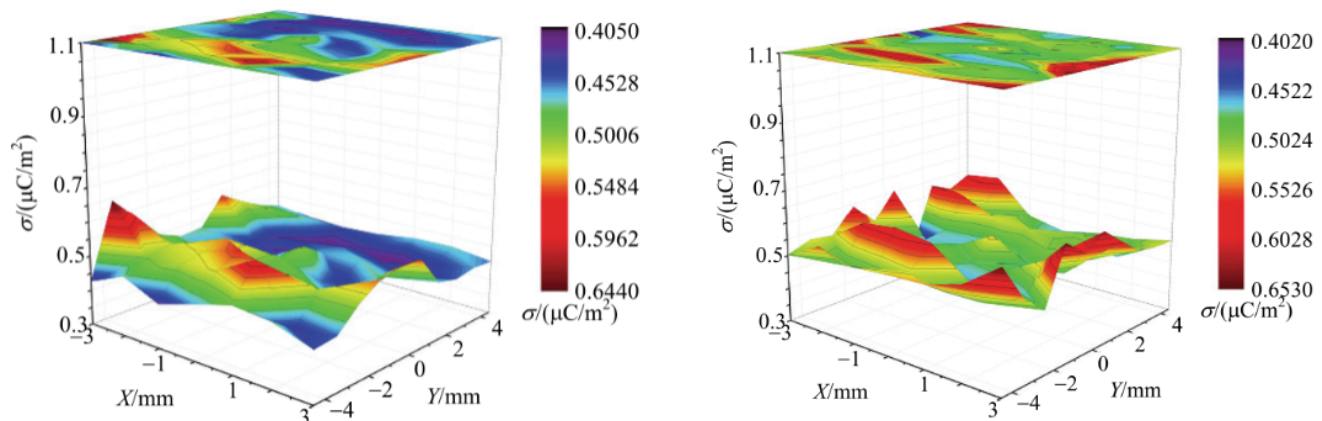


Figure 23. Surface charge distribution of samples aged by corona at 1 00h under different voltages

(a) 100 h

(b) 250h



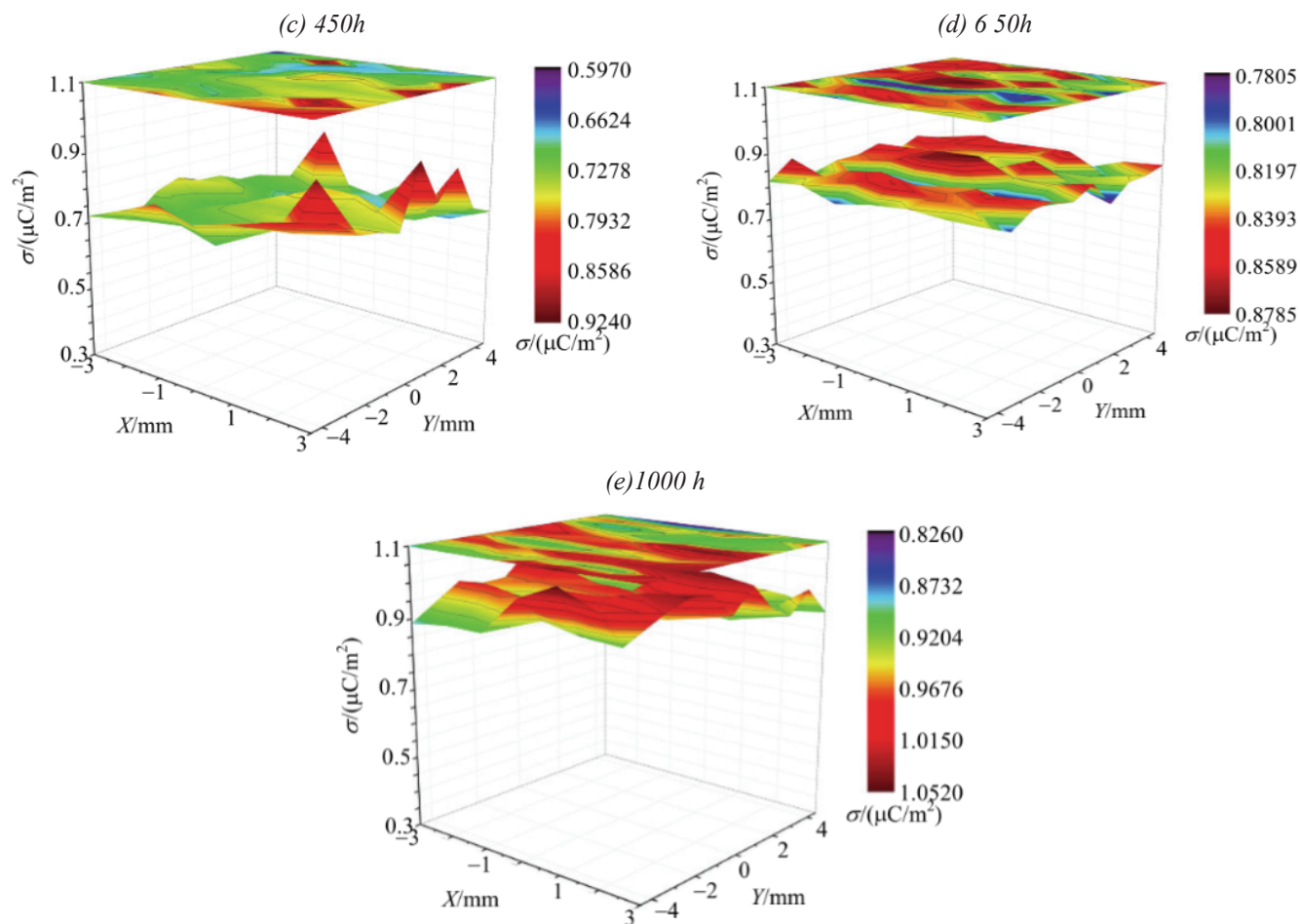
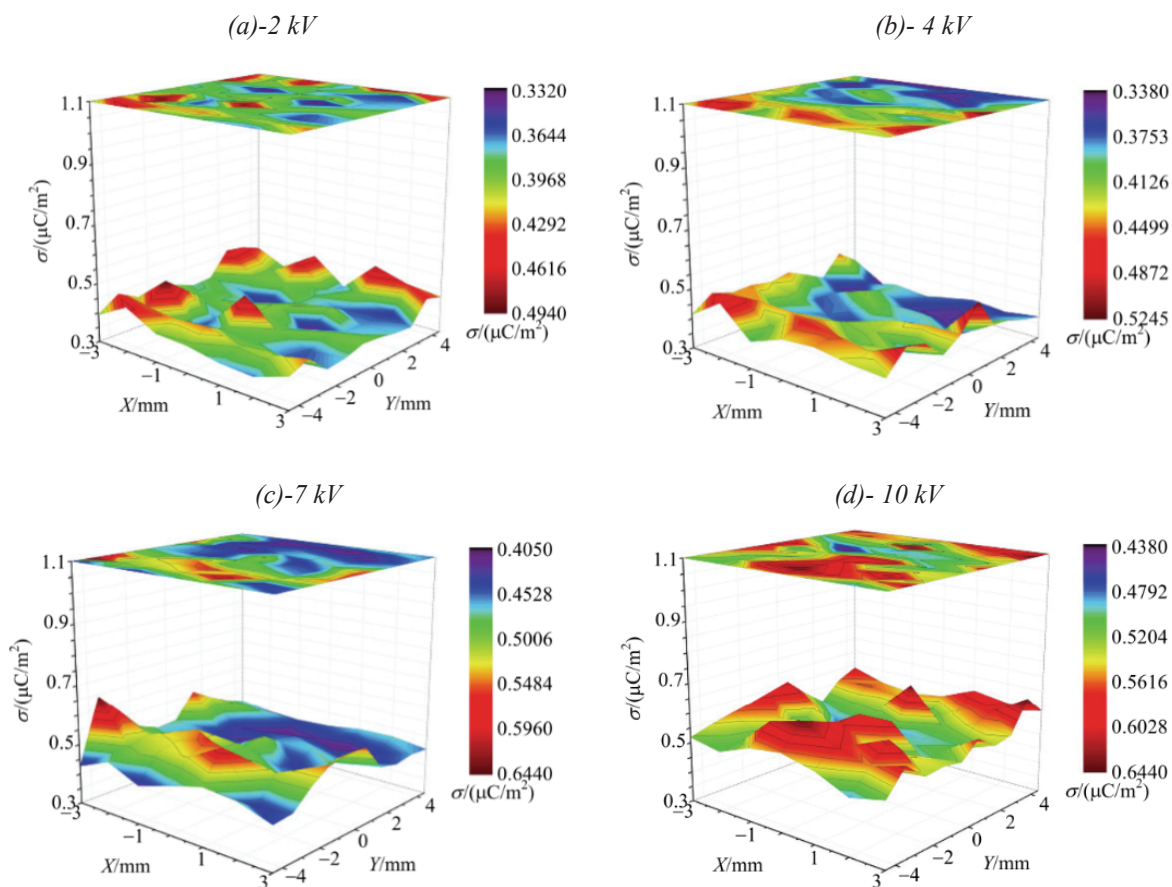
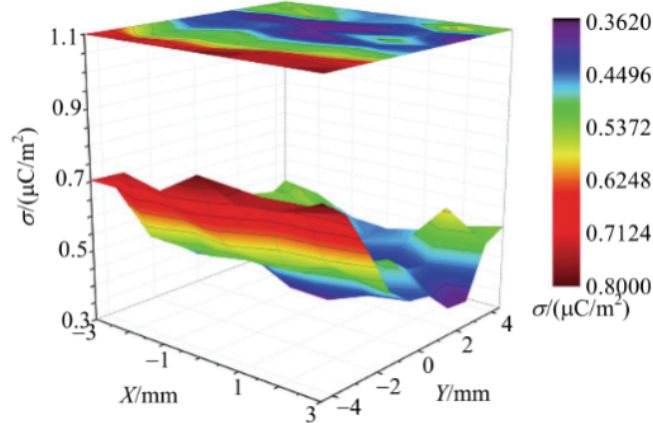


Figure 22. Surface charge distribution of samples at different aging times under 7kV



(e) 1 2.45kV (after flashover)



(b) Analysis of nitric acid action mechanism

① Causes of loss and recovery of hydrophobicity

In the main chain of silicone rubber, the electronegativity difference between silicon and oxygen atoms is 1.7, making the Si-O bond in silicone rubber polar or particulate to a degree of 41%. When silicone rubber comes into contact with high-concentration acidic or alkaline solutions, the ionic nature of the Si-O bond makes it susceptible to hydrolysis reactions. In this experiment, after nitric acid immersion, the -OH content increased, confirming that the hydrolysis reaction occurred, generating hydrophilic silanol groups. Moreover, CH was measured. The increased bending vibration indicates that nitric acid soaking causes the side groups of silicone rubber to curl inward, thus losing their shielding effect on the hydrophilic Si-O main chain. At the same time, the detection of aluminum nitrate formation suggests that nitric acid may have chemically reacted with the aluminum hydroxide filler in the silicone rubber, increasing the surface energy of the sample. SEM analysis results show that after soaking in a nitric acid solution, the smooth surface of the sample becomes rough and develops pores. These factors all contribute to making the hydrophobic silicone rubber more susceptible to moisture absorption.

The-OH, $\text{Al}(\text{NO}_3)_3$ The content gradually increases with the increase in soaking time, while the Si-O-Si structure decreases, especially noticeable in samples of series A and B. SEM analysis shows that the pores on the surface of samples from series A and B become significantly deeper as soaking time increases, with pore density gradually increasing. It is evident that as soaking time increases, both the physical and chemical structures of the samples change, not only producing more defects but also generating more hydrophilic groups. Therefore, the longer the soaking time, the more pronounced the decrease in hydrophobicity and the greater the reduction in static contact angle. On the other hand, prolonged soaking makes condensation reactions possible; when these two factors reach a certain equilibrium, the hydrophobicity of the sample remains at a certain level. At the same time, the longer the soaking time, the higher the CH_3 . The stronger the bending vibration, the weaker the hydrophobicity.

All samples contained -OH, $\text{Al}(\text{NO}_3)_3$ The content increases with the decrease of nitric acid pH value, and the Si-O-Si structure gradually decreases. Moreover, the smaller the nitric acid pH value is, the more CH in the sample. The stronger the bending vibration. At the same time, it is noted that after soaking for 180h, the changes in each group of samples are most pronounced, while those after soaking for 4h show smaller changes. SEM analysis results indicate that with longer soaking times (120h), as the pH value of nitric acid decreases, the pore density on the sample surface gradually increases, and the depth of pores also shows an increasing trend; when the soaking time is short, the pores on the sample surface are shallower, and there is no significant change in pore density. Therefore, the hydrophobicity of the samples gradually weakens or even disappears as the pH value of the nitric acid solution decreases, but when the soaking time is short, the change in hydrophobicity is minor, and when the soaking time is longer, the hydrophobicity significantly decreases as the pH value decreases.

When the action of nitric acid ceases, the methyl groups will reorient towards the surface. Meanwhile, hydrophilic groups such as silanol contain a large number of -OH groups, which, due to their high reactivity, can undergo condensation reactions to produce other products that cover the aluminum nitrate surface. These factors all contribute to the restoration of hydro-

phobicity on the silicone rubber surface. Additionally, condensation reactions may also produce polysiloxanes with cyclic structures, double-chain structures, or side-chain structures in their main chains, making it difficult for siloxanes to migrate to the silicone rubber surface, thus hindering the recovery of hydrophobicity. When these promoting and inhibiting effects reach equilibrium, the hydrophobicity of the silicone rubber remains in a stable state. In this experiment, after the action of nitric acid ceased, the methyl groups easily reoriented, representing the hydrophilic groups on the surface, and the reagents used were relatively short, so the hydrophobicity of the samples recovered quickly in the initial few hours after standing. Subsequently, due to the longer duration required for condensation reactions, the recovery process gradually slowed down. As the concentration of nitric acid increased, the time needed for condensation reactions also increased, leading to a longer time for methyl groups to flip back into the bulk phase, further extending the time required for hydrophobicity recovery, i.e., reducing the rate of hydrophobicity recovery.

② Reasons for changes in trap characteristics

According to the macroscopic morphology observation and SEM images of the samples, it can be seen that nitric acid soaking makes the silicone rubber surface rougher, more brittle, and prone to tearing, disrupting the complete structure of the silicone rubber surface and creating holes of varying sizes. These changes form structural defects in the silicone rubber. When charged particles are captured by these structural defects, they must absorb sufficient energy to break free from their constraints. Therefore, in TSC tests, trapped charges are only activated at higher temperatures, forming free charges. This is reflected in the TSC curve as higher temperatures corresponding to current peaks, indicating deep traps. The results of SEM analysis show that as the pH value of nitric acid decreases or the soaking time increases, the density of pores on the sample surface gradually increases, along with the depth of the pores. Thus, the energy level of deep traps in the sample deepens, and the amount of trap charge increases. When the soaking time is short, the changes in the structural defects on the sample surface due to increasing nitric acid concentration are minimal, so the trap energy level of the sample does not deepen with the decrease in nitric acid pH value after 4 hours of soaking. Similarly, for samples in Series E, the main trap energy level changes minimally with increasing soaking time.

FTIR analysis results show that nitric acid soaking leads to the formation of new chemical groups such as silanol and aluminum nitrate in the sample, reducing the original Si-O-Si structure. These processes introduce groups that differ from those in the original silicone rubber sample in terms of mass or electronic structure, creating chemical defects. Due to their higher electronegativity, these groups have a stronger ability to attract electrons, which in turn makes the trapped charges more easily excited. Therefore, in TSC tests, trapped charges can be excited and escape from their precarious state at lower temperatures. This is reflected in the TSC curve as lower temperatures corresponding to current peaks, indicating lower trap energy levels. FTIR analysis of the sample shows that with a decrease in nitric acid pH and an increase in soaking time, the -OH and $\text{Al}(\text{NO}_3)_3$ content gradually increases, while the Si-O-Si structure gradually decreases. Since only the chemical group content changes, the current amplitude in the low-temperature region of Sample TSC curve gradually increases, with little change in the corresponding temperature. At the same time, when the soaking time is short or the nitric acid concentration is low, the contents of silicic acid and aluminum nitrate are relatively low, so no current peak or an insignificant current peak can be observed in the low-temperature region during TSC testing.

From this, it can be seen that under the action of nitric acid, silicone rubber not only develops deep traps with higher energy levels but also forms shallower traps. The destruction of the physical structure on the sample surface and the appearance of pores are the primary causes of deep traps, while the formation of new chemical groups such as silanol is the main reason for the emergence of shallow traps. As the duration of nitric acid action increases or the pH value decreases, the trap charge in the sample shows an increasing trend, with the main trap energy level gradually deepening and the secondary trap energy level showing little change.

2.1.3 Study on the difference of corona damage mechanism to organic molecules of silicone rubber materials under different pressure and humidity

(1) Test method

To shorten the test cycle, orthogonal experimental methods are used. This involves selecting some points from comprehensive

tests that have “uniform distribution and comparable neatness” based on the orthogonality of each level number. This method can significantly reduce the test cycle and allows for the analysis of test results using appropriate range analysis and variance analysis methods to derive valuable conclusions. To study the differences in the mechanism of corona-induced damage to organic molecules in silicone rubber materials under different pressures and humidity levels, pressure, relative humidity, and aging time were determined as research factors. The test was conducted at three levels for these three factors, as shown in the table below.

Table 4 Factor level table

horizontal	factor		
	pressure /kPa	relative humidity /%	ageing time /h
1	4 5	2 5	1 00
2	6 0	5 5	4 00
3	1 00	9 0	8 00

Table 5 Orthogonal test arrangement

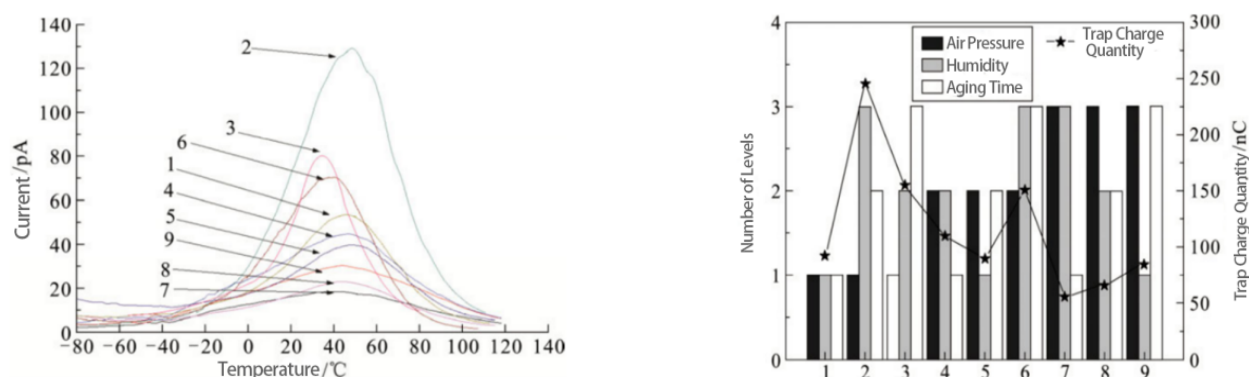
test number	condition of experiment			Air pressure levels	Relative humidity levels	Level of aging time
	pressure /kPa	relative humidity /%	ageing time /h			
1	4 5	2 5	1 00	1	1	1
2	4 5	9 0	4 00	1	3	2
3	4 5	5 5	8 00	1	2	3
4	6 0	5 5	1 00	2	2	1
5	6 0	2 5	4 00	2	1	2
6	6 0	9 0	8 00	2	3	3
7	1 00	9 0	1 00	3	3	1
8	1 00	5 5	4 00	3	2	2
9	1 00	2 5	8 00	3	1	3

(2) Test results

① TSC test result

The test results are shown in the figure below. It can be seen that the amplitude of TSC curve and the temperature corresponding to the peak current of each sample are different; at the same time, the higher the current amplitude, the narrower the width of the curve, and the shape of the curve in the low-temperature region also presents different development trends.

Figure 24. TSC test results

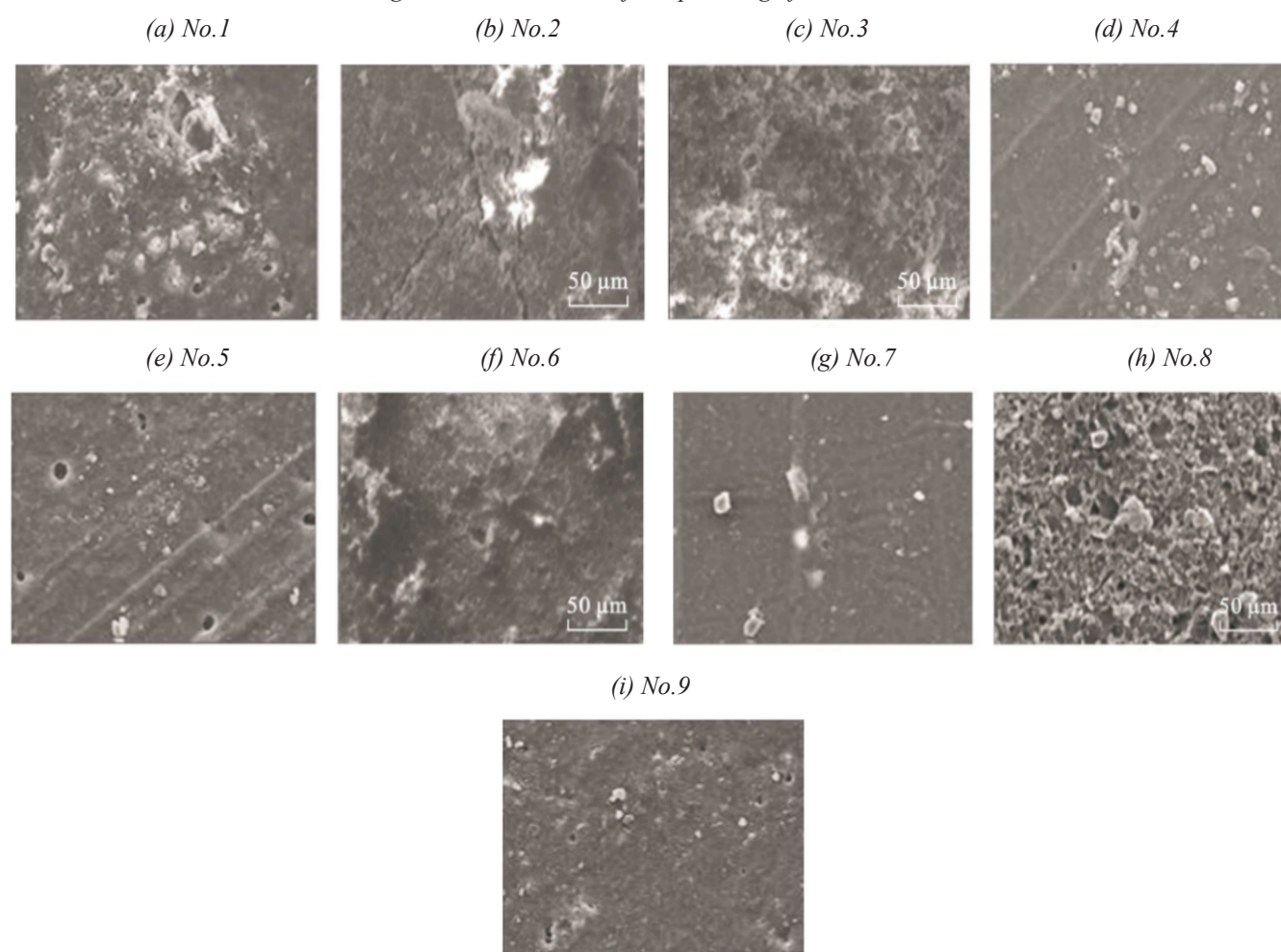


② SEM analytic result

The SEM test results of each sample are shown in the figure below. It can be observed that the surfaces of the samples exhibit

different characteristics: some are dominated by pores (Sample 9), some appear to have signs of corrosion (Sample 8), and others show cracks (Sample 2) and depressions (Sample 6). The aging caused by discharge particles mainly manifests as concentrated and regular pores and cracks, while the chemical corrosion induced by nitric acid covers a larger area with irregular and uneven pores. It is not hard to infer that if charged particles and nitric acid act together, silicon rubber surfaces will inevitably develop irregular cracks or large areas of depression. Comparing the test conditions of each group of samples, it is evident that Sample 2, which has low pressure, high humidity, and long aging time, shows a deep groove on its surface, indicating the most severe degree of aging; Sample 7, with high pressure and short aging time, although having higher humidity, only shows several small holes on its surface, indicating the least degree of aging; Sample 3, with low pressure and moderate humidity but a longer aging time, shows scattered small holes due to corrosion on its surface. Comparing Samples 1, 5, and 9, it is found that Samples 9 and 5 have a longer aging time but mainly show scattered, deeper pores, while Sample 1 shows large areas of holes. Therefore, changes in the surface morphology of samples are related to pressure, humidity, and aging time, with humidity having a more significant impact at low pressures. Significantly, the effect of aging time is greater at high pressure.

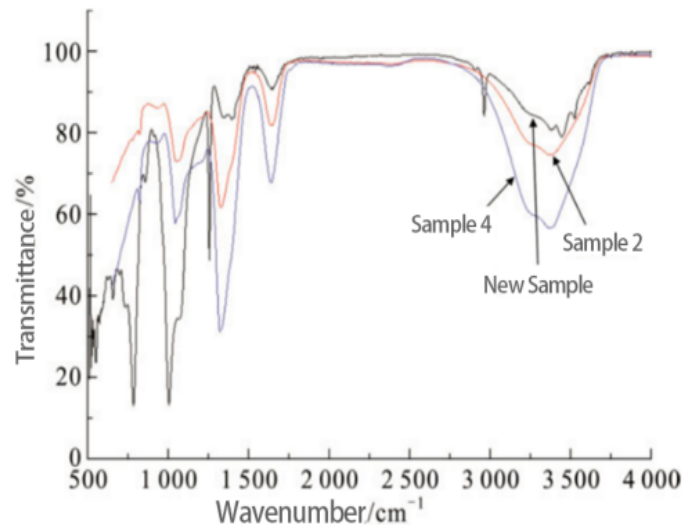
Figure 25. SEM results of samples magnified 500 times



③ FTIR analysis results

The test results are shown in the figure below. It can be seen that compared with the new sample, the FTIR spectrum of sample 2 has changed as follows: 1) wavenumber is 1007cm^{-1} The characteristic absorption peak weakened, indicating that the Si-O-Si structure decreased; 2) wavenumber was 1650cm^{-1} The absorption peak was enhanced, indicating that it was NO_3 The content increased, i.e. $\text{Al}(\text{NO}_3)_3$ Increased content; 3) wavenumber was 3500cm^{-1} The absorption peak near it becomes wider and stronger, indicating that the -OH content in the sample increases; 4) wavenumber 1260cm^{-1} The absorption peak almost disappeared, indicating Si-CH₃ Structural reduction.

Figure 26. FTIR test results of typical samples



(3) Mechanism analysis

Under the action of corona high energy particles, the molecular bonds of silicone rubber are broken, which is shown in FTIR as Si-O-Si and Si-CH₃. The corresponding characteristic absorption peak decreases. At the same time, when silicone rubber material is subjected to corona discharge in air, a certain amount of nitric acid will be formed, and the concentration of nitric acid increases with the increase of relative humidity, corresponding to wavenumber 1650 cm⁻¹ and 3500 cm⁻¹. The absorption peak of the site is enhanced, -NO₃. The increase of -OH content indicates that nitric acid is involved in the corona aging process. Nitric acid acts on silicone rubber, and due to its strong corrosiveness, it will destroy the physical structure of its surface and produce new chemical groups, thus causing the change of its trap characteristics.

Different air pressures will affect the corona intensity, which is the density of charged particles in the conductive channel. The impact of humidity on corona aging is complex. Humidity not only affects the intensity of corona discharge but also, due to high humidity, prolonged corona discharge can produce nitric acid. The strong corrosiveness of nitric acid can further age silicone rubber materials.

At low pressure (45 kPa), the air density is low, and the number density of charged particles in the discharge channel is high, with a large free travel distance. This allows charged particles to accumulate significant kinetic energy, which can bombard the material surface. When corona aging occurs for an extended period, environmental humidity increases to a certain level (Sample 3) and remains so for a sufficient duration, moisture will adhere to the material and electrode surfaces, locally increasing the electric field and making discharges more likely. Both SEM and FTIR indicate that the impact effect of charged particles becomes more pronounced at this time. When humidity reaches its maximum and aging has occurred for a certain period (Sample 2), not only does moisture distort the surface electric field of the sample, exacerbating discharges, but as nitric acid content increases, its strong corrosive properties will act on the silicone rubber material surface, causing corrosion and making it easier for charged particles to bombard the interior of the sample, accelerating aging. On the other hand, when charged particles bombard the sample surface, they form cracks, pores, and other defects, allowing nitric acid to penetrate deeper into the sample, increasing the contact area with silicone rubber and leading to more severe aging. Therefore, Sample 2 (45 kPa, 90%, 400 h) exhibits longer and deeper grooves on its surface, with a reduction in Si-O-Si structure, resulting in the most severe physical and chemical damage, generating more physical and chemical defects.

When the air pressure is high, the air density is greater, and the electron number density in the discharge channel is lower. As humidity increases, fewer water molecules are captured, thus having a smaller impact on the current. When humidity further increases and remains for an extended period, although water molecules may also adhere and distort the field strength, due to the significantly lower number of water molecules in space compared to low pressure conditions, the distortion effect is not significant, leading to minimal changes in current. At this point, there is less nitric acid content, and corrosion on the silicone

rubber surface requires a longer time. Therefore, under high air pressure, the aging effect is more pronounced. In this case, the synergistic effect between charged particles in the silicone rubber and nitric acid is difficult to manifest, resulting in fewer defects in the sample (Sample 7). Research findings from Tsinghua University also confirm this analysis.

Due to the effect of corona discharge, the microstructure of materials will change, forming many new defects or trap centers. When corona discharge occurs, injected charged particles may be captured, forming space charges. SEM results show that the surface of new samples is relatively smooth and flat, with a uniform structure. As the corona time increases, needle-like pores appear on the silicone rubber surface, then the number of pores increases and deepens, eventually developing into cracks or even fissures, leading to grooves. It is evident that the surface structure of the samples is gradually destroyed under the combined action of charged particle bombardment and nitric acid corrosion, resulting in more physical and chemical defects. The formation of these defects means that the injected charges must absorb more energy to be excited, thus the corresponding temperature during excitation is higher, and the trap energy level deepens, which is reflected in the TSC curve as an increase in current peak and a rise in the corresponding temperature.

3.Summary

To reveal the degradation mechanism of composite external insulation materials in heavily polluted areas, this section conducts experimental studies on the main factors affecting the degradation of composite external insulation—heavy pollution and corona discharge. The study tests the changes in various physicochemical micro-performance properties of silicone rubber materials under different conditions, obtaining the destruction mechanisms of organic components in silicone rubber materials under heavy pollution and corona effects. This provides a theoretical foundation for subsequent aging assessment and performance optimization. Details are as follows:

A composite material corona aging test system has been developed to conduct corona aging tests on silicone rubber materials. Using methods such as SEM, static contact angle analysis, TSC, and FTIR, the effects of corona duration, corona voltage, nitric acid exposure time, nitric acid PH value, and different pressures and humidity levels on the performance of silicone rubber materials have been studied. The differential damage mechanisms of charged particles and nitric acid on organic molecules under different pressures and humidity levels have been analyzed, refining the theory of corona aging. In subsequent optimization of silicone rubber materials, efforts can be focused on improving C-H, Si-C, and terminal Si-OH bonds. Due to the non-equivalence between artificial corona and field corona, the following conclusions will not undergo quantitative analysis.

- ① As the corona intensity increases, the hydrophobic performance of silicone rubber material rapidly decreases. Within 20 minutes at 3kV corona, the hydrophobic performance drops to around 10°, and with increasing test time, it no longer changes; the rate of hydrophobic recovery decreases as the corona intensity increases and the duration of corona exposure extends, but due to the influence of the materials microstructure, the recovery rate is not monotonically increasing or decreasing, showing fluctuations in between. Under nitric acid alone, the degradation rate of the hydrophobic performance of silicone rubber material slows down, generally only dropping to about 55°-70°. With increased nitric acid exposure time and concentration, the rate of hydrophobic recovery shows a consistent trend with corona, and can all recover to above 90°.
- ② As the corona intensity increases and the exposure time extends, changes occur on the surface morphology of silicone rubber materials, including pore-crystal layer-crack processes. Pores exhibit both smooth and irregular edges. As the corona PH value decreases and the exposure time extends, changes occur on the surface of silicone rubber materials, such as increased pore density and depth. Pores are distributed irregularly.
- ③ With the increase of corona intensity and the extension of action time, the peak TSC current and its corresponding temperature of the material increase, but mostly a single peak; with the decrease of corona PH value and the extension of action time, the peak TSC current and its corresponding temperature of the material increase, and a double current peak appears.
- ④ Under the effect of corona, the molecular breakage and oxidation of silicone rubber materials are caused by the kinetic energy of high-energy charged particles and the oxidative and corrosive effects of nitric acid. Charged particles and nitric acid

successively destroy the C-H bond, Si-C bond, and Si-O bond in silicone rubber molecules. After the destruction of the C-H and Si-C bonds, the organic components of the material decrease, its hydrophilicity increases, and its hydrophobicity rapidly declines. However, after the Si-O bond is broken, long-chain siloxanes decompose, leading to an increase in small-molecule siloxanes, which accelerates the recovery of hydrophobicity; the formation of inorganic silicon components also enhances its hydrophilicity, reducing the irreversibility of organic components and blocking the migration of hydrophobic substances. After the oxidation of silicone rubber molecules, the content of hydrophilic OH groups increases. The highly reactive OH groups condense, reducing the content of hydrophilic OH groups while increasing the molecular chain length and decreasing the content of small-molecule siloxanes. At the same time, chemical reactions at the molecular level cause surface pores, crystal layers, and cracks on the material, accelerating the rate and depth of corona damage and affecting the migration rate of hydrophobic substances. Therefore, the fluctuation in the loss and recovery of hydrophobicity in silicone rubber materials is caused by complex reactions such as molecular breakage, oxidation, and condensation, which are closely related to the surface structure of the material.

⑤ During the corona process, the breaking of molecular bonds increases electron traps. After the surface structure of the material is damaged, charged particles and nitric acid penetrate into the material, increasing the depth of electron traps. Charged particles mainly bombard the surface, so TSC curves are often unimodal; nitric acid primarily causes corrosion and oxidation, resulting in bimodal TSC curves. The increase in electron traps enhances the materials ability to capture electrons, increases the degree of surface electric field distortion, making it more prone to corona.

⑥ Under different pressures, the mechanisms of organic molecule degradation in silicone rubber materials vary. Under high pressure, water has a lesser impact on electric field distortion, and the density of charged particles and nitric acid is low, so the degradation of silicone rubber material is mainly due to charged particle damage, with less influence from nitric acid. Under low pressure, water has a greater impact on electric field distortion, and the density of charged particles and nitric acid is high, so the degradation of silicone rubber material is caused by the combined effects of charged particles and nitric acid, with higher humidity leading to more pronounced nitric acid effects.

Funding

no

Conflict of Interests

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

- [1] Bi, M., Deng, R., Jiang, T., Chen, X., Pan, A., Zhu, L., 2022. Study on Corona Aging Characteristics of Silicone Rubber Material Under Different Environmental Conditions. IEEE TRANSACTIONS ON DIELECTRICS AND ELECTRICAL INSULATION.
- [2] Hakami, M., El-Hag, A., Jayaram, S., 2021. Effects of Corona Discharges on Silicone Rubber Samples under Severe Weather Conditions. 2021 IEEE ELECTRICAL INSULATION CONFERENCE (EIC).
- [3] Wang Jianguo, Liu Yang, Fang Chunhua, Chen Junjie, Xie Congzhen, Cai Li, 2009. PERFORMANCE INFLUENCE OF HIGH TEMPERATURE VULCANIZATION SILICONE RUBBER BY ALTERNATING CURRENT CORONA DISCHARGE. ACTA POLYMERICA SINICA.
- [4] Haji, K., Zhu, Y., Otsubo, M., Honda, C., 2007. Surface Modification of Silicone Rubber After Corona Exposure. PLASMA PROCESSES AND POLYMERS.
- [5] Zhu, Y., Haji, K., Otsubo, M., Honda, C., 2006. Surface degradation of silicone rubber exposed to corona discharge. IEEE TRANSACTIONS ON PLASMA SCIENCE.
- [6] Guo, J.-k., Hao, C.-y., Yu, J.-s., Lu, L.-q., 2017. Effect of Corona Discharge Treatment on Interface Discharge Behaviour between Polypropylene and Silicone Rubber. PROCEEDINGS OF THE 2ND 2016 INTERNATIONAL CONFERENCE ON SUSTAINABLE DEVELOPMENT (ICSD 2016).
- [7] Sarathi, R., Mishra, P., Gautam, R., Vinu, R., 2017. Understanding the Influence of Water Droplet Initiated Discharges on

Damage Caused to Corona-Aged Silicone Rubber. IEEE TRANSACTIONS ON DIELECTRICS AND ELECTRICAL INSULATION.

- [8] Zhu, Y., 2019. Influence of corona discharge on hydrophobicity of silicone rubber used for outdoor insulation. POLYMER TESTING.
- [9] Zhu, Y., Otsubo, M., Honda, C., Tanaka, S., 2006. Loss and recovery in hydrophobicity of silicone rubber exposed to corona discharge. POLYMER DEGRADATION AND STABILITY.
- [10] Zhang, Z., Pang, G., Ma, X., Liang, T., Jiang, X., 2022. Mechanical Properties of High-Temperature Vulcanized Silicone Rubber Under Acid-Fog With AC Energized. IEEE ACCESS.

Design and Implementation of Time-Frequency Taming Algorithm of Rubidium Atomic Clock Based on FPGA

Sun Songfei, Hou Yuanhang, Lv Yiheng, Wang Zhiming, Chen Qun*

Henan Polytechnic, Zhengzhou 450000, China

*Corresponding author: Chen Qun, chenqun918@126.com

Copyright: 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Synchronously improving the accuracy of time and frequency sources in electronic measurement systems can enhance measurement precision. Fully combining the good long-term stability of 1pps output from satellite navigation receiver and the good terminal stability of rubidium atomic clock, a frequency and time-locked algorithm based on second-order digital phase-locked loop is designed, which integrates frequency taming and time synchronization, and is implemented based on FPGA; The test results show that the algorithm can obtain a better accuracy 1PPS time signal while taming the frequency with good long-term stability external 1pps.

Key words: Rubidium Atomic Clock; Digital Phase Locked Loop; Lpps; Frequency Taming

Published: Apr 15, 2025

DOI: <https://doi.org/10.62177/jaet.v2i2.218>

Introduction

In the process of electronic measurement, the accuracy of time and frequency sources greatly affects the accuracy of measurement results. Many instruments are equipped with external frequency sources and time/event trigger source input interfaces. In many high-precision measurement scenarios, instruments and equipment need to be connected to high-precision frequency sources such as temperature-controlled crystal oscillators and atomic clocks.

Rubidium atomic clocks offer advantages such as high short-term and long-term stability, high frequency accuracy, and low aging rates, and have been widely used in situations requiring high precision in time and frequency. Although rubidium atomic clocks are stable and accurate, the frequency will always drift and accumulate over time, leading to a drift in both time and frequency; to address this issue, the frequency drift of rubidium atomic clocks can be corrected using pulse signals (1pps) generated by satellite navigation receivers with good long-term stability^{[1][2][3]}.

In order to solve this problem, extensive research has been carried out in the industry, such as Lu Xianghong et al. proposed the design of a constant temperature crystal oscillator frequency calibration system based on FPGA^[4], Liu Tieqiang et al. proposed the design of a clock taming system based on Beidou timing^[5], and Fan Wenjing et al. improved the application of Kalman filter algorithm in clock taming technology^[6]. Most of these algorithms and applications focus on taming the frequency of the constant temperature crystal oscillator or atomic clock, overcoming the frequency drift and improving the frequency accuracy. However, in many applications, not only accurate frequency signals are required, but also 1pps (second pulses) are required to characterize accurate time signals^[7].

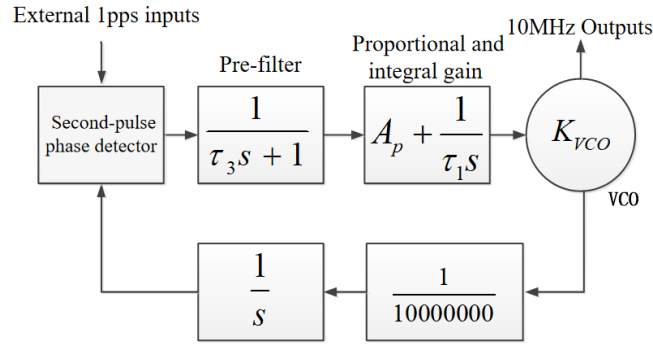
Based on the second-order digital phase-locked loop, the frequency of the rubidium atomic clock is tamed by using an

external 1pps^[8]; At the same time, the loop output is more accurate than the external 1pps, and the frequency lock and time lock are completed at the same time. The high-speed FPGA is used as the processor for the implementation of the algorithm, which has the characteristics of fast processing speed, high real-time performance, and good stability^{[9][10][11]}.

1.Design of Rubidium atomic clock frequency-locked phase-locked algorithm based on second-order digital phase-locked loop

The block diagram of the phase-locked loop is shown in Figure 1, where the “second pulse phase detector” performs phase detection of the 1pps from the external input and the 1pps generated by the frequency division of the rubidium atomic clock, with a gain of $K_{\text{det}} = 1\text{bit} / \text{ns}$. The loop filter is a digital filter that consists of a pre-filter and a standard programmable proportional-integral (PI) controller. The VCO is the rubidium atomic clock with a frequency of f , which is tuned by voltage, and the parameters adjust the frequency, with a 1pps output sensitivity of $K_{\text{vco}} = 0.001\text{ns} / \text{bit} - \text{s}$, or $(1/10^{12})$. The response function of each component of the digital phase-locked loop corresponds to the standard Laplace variable S depicted in the figure.

Figure 1 Block diagram of an external 1pps phase-locked loop



The PI controller is controlled by selecting the appropriate integrator time constant of τ_1 and a stability factor of ζ , τ_1 determines the natural time constant of the PLL τ_n , τ_n represents the step size of the tracking reference phase, while ζ determines the relative rise time and ringing of the PLL based on the step size. A value of ζ also represents the balance between the equivalent noise bandwidth and the peak passband near the natural frequency of the response function. The integration time constant is τ_1 , from 2^8 to 2^{22} seconds. The natural time constant is given by $\tau_n = \sqrt{\tau_1 / K_{\text{det}} K_{\text{vco}}} = \sqrt{(1000\text{s})\tau_1}$. Thus, the natural time constant ranges from 506 seconds to 18 hours. The integral time constant τ_1 determines the natural time constant τ_n , which is the natural time constant describing the loop response.

The stability factor ζ ranges from 0.25 to 4.0. The default value of $\zeta = 1.0$ corresponds to the critical damping response; $\zeta < 1.0$ and $\zeta > 1.0$ correspond to underdamped and overdamped responses, respectively.

Given τ_1 and ζ , the proportional gain of the controller is A_p , as given by equation $A_p = 2\zeta\sqrt{\tau_1 K_{\text{det}} K_{\text{vco}}} = 2\zeta\sqrt{(0.001\text{s}^{-1})\tau_1}$. With the default constant of $\tau_1 = 65536$ seconds and $\zeta = 1.0$, the proportional gain is approximately 0.25. In this case, the instantaneous frequency of the rubidium atomic clock will be adjusted by about 0.25 parts in every measured nanosecond of 10^{12} .

The pre-filter exponentially averages the time labels output from the “second pulse phase detector” before passing the results to the PI controller. The time constant of the pre-filter is τ_3 , which is coded as $\tau_n / 6.0$, which is to obtain maximum average benefit while ensuring that the phase-locked loop will be stable. The smoothing effect of the pre-filter greatly reduces the sensitivity of the 1pps phase jitter that occurs between 50ns and 300ns on the external reference 1pps of the digital phase-locked loop. The 1pps output of the GPS/Beidou navigation receiver has a large amount of noise associated with $1/f$; Therefore, a long time constant is required to prevent the rubidium atomic clock from getting too close to this noise. The system provides a natural time constant of up to 18.0 hours, which will allow the rubidium atomic clock to track the 1pps output of the GPS/Beidou navigation receiver on a one-day time scale, while retaining the superior short-term stability of the rubidium clock.

The digital phase-locked loop will approach the following three equations in the case of different ζ .

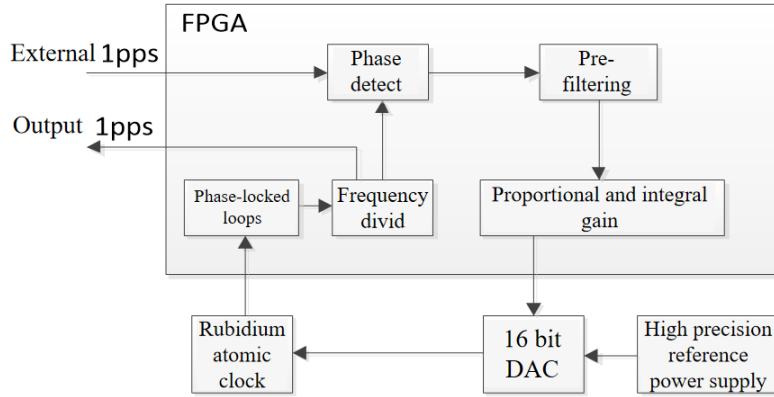
$$\begin{aligned}
\Delta T(t) &= \frac{F_0 - \zeta \Delta T(0) / \tau_n}{\sqrt{1 - \zeta^2} / \tau_n} e^{\frac{-\zeta t}{\tau_n}} \sin(\sqrt{1 - \zeta^2} t / \tau_n) \\
&+ \Delta T(0) e^{\frac{-\zeta t}{\tau_n}} \cos(\sqrt{1 - \zeta^2} t / \tau_n) \quad \text{for } \zeta < 1 \\
\Delta T(t) &= t[F_0 - \Delta T(0) / \tau_n] e^{\frac{-t}{\tau_n}} + \Delta T(0) e^{\frac{-t}{\tau_n}} \quad \text{for } \zeta = 1 \\
\Delta T(t) &= \frac{-[F_0 - (\zeta + \sqrt{\zeta^2 - 1}) \Delta T(0) / \tau_n]}{2\sqrt{\zeta^2 - 1} / \tau_n} e^{\frac{-(\zeta + \sqrt{\zeta^2 - 1})t}{\tau_n}} \\
&+ \frac{[F_0 - (\zeta - \sqrt{\zeta^2 - 1}) \Delta T(0) / \tau_n]}{2\sqrt{\zeta^2 - 1} / \tau_n} e^{\frac{-(\zeta - \sqrt{\zeta^2 - 1})t}{\tau_n}} \quad \text{for } \zeta > 1
\end{aligned}$$

$\Delta T(0)$ is the initial phase deviation between the rubidium atomic clock and the reference. F_0 is the frequency difference between the digital phase-locked loop and the reference before it is activated. $\Delta T(t)$ describes the time function when the phase of the rubidium atomic clock reaches the phase of the reference. At the default time constant of $\tau_1 = 65536$ seconds and a stabilization factor of $\zeta = 1.0$, the 1pps output produced by the rubidium atomic clock will be exponentially close to the reference 1pps phase with a time constant of $\tau_n = 8095$ seconds or 2.25 hours.

2.Implementation of Rubidium atomic clock frequency-locked phase-locked algorithm based on second-order digital phase-locked loop

The 10MHz frequency output of the rubidium atomic clock first enters the phase-locked loop inside the FPGA and is doubled to 320MHz, which is used as the master clock for internal counting and processing to improve the accuracy of processing.

Figure 2 Block diagram of time and frequency locking



The first step is 1pps phase calibration. Because the 1pps external 1pps produced by the rubidium atomic clock itself has a large phase difference, after receiving 256 consecutive “good” external 1pps inputs (e.g. the 1pps of subsequent inputs and the first 1pps phase error of less than 2048ns). The 1pps phase produced by the rubidium atomic clock assignment is corrected to the 1pps phase of the last external input. In addition, the current value of the frequency control parameters of the rubidium atomic clock (such as the calibration value at the factory) is used to initialize the integrator in ± 2000 units of $\pm 10^{12}$, and the exponential filter after the “second pulse phase detector” is reset to zero.

The second step is frequency and 1pps taming; The frequency is locked to a “good” 1pps input pulse, and a bad “1pps input (the time label of the 1pps input is 1024ns larger than the time label of the previous “good” 1pps input) will be rejected. The frequency parameter f of the frequency control command will be updated with each “good” timestamp result, $\Delta T(n)$, As shown below:

Prefilter:

$$\overline{\Delta T}(n+1) = (1.0 - \Delta t / \tau_3) \overline{\Delta T}(n) + (\Delta t / \tau_3) \Delta T(n)$$

Integral term:

$$Int(n+1) = Int(n) - (\overline{\Delta T}(n+1) / \tau_1) K_{det} \Delta t$$

Proportional term:

$$Pro(n+1) = -A_p \overline{\Delta T}(n+1) K_{det}$$

Frequency setting:

$$f(n+1) = Pro(n+1) + Int(n+1)$$

In the above equations, Δt is the time interval between phase comparisons, which is 1 second for this system. The range of the frequency control value f exceeds ± 2000 counts. If the new value of f exceeds 2000, it is set to 2000. If the new value of f is less than -2000, it is set to -2000. If the new integral term exceeds 2000, it is set to 2000. If the new integral term is less than -2000, it is set to -2000. This prevents the integrator from diverging when holding a fixed f value for an extended period to align the 1pps output with the 1pps input.

The third step is to control the frequency of the rubidium atomic clock. The output f of the digital filter serves as the frequency control word, which is converted into a voltage signal by a 16-bit high-precision DAC to control the frequency of the rubidium atomic clock, updated once per second. To ensure the precision of the control voltage, the DAC employs a high-precision reference voltage source.^{[12][13]}

If there are 256 consecutive “bad” 1pps inputs, the PLL will terminate and restart (this may occur if the 1pps input suddenly shifts by more than 1024 ns). The PLL will also terminate and restart if the measurement time tag value of the “good” 1pps input exceeds $\pm 4ns / s * \tau_1$ (a situation that may occur if the frequency of the input 1pps deviates from 10^{-9} by several units for an extended period), when the default value of τ_1 is 65536 seconds, if the cumulative phase difference of the “good” 1pps exceeds 262144 nanoseconds, the PLL will restart.

3. Test Results

The test method employed is a comparative test using two identical devices; hence, the measured error represents the relative error between the two devices. The test utilizes a frequency counter, model 53220A. To ensure stable test output results and to create a fixed phase delay between the 1pps outputs of the two devices, a cable delay method is used. This facilitates the setup of the test gate signal. During the test, only the variance and maximum range need to be monitored. The test is set to take a measurement every two seconds, and in the test graphs, one data point corresponds to a time interval of two seconds.

3.1 Performance of the 1pps Output from the Navigation Module

Figure 3: Performance of the 1pps Output from the Navigation Module

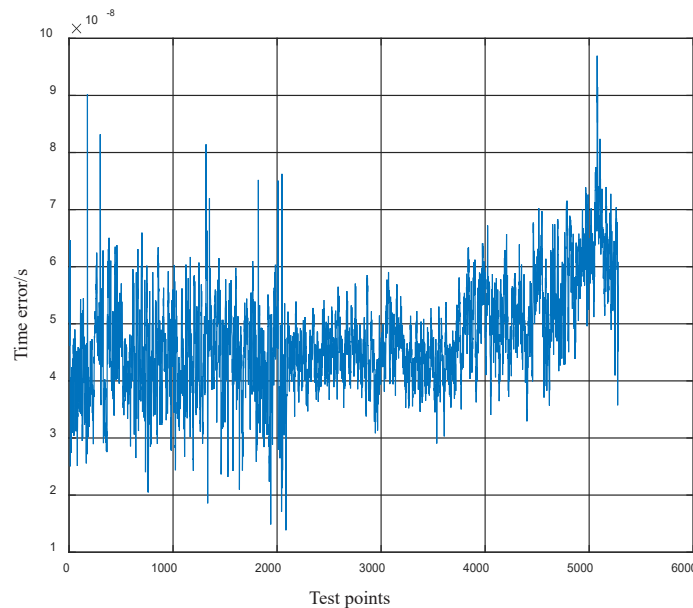


Figure 3 shows the performance of the raw 1pps output from the navigation module. The variance of the time error in the data is 9.09 ns, and the difference between the maximum and minimum values is 83.12 ns. The test data indicates that the error in the original 1pps output from the navigation module is relatively large. This test data is affected by factors such as the environment where the antenna is located and the climatic conditions. However, subsequent tests were also conducted under essentially the same conditions.

3.2 Time Phase Correction and Locking

Figure 4 illustrates the initial phase correction process. During the initial stage, the phase of the 1pps generated by the free-running rubidium atomic clock is corrected using 256 consecutive 1pps inputs with small deviations from the external navigation module. The figure shows that the initial phase deviation between the two 1pps signals was 266 ms, and after correction, the phase of the rubidium atomic clock's 1pps follows that of the last input 1pps from the navigation module (tending towards 0 in Figure 4). Of course, as shown in Figure 3, this initial phase also has a relatively large error and requires joint correction of time phase and frequency to gradually achieve stability. The time phase locking process is shown in Figure 5. The initial error was around 30 ns, and after approximately 1000 test points, which corresponds to 2000 seconds, the output 1pps signal stabilizes, completing the frequency and phase locking process.

Figure 4: Initial Time Phase Correction Process

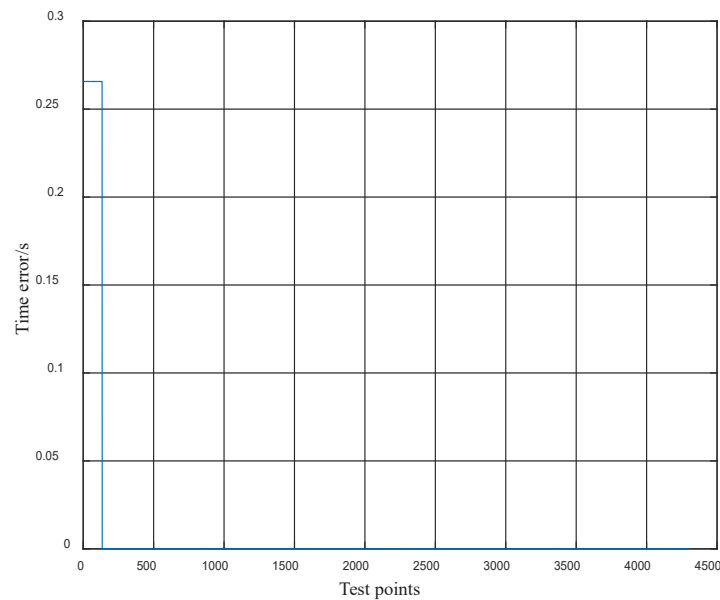
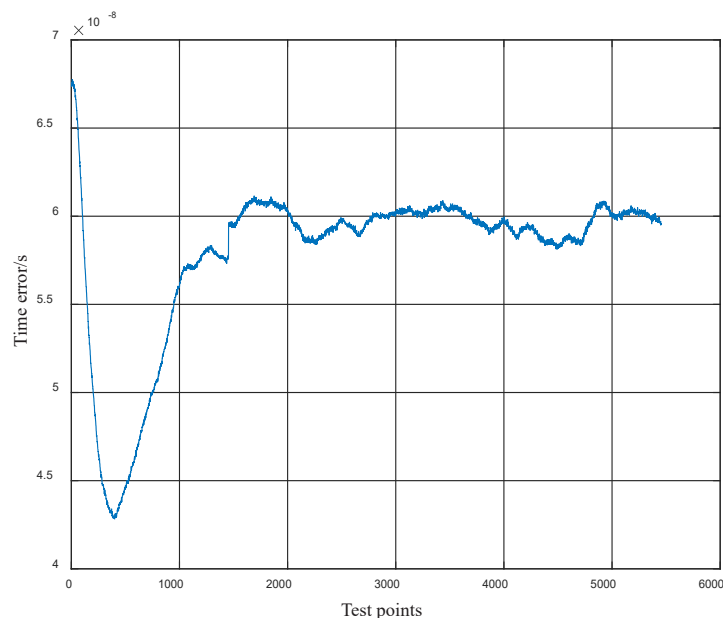


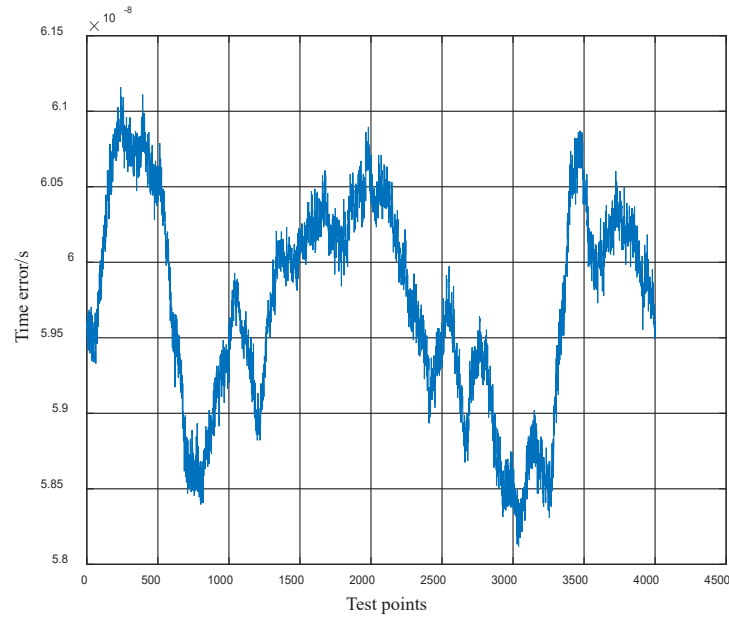
Figure 5: Time Phase Locking Process



3.3 Performance Test of 1pps After Time and Frequency Locking

The performance of the 1pps after time and frequency locking is shown in Figure 6. The variance of the data in the figure is 0.69 ns, and the difference between the maximum and minimum values is 3.04 ns. Compared to the original 1pps from the navigation module shown in Figure 3, the performance has been significantly improved. The frequency stability can be calculated as $3.04\text{ns}/8000\text{s}=3.8*10^{-13}$.

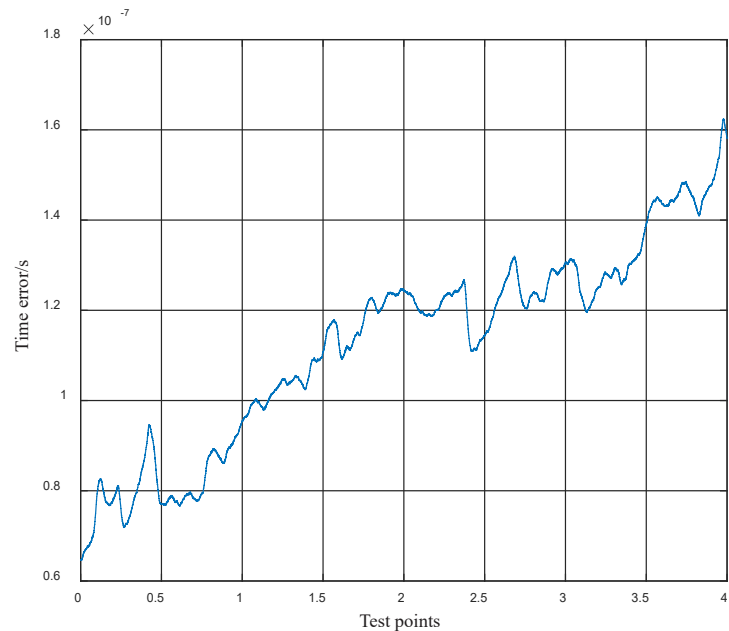
Figure 6: Performance of 1pps After Time Locking



3.4 Timekeeping Performance Test

To measure the drift after frequency locking when there is no external 1pps, a timekeeping performance test was conducted after 24 hours of locking. The test results are shown in Figure 7, with a data variance of 22.23 ns and a difference between the maximum and minimum values of 98.06 ns. This means that the error after timekeeping for 80000 seconds is about 100 ns, and the frequency error can be calculated as $1.25*10^{-12}$.

Figure 7: Timekeeping Performance After 24 Hours of Locking



4. Conclusion

Based on a second-order digital phase-locked loop, the frequency of the rubidium atomic clock is tamed using an external 1pps. Meanwhile, the loop outputs a more accurate 1pps than the external 1pps, achieving frequency and time locking

simultaneously. The algorithm is implemented using a high-speed FPGA as the processor. Test results show that both the frequency accuracy and the accuracy of the 1pps of the designed system have been significantly improved.

Funding

no

Conflict of Interests

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

- [1] Andreas B, Dirk P, Thomas P, et al. Amelioration of the usage and monitoring of GNSS signals at PTB[C]. Proceedings of the 50th Annual Precise Time and Time Interval Systems and Applications Meeting, Reston, USA, January 28-31, 2019.
- [2] Kun L, Felicitas A, Gerard P, et al. Evaluation of BeiDou time transfer over multiple inter-continental baselines towards UTC contribution[J]. Metrologia. 2018, 55(4): 513-525.
- [3] Yang Y X, Gao W G, Guo S R, et al. Introduction to BeiDou-3 navigation satellite system[J]. Navigation, 2019, 66(1): 1-12.
- [4] Lu Xianghong, Chen Rujun, He Zhanxiang. Design of a Frequency Calibration System for Oven-Controlled Crystal Oscillator Based on FPGA [J]. Application of Electronic Technology, 2010, 36(07): 101-104.
- [5] Liu Tieqiang, Huo Jing. Design of a Clock Taming System Based on Beidou Timing [J]. Radio Engineering, 2018, 48(07): 569-572.
- [6] Fan Wenjing, Wang Zhaoli, Zhou Wei. Application of Improved Kalman Filter Algorithm in Clock Taming Technology [J]. Aerospace Metrology & Measurement Technology, 2016, 36(03): 46-49.
- [7] Sun Xuelin, Qin Mingwei, Hu Guilin. Design of a High-Precision Beidou Timing System Based on TDC and ARM [J]. Aerospace Metrology & Measurement Technology, 2021, 41(01): 27-32.
- [8] Chi Huashan, Zhang Lei, Chi Wenbo, Qiang Chenghu. A Disciplining Method for Rubidium Clock Based on GPS/BeiDou Common-View Technology[J]. Computer Measurement & Control, 2017, 25(07): 306-308.
- [9] Lei Zhang, Hai-sheng Huang, Bin Zhang. Design and Implementation of FPGA-based GPS Clock Disciplining Circuit[J]. Electronic Design Engineering, 2014, 22(11): 61-64.
- [10] Liu Chang, He Yuling, Du Erwang, Xu Feng, Zhang Rongyan, Hu Jiayu, Cheng Min, Du Lijun. "Application of Digital Technology in Spaceborne Rubidium Atomic Clocks" [A]. In: Academic Exchange Center of China Satellite Navigation System Management Office (ed.). Proceedings of the 12th China Satellite Navigation Conference: S05 Space Reference and Precise Positioning [C]. Beijing: Organizing Committee of China Satellite Navigation Conference, 2021: 9.
- [11] Wang Yajun, Zhang Lei, Gu Yang, Huang Yan, Kang Tingting. Development of a Beidou Disciplined Rubidium Atomic Frequency Standard Device Based on FPGA [J]. Acta Metrologica Sinica, 2020, 41(03): 359-362.
- [12] Cui Yongjun, Liu Yang, Yang Bing. Design of a High-Precision Time Reference Measurement System Based on Rubidium Atomic Clock and Dual TDC-GP2[J]. Electronic Devices, 2017, 40(5): 1072-1077.
- [13] Yu Zhihui, Yan Shidong, Ming Gang, Mei Ganghua, Zhong Da. Design of high-precision frequency adjustment circuit for miniaturized rubidium atomic clock [J]. Chinese Journal of Magnetic Resonance, 2019, 36(1): 103-112.

Design of Lightweight, High Precision, Multi Scene Module Based on Narrowband Internet of Things

Sun Feiyu, Liu Xiangyu, Shangguan Lena, Ren Jingbo, Ma Qingsong, Chen Qun*

Henan Polytechnic, Zhengzhou 450000, China

*Corresponding author: Chen Qun , chenqun918@126.com

Copyright: 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: At present, the number of active IoT devices worldwide has exceeded 16 billion, but existing devices generally suffer from problems such as large size, short battery life, low level of intelligence, and poor stability. In response to the limitations of traditional IoT terminals in terms of size, power consumption, and scene adaptability, this paper designs a lightweight IoT intelligent module based on STM32 microcontroller and Narrowband IoT (NB IoT). Environmental data is collected through temperature and humidity sensors, as well as light sensors, and processed through the STM32 microcontroller. With the low-power, wide coverage, and strong link characteristics of narrowband IoT modules, remote data reporting is achieved. The experimental verification results show that the module has good performance and can accurately collect environmental data such as temperature and humidity. The module can be used in multiple scenarios such as bird tracking with good feedback.

Key words: Narrowband Internet of Things; STM32 Microcontroller; IoT Module; Lightweight

Published: Apr 16, 2025

DOI: <https://doi.org/10.62177/jaet.v2i2.224>

Introduction

With the increasing demand for intelligent transformation in various industries, IoT terminals have become the main tool for achieving the Internet of Things. At present, traditional IoT terminal devices generally suffer from problems such as bulky size, high power consumption, and limited communication distance, which seriously restrict their application in mobile object monitoring and long-distance data transmission scenarios^[1]. In recent years, with the development of low-power wide area network technology, a new solution has been provided for the miniaturization and intelligence of terminal devices. Through adaptive power management, wide area coverage capability has been achieved while ensuring ultra long battery life, significantly expanding the application scope of IoT technology in complex environments^[2].

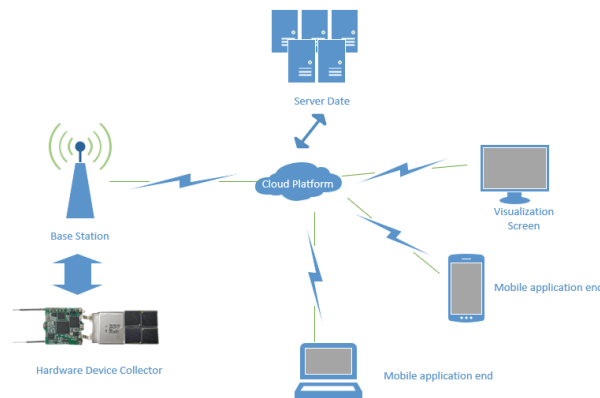
The existing IoT modules mostly adopt communication technologies such as ZigBee, WiFi, GPS/GSM combination scheme, LoRa, etc. The GPS/GSM combination scheme has the problems of high power consumption and expensive operating costs, while LoRa is limited by complex networking. ZigBee and WiFi are not suitable for use in long-distance notification scenarios due to their close communication distance^[3]. Especially in outdoor scenarios without network coverage, device endurance and data transmission reliability become the core bottlenecks restricting long-term monitoring. Due to limitations in volume, weight, energy supply, and data transmission methods, traditional IoT terminal devices commonly suffer from low positioning accuracy, short battery life, and high data transmission delays^[4]. This article designs a lightweight, multi

scenario, high-precision IoT module based on STM32 and NB IoT technology. The system uses STM32 as the core control unit, combined with high-precision sensor matrices such as temperature and humidity sensors and light intensity sensors, to achieve multidimensional data acquisition.

1. Overall System Architecture

The system architecture consists of a perception layer, a network layer, a platform layer, and an application layer^[4]. The perception layer adopts Cortex-M4 and Cortex-M0+dual core STM32WB55 multi-core processors, which manage nine axis sensors, environmental sensors, and positioning modules through SPI/I2C bus to achieve synchronous acquisition of motion trajectory and environmental parameters. The network layer adopts Bluetooth 5.0 and Cat-M1/NB IoT dual-mode communication, and establishes a dedicated APN channel through the China Telecom IoT platform. The platform layer is deployed on the Alibaba Cloud IoT platform. The application layer provides web and mobile visual interfaces, supporting geofencing settings and abnormal warning functions. The module system architecture is shown in Figure 1.

Figure 1 System Architecture Diagram

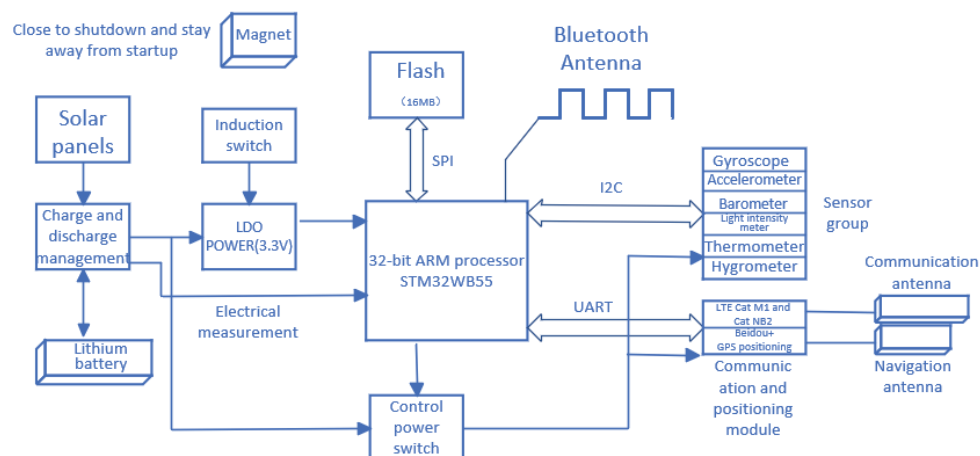


Next, the design and implementation process of the system will be detailed from the aspects of overall system design scheme, hardware circuit design, software program design, and system testing and verification. Validate and evaluate the performance and stability of the system through experiments.

2. Hardware Circuit Design

The system hardware circuit mainly includes STM32 core processing unit, power management unit, storage unit, sensor group, communication positioning module and antenna. The hardware circuit design fully considers the lightweight requirements, and surface mount components are used in the hardware selection stage. The hardware architecture of the system is shown in Figure 2.

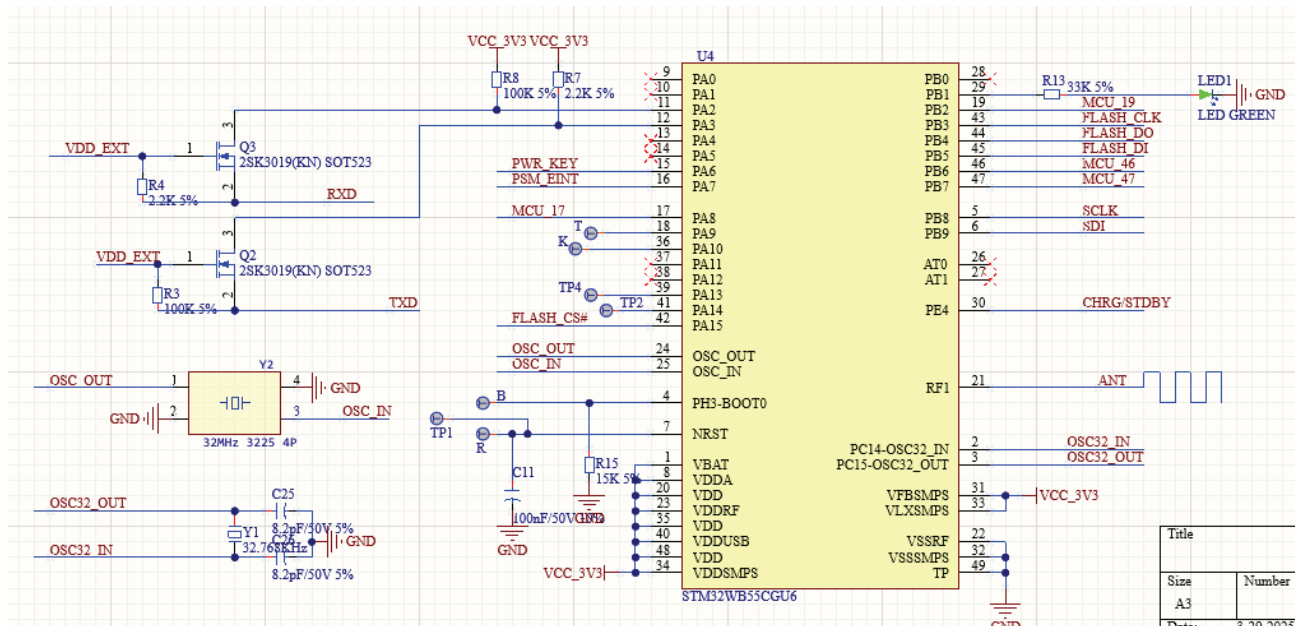
Figure 2 Hardware Architecture Diagram



2.1 Core Processing Unit

The core processing unit mainly undertakes the core tasks of data scheduling and protocol parsing. STM32 is a high-performance, low-cost, and low-power 32-bit microcontroller unit developed by STMicroelectronics based on the ARM Cortex-M core. It is widely used in smart homes, wearable devices, automotive electronics, medical equipment, and other fields. The module uses STM32WB55CGU6 dual core microcontroller, with Cortex-M4 core running at a frequency of 64MHz, responsible for the main program and wireless protocol stack processing; The Cortex-M0+coprocessor is dedicated to low-power task scheduling^[5]. This chip integrates 16MB Flash and 128KB SRAM, and supports dual-mode communication of Bluetooth 5.2 and Zigbee 3.0. The STM32 minimum system includes power circuit, crystal oscillator circuit, reset circuit, etc. The circuit principle of the core processing unit is shown in Figure 3.

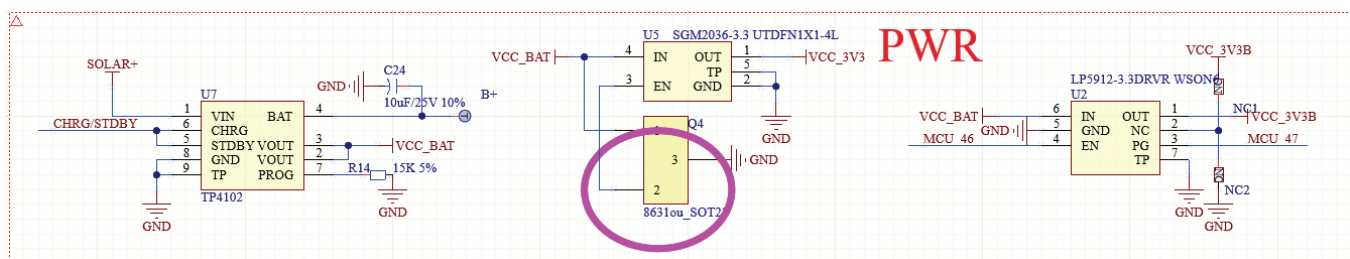
Figure 3 Schematic diagram of core processing unit circuit



2.2 Power Management Unit

The power management unit is the foundation for the stable operation of the system. The module charge and discharge management part adopts TP4102 chip, which supports a maximum charging current of 1A. Combined with UFX302020 solid-state lithium battery, it can suppress leakage current to below 1 μ A in sleep mode. The LP5912 low dropout linear regulator provides a stable output of 3.3V, with a noise suppression ratio exceeding 60dB, significantly reducing the interference of power ripple on sensitive circuits. To solve the problem of module endurance, the system is connected to MSCM solar panels, which can output 4.5V voltage and 14.8mA current under lighting conditions, achieving adaptive collection and storage of environmental energy^[6]. In order to reduce the size of the Hall element in the module, when the magnet is close to the module, the SGM2036 element's pin 3 enables the module to enter normal working state. The principle of the power circuit diagram is shown in Figure 4.

Figure 3 Schematic diagram of core processing unit circuit

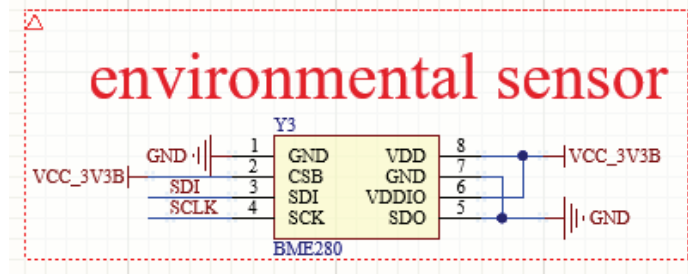


2.3 Sensor Array

The sensor array design aims to fuse multi-source data. The environmental perception part adopts BME280 sensor, and

its internal compensation algorithm effectively suppresses the temperature drift effect, which can synchronously obtain temperature, humidity, and air pressure parameters^[7]. The motion state detection adopts the MPU9250 nine axis sensor, and the combination of accelerometer, gyroscope, and magnetometer can accurately analyze the three-dimensional spatial posture^[8]. The ALS-DPDIC17-78C light intensity sensor provides wide range detection capability, and its I²C digital interface and 10ms fast response feature ensure the real-time and accuracy of light data. The schematic diagram of the BME280 sensor circuit is shown in Figure 5.

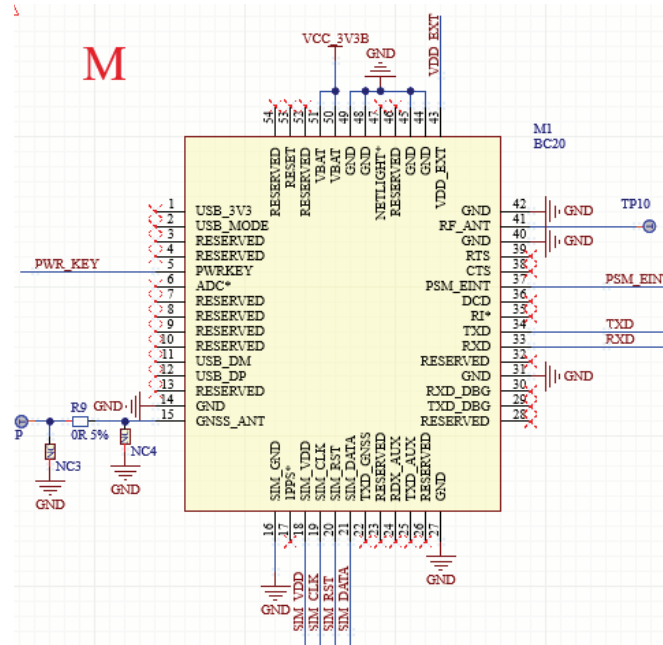
Figure 5 Schematic diagram of BME280 circuit



2.4 Wireless Communication Module

Wireless communication modules need to consider wide area connectivity and precise positioning requirements. The narrowband IoT chip adopts Shanghai Mobile Remote Communication BC20 module, which supports Cat M1/NB IoT dual-mode and B1/B3/B5/B8 multi frequency bands. The receiving sensitivity can reach -129dBm at a transmission power of 23dBm^[8]. The positioning system adopts a dual-mode solution of Beidou III and GPS, combined with CA-G01 navigation antenna, which can achieve a positioning accuracy of 2.5 meters in open environments and shorten the cold start time to less than 35 seconds. The communication antenna adopts an inverted F-shaped ceramic design, with a size of only 5×3mm², and a radiation efficiency of over 48% in the 900MHz frequency band. At the same time, in order to minimize module size as much as possible, the module adopts ESIM technology, which enables seamless access to global operator networks without the need for physical card slots. The BC20 circuit schematic is shown in Figure 6.

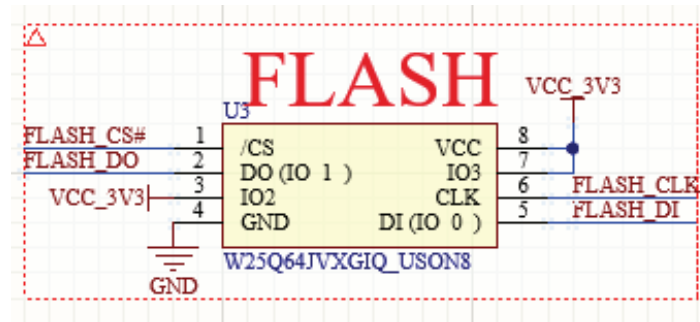
Figure 6 Schematic diagram of BC20 circuit



2.5 Storage Unit Design

The design of storage units focuses on considering data capacity and data security. The module storage unit adopts an external W25Q64JVXG IQ_SON8 to provide 16MB of storage space, supporting XIP mode and 128 bit AES encryption storage, with a data transfer rate of up to 100Mbps^[9]. The circuit diagram of the storage unit is shown in Figure 7.

Figure 7 FLASH Electrical Schematic Diagram



3. Embedded software design

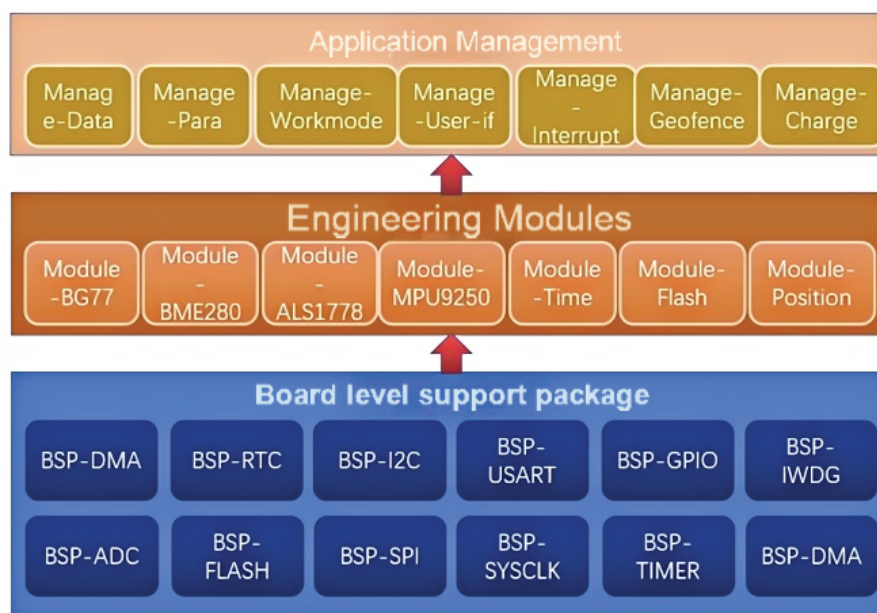
3.1 Development Environment Construction

The modular embedded software adopts a software architecture design that combines hierarchy and modularity. By using Keil MDK-ARM integrated development environment and STM32CubeMX configuration tool for project basic configuration, a full stack software solution was built from low-level hardware drivers to upper level application logic^[10]. Keil MDK-ARM integrated development environment is an embedded system development tool specifically designed for ARM series microcontrollers. The software is easy to learn and use, with powerful features that can meet most embedded applications.

3.2 Embedded Software Architecture

At the overall architecture level, the system is divided into three levels: application management layer, functional module layer, and board level support package, as shown in Figure 8.

Figure 8 Embedded software design diagram

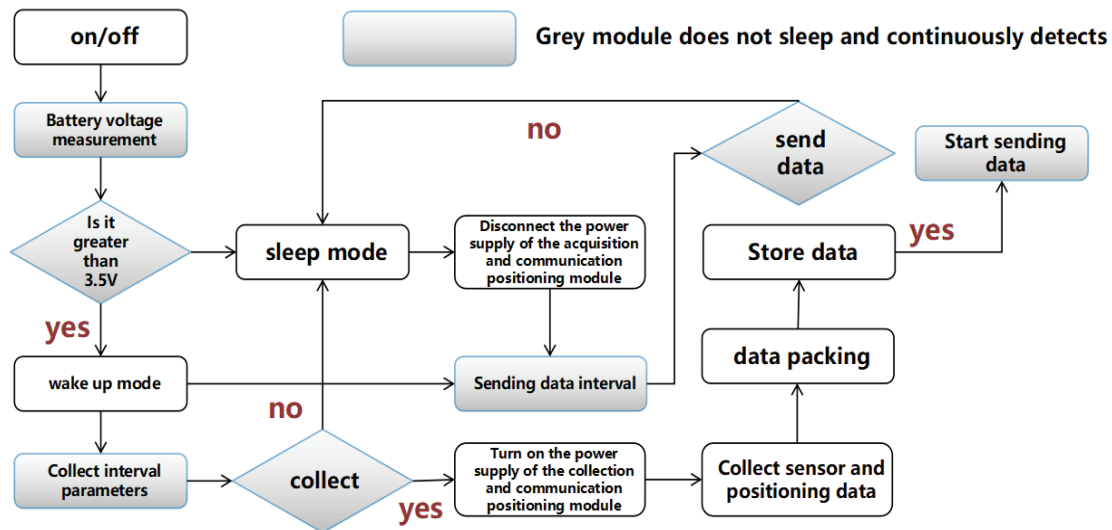


3.3 Main Program Design

After setting up the work environment, use C language to write embedded code, compile the code and burn it into the main control chip. The workflow of the module when it is powered on and running is mainly as follows:

- (1) Module embedded program initialization, initialization of system clock peripherals and external module initialization;
- (2) Battery voltage measurement, voltage greater than 3.5V, normal operation; When the voltage is less than 3.5V, enter sleep mode;
- (3) Collect data from temperature and humidity sensors, illuminance sensors, nine axis sensors, etc. at regular intervals. The default collection interval is 1 hour, and the time interval can be customized. The sensor data is transmitted to the microcontroller for analysis and processing;
- (4) Determine whether the data transmission interval, data packaging, and data transmission are met;

Figure 9 Program flowchart

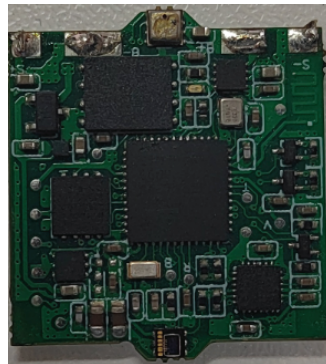


4. Experimental verification

4.1 Hardware physical display

The core hardware achievement of this study is a self-designed narrowband IoT module. The physical size of the module is 20mm × 20mm × 5mm, with a weight of only 7g. Through 3D printing of the packaging shell and IP67 protection design, it meets the deployment needs of complex scenarios such as outdoor and industrial environments. The physical display of the module is shown in Figure 10.

Figure 10 Physical display of module



4.2 Module software and hardware debugging

The core processing unit, power management unit, sensor array, wireless communication module, and storage unit of the system are soldered onto the PCB. Before powering on the module, the hardware soldering should be checked for short circuits, virtual soldering, and polarity according to the PCB design diagram. Perform programming and unit testing based on the schematic diagram to verify if the module hardware is functioning properly. According to the system functional requirements, complete the module embedded code programming, compile and burn it into the system.

4.3 Module Performance Testing

To verify the stability of module operation, perform module performance testing. The cloud platform for this test is Alibaba Cloud platform, and the test results show that the module voltage detection range covers 3.0-4.2V, accurately reflecting the charging and discharging characteristics of lithium batteries. The temperature and humidity sensor achieves accuracy of ± 0.3 °C and ± 3% RH within the range of -20~60 °C and 0-100% RH, respectively; The positioning accuracy is 3.5 meters; The atmospheric pressure detection module supports wide range measurement from 300-1100 hPa, adapting to diverse needs from low altitude plains to high-altitude areas. The experimental data validated the stability and environmental perception accuracy of the module, providing reliable support for multi scenario deployment.

5. Conclusion

This article is based on narrowband IoT and STM32 technology, and designs and implements a lightweight multi scene module. The module weighs 7g and has a volume of $2\text{cm} \times 2\text{cm} \times 0.5\text{cm}$, which is an ultra micro volume. The system functions normally and meets the design requirements. The module can be widely used in ecological monitoring, smart cities and other fields, and has good application value.

Funding

no

Conflict of Interests

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

- [1] Li Huqun, Zhu Xiaomang, Wang Fang, Wu Feilong, Su Leilei Research and Implementation of Smart Campus IoT Door Lock System Based on NB IoT [J]. IoT Technology, 2025, 15 (07): 71-73+80.
- [2] Li Xingli, Liu Dongxu, Gu Bingling, peaceful A new intelligent system for electric vehicle charging piles based on NB IoT technology [J]. Bonding, 2025, 52 (03): 146-149.
- [3] Zhou Yuanju Design of Wireless Hydrological Monitoring System Based on NB IoT [J]. Internet of Things Technology, 2025, 15 (05): 9-11.
- [4] Li Huqun, Zhu Xiaomang, Wang Fang, Wu Feilong, Su Leilei Research and Implementation of Smart Campus IoT Door Lock System Based on NB IoT [J]. IoT Technology, 2025, 15 (07): 71-73+80.
- [5] Tang Yingjie, Chang Xiangxiang, Fan Changsheng Design of Kitchen Gas Security Monitoring and Alarm System Based on STM32 [J]. Instrument Technology, 2025, (02): 22-24.
- [6] Chen Hongru, Fan Zai, Fan Changsheng Design of Automatic Water Flow Control System Based on STM32 [J]. Instrument Technology, 2025, (02): 38-40.
- [7] Ma Qiufang Development of Thermal Control Billing Instrument Based on STM32 [J]. Instrument Technology, 2025, (02): 19-21.
- [8] Chen Xiaodan, Li Haoyuan, Zeng Weichao, Lin Jiarui, Chen Xinxin, Zhang Min Research on Human Health Monitoring System Based on Ultra Wideband Radar Deep Learning [J]. Instrumentation Technology, 2025, (02): 58-62
- [9] Liu Jie, Yuan Liangwei, and Yin Xuetao Design of Unmanned Self Balancing Bicycle Based on STM32. "Industrial Control Computer 38 03 (2025): 167-168.
- [10] Jia Rongcong Design and Implementation of Vacuum Cleaning Robot Based on STM32. "Industrial Control Computer 38 03 (2025): 169-170.

Analysis of the Development Path of Artificial Intelligence Technology in the Big Data Era

Yusen Huang*

Santiago Canyon College, California, 92869, USA

**Corresponding author: Yusen Huang*

Copyright: 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: This paper systematically analyzes the development path of artificial intelligence technology in the big data era, and explores the integration mechanism of the two and their role in promoting technological progress. By sorting out the definition and characteristics of big data, the basic concepts of artificial intelligence technology and the interactive relationship between the two, the study reveals how the massive resources provided by big data for artificial intelligence drive breakthroughs in deep learning, natural language processing and other fields. At present, artificial intelligence technology has made significant progress in model optimization and cross-modal fusion, but it still faces challenges such as high computing costs and privacy security. In the future, efficient algorithms, edge intelligence and ethical governance will become the focus of development, and are expected to achieve greater value in medical care, education and urban management. This study provides theoretical reference and practical guidance for the optimization and application of artificial intelligence technology.

Keywords: Big Data; Artificial Intelligence; Technology Integration; Development Trend; Ethical Challenges

Published: May 8, 2025

DOI: <https://doi.org/10.62177/jaet.v2i2.340>

1.Introduction

With the rapid development of information technology, big data and artificial intelligence (AI) have become the core driving force for promoting social progress and economic transformation. The explosive growth of big data has provided artificial intelligence technology with rich training resources, enabling it to achieve significant breakthroughs in image recognition, natural language processing, autonomous driving and other fields. For example, the application of convolutional neural networks (CNNs) in 3D face reconstruction demonstrates the powerful potential of data-driven models ^[1]. Big data, with its massiveness, diversity, and high speed, provides a basis for optimizing AI algorithms and solving complex problems, while AI further explores the potential value of big data through technical means such as deep learning and reinforcement learning ^[2]. The two complement each other and jointly promote the comprehensive upgrade of intelligence. However, the era of big data also brings challenges such as data privacy, algorithm bias, and computing resource bottlenecks. These problems have put forward new requirements for the development path of AI technology. Therefore, systematically analyzing the integration mechanism and development laws of big data and AI will not only help understand the current technology ecology, but also point out the direction for future technological breakthroughs. This study focuses on this and aims to provide theoretical reference and practical guidance for academia and industry, which has important academic value and practical significance. The interaction between big data and AI has also profoundly affected many fields such as social governance, medical health, and financial services. For example, the application of big data-based predictive models in medical diagnosis has significantly

improved the accuracy of disease screening^[3]. However, with the advancement of technology, data security and ethical issues have become increasingly prominent, and it is urgent to address them through the coordination of technology and policy. Therefore, exploring the development path of AI technology in the era of big data is not only a technical demand, but also an inevitable choice to cope with social challenges.

2. Overview of big data and artificial intelligence technology

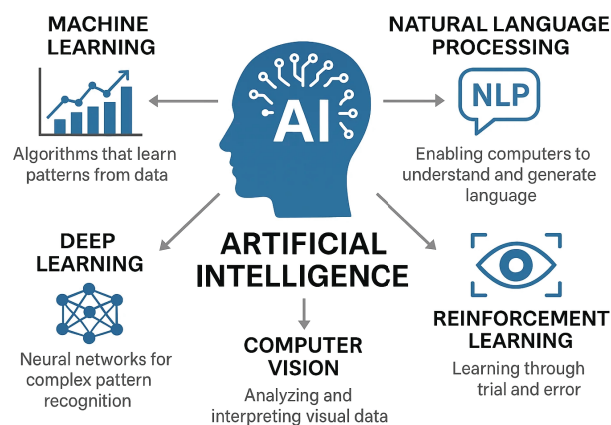
2.1 Definition and characteristics of big data

Big data refers to a collection of data with huge volume, complex types, fast generation speed and difficult to effectively analyze using traditional data processing methods^[4]. Its core characteristics include large volume, multiple types, fast speed and high value. Large volume is reflected in the exponential growth of data scale, such as social media, IoT devices, etc., which generate massive amounts of data every day; multiple types refer to the diverse forms of data, including structured data such as database records, unstructured data such as images and videos, and semi-structured data such as log files; fast speed reflects the real-time nature of data generation and processing requirements, such as financial transactions or online recommendation systems; high value emphasizes the insights mined through deep analysis, which can drive decision optimization and business innovation^[5]. These characteristics of big data provide rich raw materials for intelligent applications, especially in the fields of pattern recognition and trend prediction, which greatly expands the boundaries of technology application. However, the complexity of big data also places higher demands on storage, computing and analysis capabilities, prompting the continuous innovation of the technology system to cope with the opportunities and challenges brought by the data flood.

2.2 Basic concepts of artificial intelligence technology

Artificial intelligence technology refers to a collection of technologies that enable computer systems to perform tasks such as perception, reasoning, learning and decision-making by simulating human intelligent behavior (Figure 1). Its core branches include machine learning, deep learning, natural language processing, computer vision and reinforcement learning^[6]. Machine learning extracts patterns from data through algorithms to drive prediction and classification tasks; deep learning uses neural networks to handle complex nonlinear problems and performs well in image and speech recognition; natural language processing gives machines the ability to understand and generate human language; computer vision focuses on the analysis of image and video content; reinforcement learning optimizes decision-making strategies through trial and error and is suitable for dynamic environments. The essence of artificial intelligence is to automate intelligent behavior based on data and algorithms. Its development relies on powerful computing power and diverse training data, aiming to solve complex problems and improve efficiency and accuracy in various fields^[7].

Figure 1 Basic concepts of artificial intelligence technology



2.3 Integration and interaction between big data and artificial intelligence

The integration of big data and artificial intelligence is an important driving force for technological progress. The two have formed a dynamic relationship of mutual promotion. Big data provides artificial intelligence with a huge amount of training materials, enabling algorithms to learn complex patterns from diverse scenarios, thereby improving the generalization ability of the model^[8]. For example, recommendation systems achieve accurate personalized services by analyzing user behavior

data. In turn, artificial intelligence uses efficient data processing and analysis technologies to explore the hidden value in big data and solve scale and complexity problems that traditional methods cannot cope with. This interaction not only accelerates technology iteration, but also promotes cross-domain innovative applications, such as disease prediction in smart healthcare and traffic optimization in smart cities^[9]. However, the integration process also faces problems such as uneven data quality and insufficient algorithm transparency, which require the coordinated solution of technology and management methods to maximize the advantages of both.

3. Development path and trend of artificial intelligence technology in the big data era

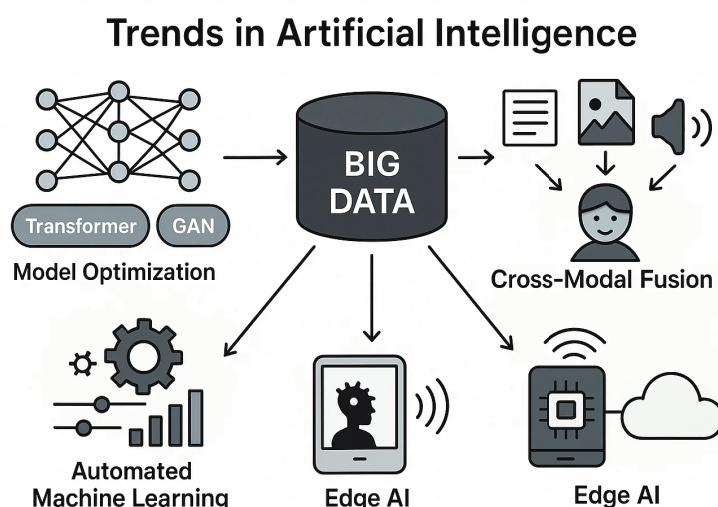
3.1 Current status and breakthroughs in technology development

Driven by big data, artificial intelligence technology has made significant progress in recent years, especially in the fields of deep learning, natural language processing and computer vision. Deep learning processes massive amounts of data through large-scale neural networks, reaching or even surpassing human performance in tasks such as image recognition and speech transcription, such as the widespread application of face recognition systems. Breakthroughs in natural language processing technology in machine translation and dialogue systems have made cross-language communication and intelligent customer service a reality, such as the popularization of multilingual real-time translation tools^[10]. Computer vision has shown great potential in autonomous driving and medical image analysis, such as assisting disease diagnosis through scanning images. The volume and diversity of big data provide the basis for these breakthroughs, enabling models to learn more sophisticated patterns from complex data. However, current technology is still limited by high computing costs and data dependence, and the robustness and interpretability of models in some scenarios are insufficient. These current situations show that artificial intelligence has entered a rapid development stage driven by big data, but technical bottlenecks still need to be overcome to achieve wider application.

3.2 Main Development Trends and Directions

Artificial intelligence technology has shown a diversified development trend in the big data era. First, the optimization of model architecture has become a focus, and new algorithms such as Transformer and generative adversarial networks continue to improve performance and efficiency. Second, automated machine learning technology is emerging, which reduces the development threshold by automatically adjusting parameters and selecting features, and promotes the popularization of artificial intelligence. Third, cross-modal fusion has become a hot topic, combining multimodal data such as text, images, and voice to achieve more comprehensive intelligent perception. For example, smart assistants can understand voice commands and visual environments at the same time. In addition, edge intelligence has gradually attracted attention, reducing dependence on the cloud by running lightweight models on the device side, improving real-time performance and privacy protection. These trends show that artificial intelligence is evolving in the direction of efficiency, universality, and multi-dimensionality, striving to achieve more flexible and intelligent application scenarios with the support of big data.

Figure 2 Development direction of artificial intelligence



3.3 Potential and Challenges of Future Technologies

Artificial intelligence has great potential in the era of big data and is expected to further reshape all walks of life. In the medical field, accurate diagnosis and personalized treatment plans may greatly improve the level of health management; in urban management, intelligent transportation and energy optimization based on real-time data will improve resource utilization efficiency; in the field of education, adaptive learning systems can provide students with customized educational experiences. However, these potentials are accompanied by significant challenges. First, data privacy and security issues are becoming increasingly serious, and large-scale data collection may trigger a crisis of user trust. Second, algorithmic bias may be exacerbated by the imbalance of training data, affecting the fairness of decision-making. In addition, the demand for computing power for high-performance artificial intelligence continues to rise, and energy consumption and environmental impact cannot be ignored. Future technological breakthroughs need to balance efficiency and ethics, explore models with strong interpretability and low resource consumption, and jointly address privacy and fairness issues through policies and technologies to ensure the sustainable development of artificial intelligence.

4. Conclusion and Prospects

4.1 Research Summary

This study systematically analyzed the development path of artificial intelligence technology in the era of big data and revealed the mechanism and significance of the integration of the two. Big data, with its massiveness, diversity and real-time nature, provides a key driving force for artificial intelligence, and promotes breakthroughs in deep learning, natural language processing and other fields. Artificial intelligence mines data value through efficient algorithms and realizes a wide range of applications such as personalized recommendations and precision medicine. Studies have found that current technologies have made significant progress in model performance and application scenarios, but still face limitations such as high computing costs and insufficient interpretability. In terms of development trends, model optimization, automated machine learning and multimodal fusion are leading technological innovation, while edge intelligence has expanded the application boundaries. Overall, the deep interaction between big data and artificial intelligence has not only accelerated technological iteration, but also injected new vitality into the social economy, laying a solid foundation for subsequent research and practice.

4.2 Outlook on future development paths

In the future, artificial intelligence will make further breakthroughs with the support of big data, deeply penetrate into fields such as medical care, education and urban management, and help accurate decision-making and resource optimization. At the technical level, efficient algorithms and low-power models will become the focus to cope with computing bottlenecks and environmental challenges. At the same time, privacy protection and algorithm fairness need to be solved through the coordination of technology and policy to ensure that the benefits of technology benefit a wider group of people. Interdisciplinary collaboration will also promote the development of explainable artificial intelligence and enhance public trust. Looking ahead, the integration of artificial intelligence and big data will continue to drive intelligent transformation and shape a more efficient and fair digital society.

Funding

no

Conflict of Interests

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

References

- [1] Zhang, F., Liu, Q.: Innovative Analysis of Student Management Path Based on Artificial Intelligence and Big Data Integration. *International Journal of e-Collaboration*, pp. 1-10 (2024).
- [2] Luo, F.: Industry Iterative Transformation Based on Big Data Intelligent Platform. *Advances in Computer, Signals and Systems*, pp. 1-15 (2023).

- [3] Shandilya, S., Sangwan, S., Yadav, R.: Artificial Intelligence and Big Data. 2015 IEEE International Conference on Artificial Intelligence for Industries (AI4I), pp. 1-12 (2015)
- [4] Ostrowski, D.: Artificial Intelligence with Big Data. 2018 First International Conference on Artificial Intelligence for Industries (AI4I), pp. 125-126 (2018).
- [5] Qiu, B., Zhu, Y., Du, L., Feng, C.: Analysis and Reflection on the Teaching Application of Artificial Intelligence Technology in the Context of Big Data. Curriculum and Teaching Methodology, pp. 1-15 (2024).
- [6] Kumar, P., Singh, L.: Analysis of the Impact of Artificial Intelligence on Big Data Analysis. International Journal of Engineering Science and Humanities, pp. 1-15 (2024).
- [7] Kersting, K., Meyer, U.: From Big Data to Big Artificial Intelligence? KI - Künstliche Intelligenz, vol. 32, pp. 3-8 (2018).
- [8] Soni, R., Baghel, A., Paliya, S., Mamtani, R., Gupta, L.: Connection of Big Data Analytics & Artificial Intelligence. 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-6 (2023).
- [9] Johri, S., Rawal, K. V., Aishwarya, B., Singh, N., Shaaker, A. M., Revathi, V.: Big Data and Artificial Intelligence: Revolutionizing Business Decision-Making. 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), pp. 1689-1693 (2023).
- [10] Bag, S., Gupta, S., Kumar, A., Sivarajah, U.: An Integrated Artificial Intelligence Framework for Knowledge Creation and B2B Marketing Rational Decision Making for Improving Firm Performance. Industrial Marketing Management, vol. 92, pp. 178-189 (2021).

Research on APP Design for Albinism Family Psychological Therapy Based on User Needs

Fan Wu¹, Jiaming Zeng¹, Congrong Xiao^{2*}

1. Software Engineering Institute of Guangzhou, Guangzhou, 510000, China

2. Guangdong Baiyun University, Guangzhou, 510000, China

*Corresponding author: Congrong Xiao

Copyright: 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Albinism is a rare genetic disorder caused by genetic mutations that result in impaired melanin synthesis, often accompanied by significant skin, hair, and eye pigment loss. Due to the unique appearance characteristics, individuals with albinism and their families often face prejudice, discrimination, and isolation from society, which not only affects their physical and mental health, but also causes long-term and profound psychological pressure on their family members. Research has shown that families with sick children generally suffer from psychological problems such as anxiety, depression, self blame, and feelings of isolation, and there is an urgent need for professional and systematic psychological support and intervention measures. However, current psychological healing resources for families with albinism are extremely limited, and related service platforms lack specificity, accessibility, and continuity, making it difficult to meet the actual needs of this special group. With the rapid development of mobile Internet and digital health technology, the application of APP in mental health intervention has gradually become an important trend. By relying on mobile applications, we can provide users with readily accessible, personalized, and privacy protected psychological therapy.

Keywords: Albinism; Psychological Therapy; User Demand; App Design

Published: May 12, 2025

DOI: <https://doi.org/10.62177/jaet.v2i2.314>

1.Introduction

Albinism is a rare genetic disorder caused by genetic mutations that result in impaired melanin synthesis, often accompanied by significant skin, hair, and eye pigment loss^[1]. Due to the unique appearance characteristics, individuals with albinism and their families often face prejudice, discrimination, and isolation from society, which not only affects their physical and mental health, but also causes long-term and profound psychological pressure on their family members. Research has shown that families with sick children generally suffer from psychological problems such as anxiety, depression, self blame, and feelings of isolation, and there is an urgent need for professional and systematic psychological support and intervention measures. However, current psychological healing resources for families with albinism are extremely limited, and related service platforms lack specificity, accessibility, and continuity, making it difficult to meet the actual needs of this special group. With the rapid development of mobile Internet and digital health technology, the application of APP in mental health intervention has gradually become an important trend. By relying on mobile applications, we can provide users with readily accessible, personalized, and privacy protected psychological therapy.

2. Research on User Psychological Analysis

Stigma of Illness refers to a series of psychological and behavioral reactions, such as shame, inferiority, exclusion, and social isolation, caused by individuals or families being labeled negatively by social groups due to illness^[2]. Erving Goffman first systematically elucidated the social psychological mechanism of disease shame in his book “Stigma: Notes on Managing Impaired Identity”, pointing out that disease is not only an abnormality in physical condition, but also evolves into the “stigma” of individual identity in social cognition. In the field of rare diseases, the phenomenon of disease shame is particularly prominent, and patients and their family members often experience stronger social exclusion and psychological pressure due to differences in appearance, social misunderstandings, or lack of information.

2.1 Characteristics of psychological needs of albino families

In families with albinism, shame of illness exhibits unique and complex characteristics. Due to the obvious physical features of albinos, they often attract the attention, misunderstanding, and even discrimination of bystanders. Family members (especially parents and direct relatives) are prone to internalizing social biases, forming emotional experiences of self blame, shame, or avoidance. This sense of shame not only exacerbates the psychological burden within the family, but also weakens their enthusiasm for seeking social support and medical resources, further plunging them into a situation of isolation and helplessness.

From the perspective of psychological needs, families with albinism urgently need to break the negative cognitive cycle caused by shame and rebuild their sense of self-esteem and social identity. Effective psychological therapy interventions should focus on the deep needs of family members in emotional adjustment, social adaptation, and information cognition, helping them understand the scientific nature of the disease, resist external biases, and enhance their self acceptance and social communication abilities. Therefore, when designing a psychological therapy app for families with albinism, the influence mechanism of shame theory should be fully considered, and targeted functional modules such as emotional counseling, mindfulness training, peer assistance, and knowledge dissemination should be established to achieve comprehensive and progressive psychological support^[3].

2.2 Family System Theory

The Family Systems Theory was proposed by psychologist Murray Bowen, emphasizing that family members are a interconnected and interacting whole system^[4]. According to this theory, an individual's emotions and behaviors are not only a reflection of their personal traits, but also deeply influenced by the internal relationship structure, interaction patterns, and emotional communication methods within the family system. When a member of the family experiences a special problem, such as a major illness, the entire family system will be impacted and make corresponding adaptive or defensive responses.

In families with albinism, the rare disease characteristics of patients often become an important source of stress within the family system, affecting the emotional state, role division, and intimate relationships of family members. Research has shown that in the face of disease pressure, families are prone to the following typical response patterns: overprotection, avoidance of responsibility, emotional alienation, or escalation of conflicts. If these reactions cannot be effectively regulated, they will further deteriorate the family atmosphere, increase the psychological burden on family members, and form a vicious cycle.

From the perspective of needs analysis, albino families not only need support for individual psychological issues, but also need to promote emotional communication, role identification, and resource sharing among members from the perspective of the family as a whole. Effective psychological therapy interventions should be based on the theory of family systems, helping families establish a positive interaction mechanism, such as improving parenting education, strengthening family cohesion, and supporting emotional adjustment within the family. Therefore, in app design, it is necessary to introduce a family system perspective and set up functional modules such as family emotional exercises, parent-child communication guidance, and family task management to enhance the systematicity and overall effectiveness of the healing process, truly achieving comprehensive support for albino families.

2.3 Deduction of Core Needs for Psychological Healing

2.3.1 Security Requirements

In the process of psychological therapy for families with albinism, a sense of security is an important prerequisite for users to

actively participate in healing activities and continue using the platform. According to Maslow's hierarchy of needs theory, safety needs are located within basic psychological needs and encompass individuals' basic expectations for their privacy, dignity, and personal security. For individuals with albinism and their families, the social prejudice and shame caused by the characteristics of the disease make them more sensitive and defensive when seeking psychological support, particularly emphasizing the protection of personal information and the anonymity of the communication environment.

Anonymity, as an important manifestation of security requirements, is a key factor in reducing user psychological vigilance and promoting authentic expression. If users can ensure that their identity information is not leaked during open communication, self disclosure, or emotional release, it can significantly reduce shame and exposure anxiety, and enhance their willingness to participate in healing activities. At the same time, privacy protection is also an indispensable basic element in the design of psychological healing apps. It is necessary to ensure that users' personal information, communication records, psychological assessment data, and other information are encrypted and strictly protected to avoid the risk of leakage and safeguard users' trust in the platform.

Therefore, when designing a psychological therapy app for families with albinism, safety needs must be given top priority. Specific measures include providing anonymous registration and usage mechanisms, strengthening data encryption technology, setting up privacy protocol prompts, allowing users to independently control the scope of information disclosure, and establishing secure exit functions. By creating a safe, private, and controllable healing environment, we can effectively break down users' psychological defense barriers and promote their active and continuous participation in the healing process on the platform.

2.3.2 Attribution Requirements

Sense of belonging is an important support for individual mental health and is also the core content of Maslow's hierarchy of needs theory's social needs hierarchy^[5]. For individuals with albinism and their families, long-term exposure to social prejudice, isolation, and misunderstanding, as well as a lack of sense of belonging, can easily lead to negative psychological states such as loneliness, helplessness, and self denial. Therefore, rebuilding a sense of belonging and promoting recognition and support for patients and family members within the group are essential core needs in the process of psychological healing.

The establishment of patient communities is an important path to meet the need for belonging. By building an online community platform exclusively for albinos and their families, users can establish emotional connections, share experiences, support each other, and reduce the impact of isolation and shame. Patient communities not only provide users with a space for emotional release and empathetic communication, but also stimulate the motivation to actively respond to disease challenges and enhance psychological resilience through the role model effect and mutual assistance mechanism.

At the same time, family collaboration is also an important dimension in the need for belonging. As the most direct support system for patients, the quality of communication and collaboration among family members directly affects the effectiveness of treatment. APP design should encourage family members to participate in the healing process together, such as setting up functions such as family task check-in, emotional recording and mutual assistance, parent-child communication training, etc., to promote emotional connection and support network construction within the family, form a small "belonging system", and help patients and families grow and recover together.

Therefore, in the design of psychological healing apps, the need for belonging should be fully considered. By building a collaborative mechanism between patient communities and families, a warm and belonging digital environment should be created, allowing users to receive continuous companionship, understanding, and support psychologically, further enhancing the depth and effectiveness of healing.

2.3.3 Self actualization needs

Self actualization needs are located at the highest level of Maslow's hierarchy of needs theory, referring to the internal motivation of individuals to pursue personal potential and value realization after meeting basic survival, safety, and belonging needs. For individuals with albinism and their family members, gradually establishing self-identity and actively facing life challenges after experiencing initial emotional fluctuations and social adaptation becomes a deep-seated goal in the process

of psychological healing. Therefore, meeting self actualization needs is an important way to promote long-term psychological recovery and improve quality of life.

Psychological education, as an important means of achieving self growth, can help users deepen their understanding of albinism related knowledge, master psychological strategies for coping with stress and negative emotions, and enhance their self-awareness and emotional regulation abilities. Through systematic and hierarchical psychological education content, users can not only correct misunderstandings about diseases and reduce shame, but also gain cognitive growth, enhance psychological resilience and autonomous coping abilities.

Empowering tools are the key to helping users realize their self-worth at the practical level. By setting up functional modules such as emotion management exercises, mindfulness meditation guidance, self goal setting and tracking, and small achievement recording, the app can provide users with actionable psychological self-help paths, inspiring them to actively practice psychological adjustment skills in daily life and gradually achieve a transition from passive acceptance to active healing. The existence of empowerment tools not only meets the individual's need for autonomy in the process of psychological healing, but also promotes lasting and deep psychological growth.

Therefore, in the design of psychological therapy apps for families with albinism, full attention should be paid to self actualization needs. By building a systematic psychological education system and diversified empowerment tools, users can explore their potential, achieve psychological independence and positive living goals, and ultimately promote the transition from emotional healing to self growth.

2.4 Theoretical Model of Requirement Priority

In the design process of psychological therapy apps, facing multi-level and multi-dimensional user needs, determining the priority of needs reasonably is a key step to ensure the effectiveness of product functions and optimize user experience. Based on Maslow's hierarchy of needs theory and practical research results, this study divides the psychological healing needs of albino families into three levels: safety needs, belonging needs, and self actualization needs. Based on their importance and implementation order, a needs priority theory model is constructed.

Firstly, security requirements are at the most fundamental and prioritized level of satisfaction. For families with albinism, ensuring anonymity and privacy security is a prerequisite for users to dare to enter the platform, open their hearts, and participate in interactions. If there is a lack of sufficient guarantee for a sense of security, other levels of needs will be difficult to effectively expand. Therefore, APP design should first invest key resources and development efforts in information security protection, anonymous usage mechanisms, and privacy management functions.

Secondly, as a mid-level requirement, the need for belonging becomes a key factor affecting users' continued use of the app and the formation of emotional stickiness once they have gained a basic sense of security. By building patient communities and promoting family collaboration, the app can help users establish emotional connections and support networks, alleviate feelings of loneliness and isolation, and further enhance the platform's activity and user loyalty.

Finally, as a high-level developmental need, self actualization needs may not constitute a usage threshold in the early stages, but they are of great significance in the later stages of user psychological healing and personal growth. By continuously building a psychological education system and empowering tools, users can achieve cognitive enhancement, emotional regulation, and self-worth exploration, which can effectively prolong their life cycle and promote the depth and breadth of psychological rehabilitation.

In summary, based on the priority theory model of demand hierarchy, this study proposes that the order of demand implementation for albino family psychological therapy APP should follow the logical path of "safety guarantee → belonging establishment → self realization". By gradually meeting the psychological needs of different levels, the user experience can be effectively improved, and the systematicity and sustainability of psychological therapy can be enhanced, providing theoretical basis for subsequent functional design and service strategies.

3.Design Strategy for Psychological Healing APP

3.1 Inclusive Design

The principle of inclusive design is crucial when designing psychological therapy apps for families with albinism^[6]. Inclusive design emphasizes providing equal, accessible, and user-friendly experiences for all user groups under different physiological, psychological, and cognitive conditions. Considering that albino patients generally have a certain degree of visual impairment, such as low visual acuity, sensitivity to strong light, and abnormal color vision, APP design needs to be optimized in terms of visual adaptation, color usage, interface layout, etc., to ensure their convenience and comfort in use.

In terms of visual impairment adaptation, interface design should focus on clear information hierarchy, sufficient contrast between text and background, moderate and adjustable font size, avoiding the use of low contrast or high-frequency flickering elements, and reducing visual fatigue^[7]. In addition, strong light stimuli such as large areas of pure white background or high brightness color blocks should be minimized as much as possible to prevent visual discomfort. At the same time, the interface interaction process should be concise and intuitive, reducing complex operation steps, improving usability, and helping visually impaired users smoothly complete various functional operations.

In terms of color strategy, the overall visual tone of the app is recommended to be mainly orange yellow. The orange yellow color scheme not only has good visibility and recognition, which can adapt to the sensitivity of albinos to light, but also has a warm, positive, and healing emotional symbol, which helps to convey a caring and supportive psychological atmosphere. The application of orange yellow color in the interface should be paired with soft and neutral auxiliary colors to ensure overall visual comfort and coordination, avoiding colors that are too bright or fatigued. At the same time, the application of color in functional areas should be combined with color block partitioning, icon prompts, and text assistance to improve the readability and operational accuracy of the interface through multi-channel information prompts.

In summary, based on the principle of inclusive design, through visual barrier adaptation and reasonable color usage, not only can the user experience of albinism patients and their family members be improved, but also the brand concept of caring and inclusiveness of the APP can be conveyed, further enhancing users' psychological security and sense of belonging, laying the foundation for the effective development of subsequent psychological healing functions.

3.2 Lowering the threshold for use

Barrier free design is an important principle in the design of psychological therapy apps to increase users' willingness to engage and continue using them for the first time^[8]. Especially for people with albinism and their family groups, they already bear a lot of physical, psychological, and social pressure in their daily lives. If the APP operation process is cumbersome and complex, it can easily cause cognitive burden, reduce user experience, and even lead to user churn. Therefore, simplifying the operation process and reducing learning costs are the fundamental guarantees for ensuring the effective transmission of healing functions. In terms of simplifying the operation process, the APP should follow the Minimum Steps Principle, which means using concise and intuitive interactive design as much as possible to help users complete the required operations with the least number of clicks and thoughts. For example, for commonly used functions such as emotion recording, community communication, and psychological exercises, a one click direct entrance should be set up to reduce hierarchical jumps; When users first use it, necessary information can be presented in stages through a guided interface (Onboarding) to avoid information overload.

In addition, the interface layout should follow the principles of consistency and predictability, maintaining clear logic and stable positioning of functional modules, allowing users to form operational habits in a short period of time, and reducing learning difficulty. When entering content, try to minimize the burden of free input and use simple interactive methods such as multiple-choice, slider, and icon clicking to improve filling efficiency and operational comfort. In order to further reduce the threshold for use, the APP should also support personalized and simplified settings, such as intelligently recommending commonly used functions based on user usage frequency, or providing custom shortcut entrances on the interface to help users with different needs optimize their operation paths according to personal habits, truly achieving a "user centered" convenient experience.

In summary, by simplifying the operation process and optimizing the interface interaction, the psychological therapy APP can not only meet the actual usage needs of the albino family user group, but also effectively improve the overall user experience, reduce psychological and cognitive burden, and thus assist in the smooth achievement of psychological therapy goals.

3.3 Emotional Design

Emotional design emphasizes that products should not only meet functional needs, but also actively connect with users on an emotional level, evoking joy, trust, and a sense of belonging^[9]. For albino family psychological therapy apps, a good emotional experience is an important factor in enhancing user stickiness and promoting the smooth progress of psychological therapy. Therefore, in the design process, Norman's three-level theory should be systematically introduced, starting from the instinctive level, behavioral level, and reflective level, to construct a comprehensive emotional design system. At the intuitive level, design should immediately evoke a sense of pleasure and security in users through sensory elements such as visual and auditory senses. For example, the interface color is mainly warm orange yellow, with soft auxiliary tones to create a warm and inclusive visual atmosphere; The interface animation and prompt sound effects should be gentle and soothing, avoiding abruptness and stimulation, and enhancing the overall sensory experience's friendliness.

At the behavioral level, design should focus on the user's operational experience and functional feedback during use. The interface interaction should be intuitive and smooth, with timely and accurate feedback, allowing users to feel a sense of control and achievement during use. For example, after completing an emotional recording, providing positive psychological feedback in a timely manner through warm animated feedback or small reward prompts can enhance the user's motivation to continue using. At the Reflective Level, design should inspire users to deeply identify with and emotionally resonate with the meaning of the product. Through elements such as emotional stories, user case sharing, and achievement medal systems, guide users to continuously perceive their own growth and changes during use, and establish emotional identification with the APP brand concept. At the same time, personalized ritual experiences can be set up, such as important date reminders (rehabilitation anniversaries, emotional improvement days), to help users generate positive self-identity and a sense of value during the review process.

In summary, through the emotional design strategy based on Norman's three-level theory, the psychological therapy app can establish a sustained and warm relationship with albino family users in three dimensions: sensory experience, behavioral interaction, and deep emotional connection, injecting positive emotional motivation into the psychological therapy process and enhancing overall healing effectiveness and user loyalty.

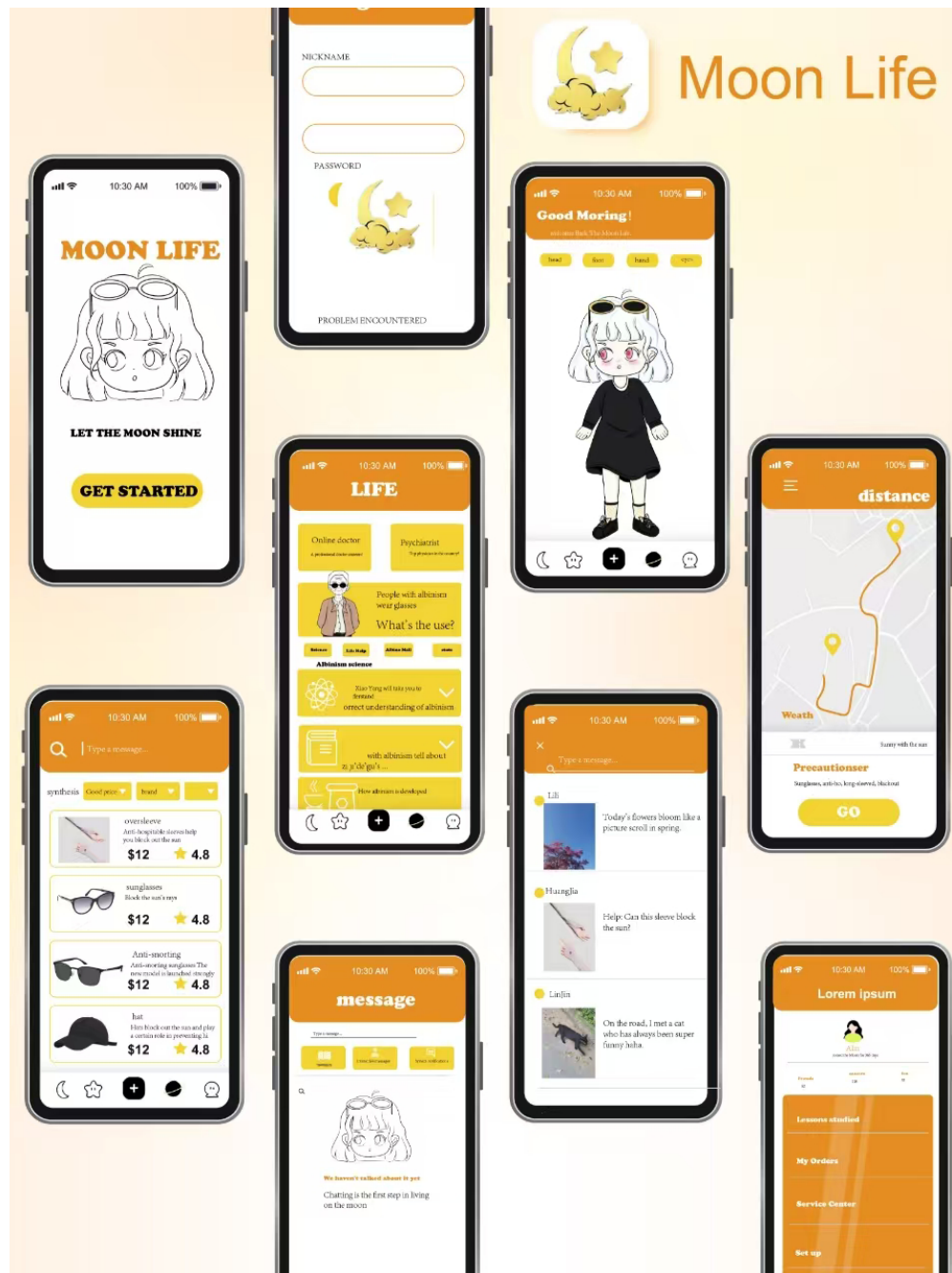
3.4 APP Design Practice

Based on the design strategies proposed in sections 3.1 to 3.3, the APP design practice for family psychotherapy for albinism combines inclusive design, low threshold interaction, and emotional design principles to construct a functional system centered on user needs. The interface adopts a high contrast layout and adjustable fonts to avoid strong light elements. The main color tone is orange yellow to meet the visual sensitivity needs of albino patients, supplemented by neutral and soft colors to ensure visual comfort. The functional modules are implemented around the core user requirements in a layered manner: ensuring security needs through anonymous registration, data encryption, and privacy protocols; Building patient community forums and family collaboration tools (such as shared task check-in and parent-child communication training) to meet the needs of belonging; Provide a structured psychological education resource library (disease knowledge, emotion regulation courses) and self empowerment tools (goal tracking, mindfulness meditation) to support self actualization needs. The specific design scheme is shown in Figure 1.

4. Conclusion

Based on the integration of user needs and design strategies, this psychological therapy app has built a multidimensional support platform with safety, belonging, and self realization as its core through inclusive visual design (high contrast interface, orange yellow tone adaptation for visual sensitivity), low threshold interaction process (simplified operation, preset input methods), and layered functional modules (anonymous security mechanism, patient community and family collaboration tools, psychological education resource library, and self empowerment tools). It combines emotional design elements (warm color tones, achievement feedback, personalized milestones) to enhance user participation and comprehensively respond to the multi-level needs of albino families in psychological rehabilitation, ensuring an effective balance between functional practicality and emotional connection.

Figure 1: APP Design for Families with Leukemia



Funding

Guangdong Provincial College Students' Innovation and Entrepreneurship Training Program Project (S202412618014)

Conflict of Interests

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

- [1] King, R. A. (1987). Albinism. *Neurocutaneous Diseases*, 311-325.
- [2] Corrigan, P. W., & Fong, M. W. (2014). Competing perspectives on erasing the stigma of illness: What says the dodo bird?. *Social science & medicine*, 103, 110-117.
- [3] Corrigan, P. W., & Wassel, A. (2008). Understanding and influencing the stigma of mental illness. *Journal of psychosocial nursing and mental health services*, 46(1), 42-48.
- [4] Broderick, C. B. (1993). *Understanding family process: Basics of family systems theory*. Sage.

- [5] Maslow, A., & Lewis, K. J. (1987). Maslow's hierarchy of needs. Salenger Incorporated, 14(17), 987-990.
- [6] Clarkson, P. J., Coleman, R., Keates, S., & Lebbon, C. (2013). Inclusive design: Design for the whole population.
- [7] Blair-Early, A., & Zender, M. (2008). User interface design principles for interaction design. *Design Issues*, 24(3), 85-107.
- [8] Miura, T., Yabu, K. I., Ikematsu, S., Kano, A., Ueda, M., Suzuki, J., ... & Ifukube, T. (2012, October). Barrier-free walk: A social sharing platform of barrier-free information for sensory/physically-impaired and aged people. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 2927-2932). IEEE.
- [9] Ho, A. G., & Siu, K. W. M. G. (2012). Emotion design, emotional design, emotionalize design: A review on their relationships from a new perspective. *The Design Journal*, 15(1), 9-32.

Synthesis of Cobalt Doping Titanium Dioxide Fiber Supported by Reduced Graphene Oxide for Photocatalytic Degradation of Toluene in Air

Guilan Gao¹, Xiran Zhang¹, Li Zhang¹, Shuai Chen^{1*}, Pinnan Li^{2*}

1. School of Minerals and Environmental Engineering, Shanghai Polytechnic University, Shanghai, 201209, China

2. UMR 8000 CNRS, Institut de Chimie Physique, Université Paris-Saclay, Orsay, 91400, France

*Corresponding author: Shuai Chen, chenshuai@sspu.edu.cn; Pinnan Li, pinnan.li@universite-paris-saclay.fr

Copyright: 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Titanium dioxide has been widely used for photocatalytic degradation of organic pollutants in air, while problems like low utilization rate of sunlight and easy recombination of photogenerated electrons and holes are the main drawbacks for its application. In this study, a combination of Co doping and graphene supporting was used to synthesize cobalt doping titanium dioxide fiber supported by reduced graphene oxide (Co-TiO₂/RGO) using processes including electrospinning, heating and freeze-drying. The structural and textural features of Co-TiO₂/RGO were characterized by different techniques, and toluene was used as a model pollutant to test its photocatalytic performance. Results show that Co-TiO₂ is uniformly dispersed in reduced graphene oxide and the specific surface area of Co-TiO₂/RGO is higher than that of TiO₂ and Co-TiO₂. Further, under the conditions of Co doping amount of 1%, calcination of 600°C for 4 h, gas residence time of 100 min, relative humidity of 100%, and the Co-TiO₂ loading amount of 1 g, the degradation percentage of toluene by Co-TiO₂/RGO was the highest, up to 99.1%. This study confirms the possibility of synthesizing Co-TiO₂/RGO for the degradation of organic pollutants in air.

Keywords: Titanium Dioxide; Electrospinning; Photocatalytic Degradation; Toluene

Published: May 19, 2025

DOI: <https://doi.org/10.62177/jaet.v2i2.349>

1. Introduction

Volatile organic compounds (VOCs) in air have aroused wide concern because of their great variety and harm^[1]. Among these, toluene is one of the most representative substances, which is toxic and carcinogenic, may cause serious damage to the nervous system and internal organs^[2]. Photocatalytic oxidation, with the advantages of few by-products, and easy to operate, is currently one of the widely used methods to treat VOCs in air. The commonly used photocatalysts mainly include TiO₂^[3], g-C₃N₄^[4], BiOBr^[5] etc., among which TiO₂ is one of the promising photocatalysts due to its strong photocatalytic ability and cost-effectiveness^[6]. However, problems, such as large forbidden band gap, only being excited under UV-light, and low specific surface area, hinder its application^[7].

As so far, many methods have been used in the modification of TiO₂. For example, H₂ could be employed to manufacture oxygen vacancy, which can decrease the band gap of TiO₂^[8]. What's more, doping metal or non-metal elements is also a promising idea, such as S doping^[9], Cu doping^[10], and even Co-doping^[11]. Karafas et al.^[12] prepared cobalt-doped TiO₂

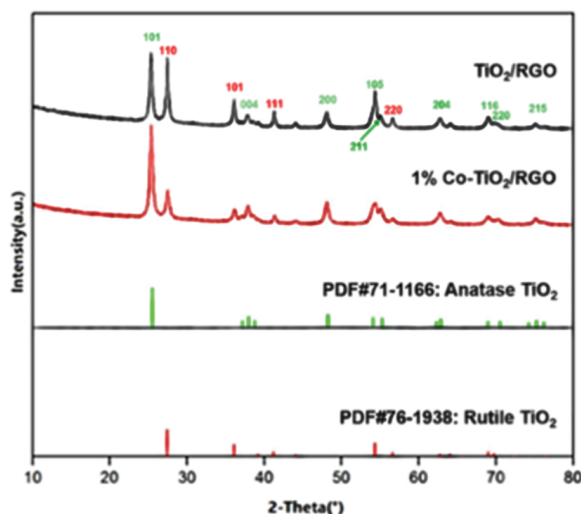
array film, and found that the incorporation of Co improved the absorption of visible light and promoted the separation of photo-generated electro-hole pairs, leading to the better photocatalytic reduction performance compared with the film of TiO_2 . About the problem of carrier recombination, Z-scheme heterojunction could be used to promote the separation of photoexcited carriers and reduce their recombination efficiency, leading to the enhancement of photocatalyst^[13]. Additionally, coupling TiO_2 with semiconductors^[14], and the recombination with carbon-based materials like carbon nanotubes^[15], graphite oxide^[16], and graphene^[17] are also effective to enhance its activity. Because of the good conductivity due to the unique structure of graphene, it is of great interest to synthesize semiconductor-graphene^[18]. Song et al.^[19] used poly(3-hexylthiophene) (P3HT) to modify graphene, and added TiO_2 powder to the P3HT/graphene solution to form a P3HT/ TiO_2 /graphene composite. The presence of graphene improves the charge transport capacity, reduces the e^-h^+ recombination rate and further enhances its light catalytic activity of TiO_2 . In addition, graphene combined with metal-modified TiO_2 has shown the stable and efficient photocatalytic degradation property. Xie et al.^[20] prepared graphene oxide/ TiO_2 - Bi_2WO_6 to degrade ethylene, and found that the catalytic property of TiO_2 - Bi_2WO_6 was significantly improved. However, the aggregation of TiO_2 or metal doped TiO_2 particles will lead to a decrease of the reactivity. Electrospinning has been a worth-trying way to synthesize TiO_2 ^[21-24], especially in the area of photocatalysis, due to its cost-effectiveness, and controlling of fiber dimensions which can get a larger surface area and then enhance the photocatalytic activity, and so on in a similar fashion.

So, in this work, the electrospinning was employed to compound TiO_2 , which can successfully reduce the agglomeration of TiO_2 particles. Moreover, after Co-doping, the biggish band gap, one of the greatest Achille's heels of TiO_2 , could be solved and more visible light can be absorbed. The graphene can promote the transfer of electrons under the effect of its high conductivity, so the recombination of carriers could be restrained^[25-27]. In the end, toluene was chosen as a typical pollutant to test the catalytic activity of this photocatalyst.

2. Results and Discussion

2.1 Morphology, phase and optical characterization

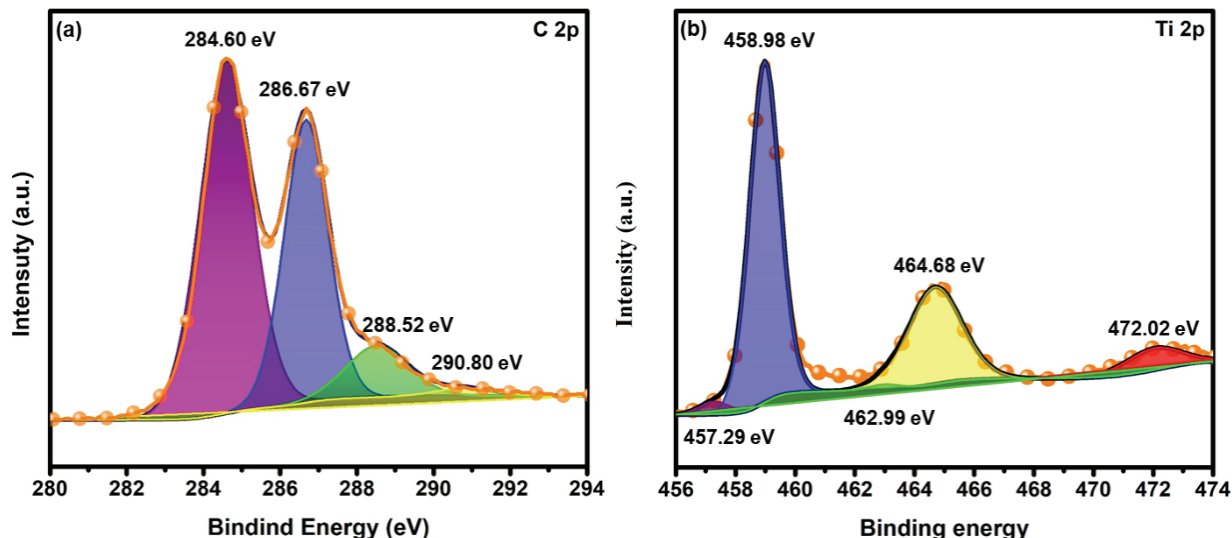
Figure 1. XRD spectra of TiO_2/RGO and 1% $\text{Co-TiO}_2/\text{RGO}$



To explore the meaning of science behind the 1% $\text{Co-TiO}_2/\text{RGO}$ which was synthesized by the electrospinning, a series of characteristics have been tested. First of all, the X-ray diffraction (XRD) patterns of the 1% $\text{Co-TiO}_2/\text{RGO}$ have shown in Figure 1, suggesting that the prepared samples, TiO_2/RGO and 1% $\text{Co-TiO}_2/\text{RGO}$, both remain the structure of TiO_2 . In other words, the 1% Co element could not change the formation of TiO_2 . Compared with the work done by DeawAphairaj^[28], a Rutile TiO_2 emerged after calcined at the temperature of 600°C . What's more, to examine the states of the product 1% $\text{Co-TiO}_2/\text{RGO}$, the X-ray photoelectron spectroscopy (XPS) was monitored in Figure 2. In Figure 2(a), the 284.60 eV is assigned to the C-C sp^2 , and the 286.67 eV and 288.52 eV are represented the C-O-C and O-C=O, respectively. It means that the GO is not exactly reduced to graphene, and still have some functional groups which containing oxygen. In Figure 2(b), the peaks at 458.98 eV, 464.68 eV show a feature of TiO_2 , but with a little red shift compared to the standard TiO_2 ^[29, 30]. It is a result of

the Co doping, decreased the electron density on Ti^{4+} . After modified by Co element, the Ti^{3+} emerges at the peaks of 457.29 eV, 462.99 eV, which means oxygen vacancy was fabricated [31]. As shown in the Figure S2, the Co element can be observed at the peaks of 781 eV, 787 eV, 797 eV and 803 eV. The peaks reported at 781 eV and 797 eV corresponding to the $\text{Co}^{2+} 2p^{3/2}$ and $\text{Co}^{2+} 2p^{1/2}$, respectively. And the ones at 787 eV and 803 eV are satellite peaks which are on behalf of the formation of CoTiO_3 [32-34]. In conclusion, Co element successfully intercalated in the lattice of TiO_2 .

Figure 2. XPS spectra of C 1s (a), Ti 2p (b) of 1% $\text{Co-TiO}_2/\text{RGO}$.



The SEM results of 1% $\text{Co-TiO}_2/\text{RGO}$ is shown in Figure 3a-c. The morphology of Co-TiO_2 fiber is rod-like structure, with regular shapes and uniform protrusions on the surface. The RGO sheet is almost covered by the Co-TiO_2 rod-shaped particles trimly and uniformly as a result of the interaction between the RGO and TiO_2 . What's more, the EDS figures, Figure 3d-f, show that the Co element anchoring in the TiO_2 fiber structure, so the result of EDS and XRD can corroborate each other.

Figure 3. SEM (a, b, c) and SEM-EDS (d, e, f) results of $\text{Co-TiO}_2/\text{RGO}$.

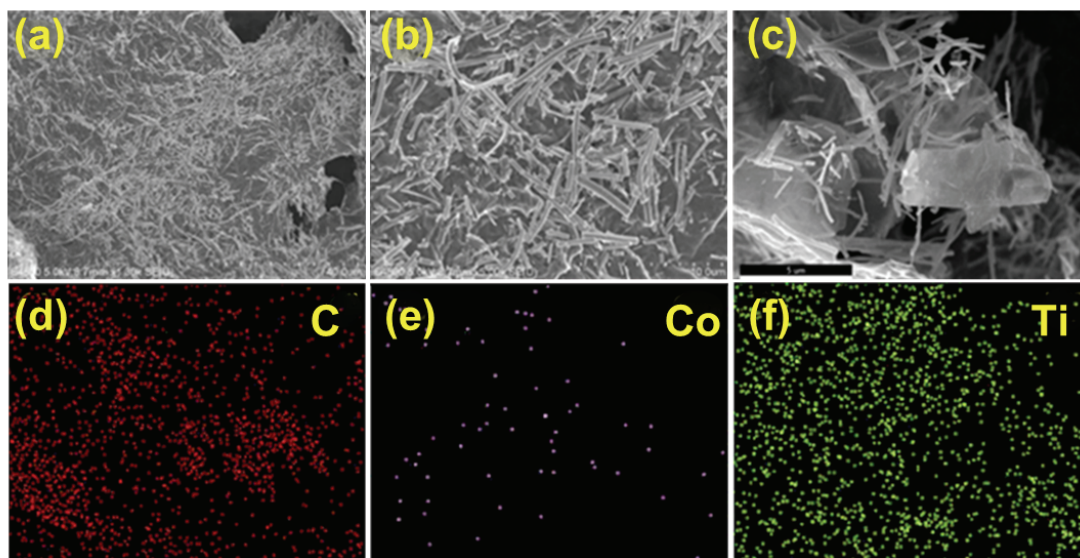


Figure 4 and Table S1 show the N_2 adsorption/desorption results and the pore size results of TiO_2 (a), Co-TiO_2 fibers (b) and $\text{Co-TiO}_2/\text{RGO}$ (c). All the samples exhibit IV-type isotherms with H3-type hysteresis loops which demonstrates the present of mesoporous. As shown in Table S1, the specific surface area of the samples could obtain an obvious increasing after loading onto RGO, because the photocatalysts uniformly disperse onto the surface of RGO. It could let more reactive centers expose and improve its catalytic activity.

Figure 5 is the UV-vis absorption spectra of the samples at different calcination temperatures. It can be seen that the absorption edge of all samples has a significant red shift compared with TiO_2 calcined at the temperature of 500°C , showing

that the addition of Co element can effectively extend the light absorption to the visible light range. This is because the changes of electronic transition. Due to the doping of Co, not only the Ti3d but also the Co3d can become conduction band when O2p become valence band. What's more, the electrons can also excite from Co3d- t_{2g} to Co3d- e_g . In summary, the long wavelength light can be absorbed [35-38]. Specifically, due to the influence of the wide band gap, TiO₂ has a significantly low light applicability, and as a result, the electrons can only be excited under the UV light. However, after the introduction of Co element, an impurity level emerged, which could expand the light absorption of TiO₂, because the edge of the band tail is extended into the forbidden band. The samples with Co doping which is calcined at the temperature of more than 500°C has an obvious blue shift compared with the one calcined at 500°C. This is the reason that the rutile phase of TiO₂ was created after calcined at the temperature of 600°C^[39].

Figure 4. N₂ adsorption/desorption of TiO₂ (a), Co-TiO₂ (b), Co-TiO₂/RGO (c).

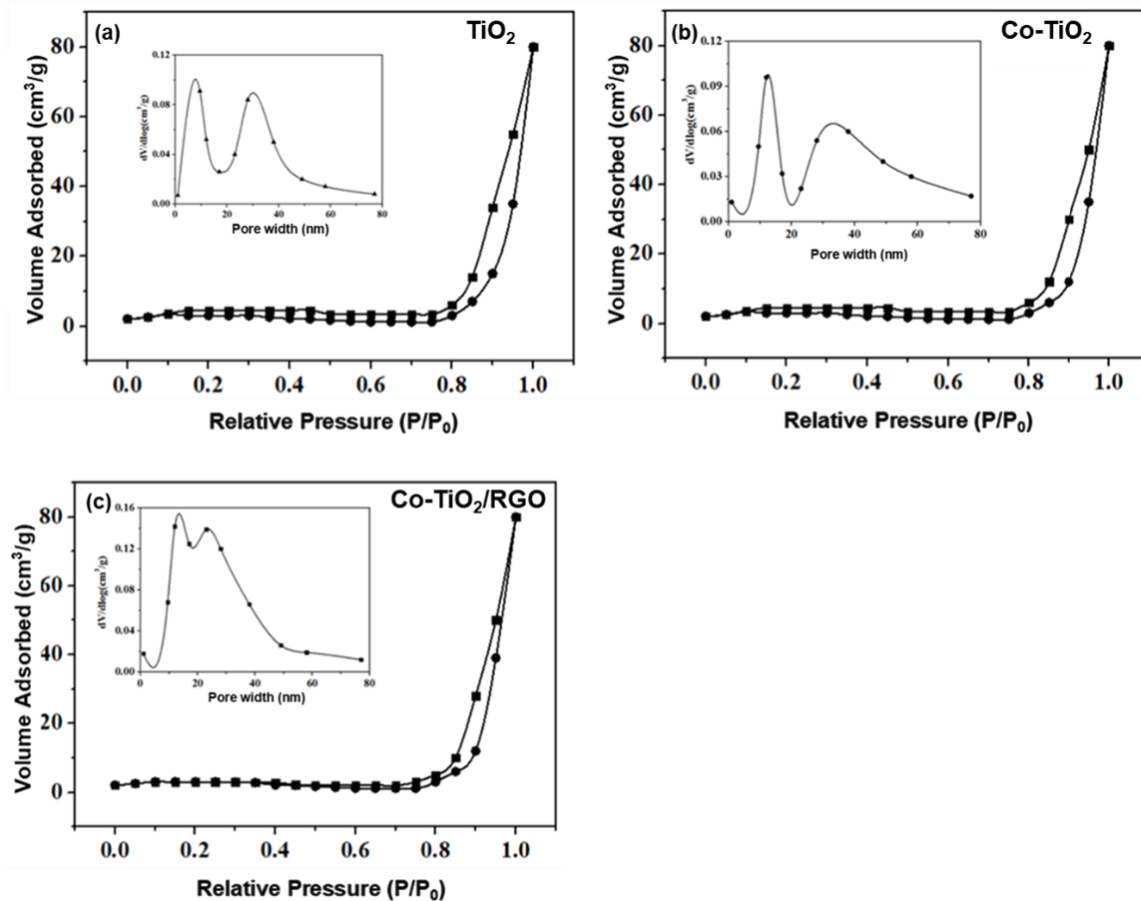
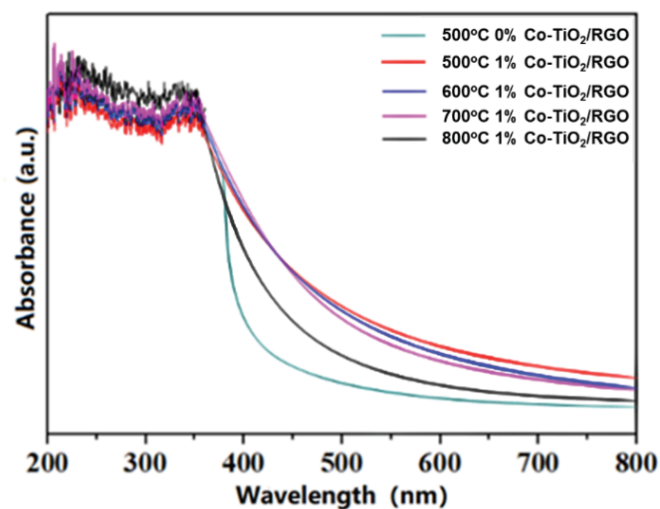


Figure 5. The UV-vis absorption spectra.



2.2 Photocatalytic degradation of toluene

Figure 6a shows the effects of Co-TiO₂/RGO photocatalysts with Co/TiO₂ molar ratios of 0, 0.005, 0.010, 0.015, and 0.020 on the degradation of toluene under the conditions of calcination at 500°C and degradation for 40 minutes. As the Co doping ratio increases, the photocatalytic performance of Co-TiO₂/RGO firstly increases and then decreases. When the molar ratio of Co: TiO₂ = 1: 100, the degradation performance of toluene was the best, reaching 82.6%. With a further increase of the Co doping ratio, the degradation efficiency does not increase but decrease. The photocatalytic activities of Co-TiO₂ in low concentration were higher than the non-doping TiO₂ samples. When the concentration of Co exceeds 2.5%, the photocatalytic rate was lower than non-doping TiO₂. This is because the excess dopants would lead to photon scattering^[40].

Figure 6. Degradation of toluene with different Co doping (a), calcination temperatures (b), gas residence time (c), relative humidity (d), and catalysts.

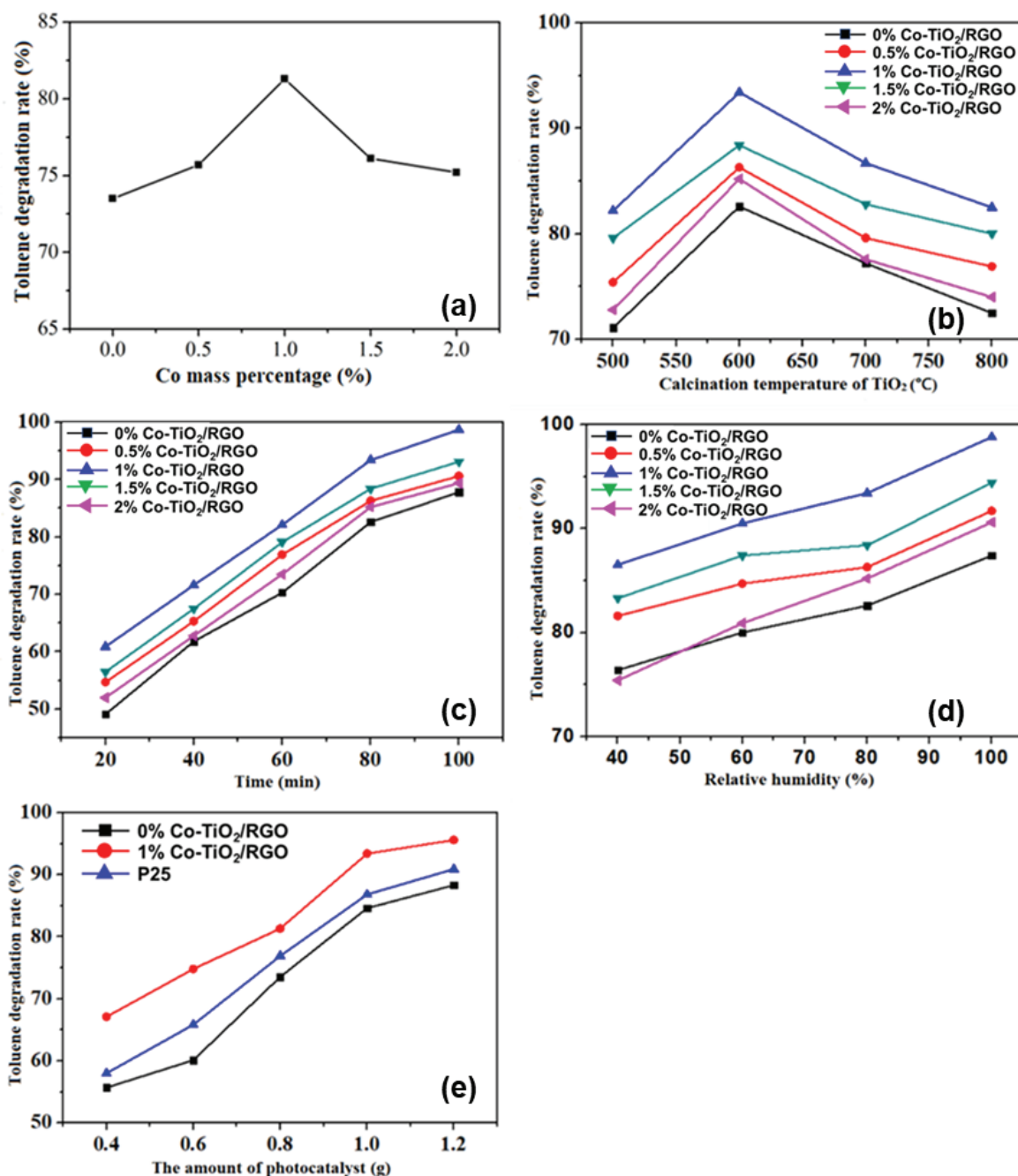


Figure 6b shows the degradation performances of toluene, in 60 min, by catalysts synthesized with different Co/TiO₂ molar ratios at different calcination temperatures. With the increase of calcination temperature, the photocatalytic performance of Co-TiO₂/RGO first increases and then decreases. The best degradation performance was reached when the calcination

temperature is 600°C, and 1% Co-TiO₂/RGO, being 93.4%. When the calcination temperature is lower than 600°C, the PVP fiber is not completely carbonized, and the organic components attach to the Co-TiO₂ fiber, hindering the contact between the catalyst and light, thereby reducing the performance of photocatalytic degradation of toluene. When the calcination temperature is higher than 600°C, the rutile TiO₂ is formed and it can make the activity of the catalyst decline due to being unable to prevent the recombination of the carriers^[41].

Figure 6c shows the degradation of toluene by photocatalysts synthesized with different Co/TiO₂ molar ratios at a calcination temperature of 600°C under different gas residence times. The photocatalytic performance of Co-TiO₂/RGO gradually increases with the increase of gas residence time. After 80 min, the degradation percentage of toluene infinitely approaches 100%, showing that with the increase of gas residence time, the photocatalytic reaction of Co-TiO₂/RGO with toluene will be completed. However, considering the cost, practicality and other factors, 100 minutes of degradation should be enough.

Figure 6d shows the effect of relative humidity on the degradation of toluene in 60 min, by Co-TiO₂/RGO synthesized under the conditions of calcination at 600°C. As the relative humidity increases, the catalytic activity of Co-TiO₂/RGO increases gradually, showing that the increasing of relative humidity is positively correlated with the degradation rate of toluene. The increase can enhance the activity of the catalyst hydroxyl radicals, thereby improving the photocatalytic activity of the catalysts. The degradation percentage of 1% Co-TiO₂/RGO for toluene increased from 86.5% to 99.1%.

Figure 6e shows the effect of TiO₂/RGO, 1% Co-TiO₂/RGO, and TiO₂ on the degradation of toluene under conditions of calcination at 600°C and degradation for 40 min. The 1% Co-TiO₂/RGO has the highest degradation percentage, reaching 95.6%, showing the loading improves the photocatalytic reaction of toluene. Nevertheless, considering the cost and practicability, the loading of the Co-TiO₂ should be selected at about 1 g.

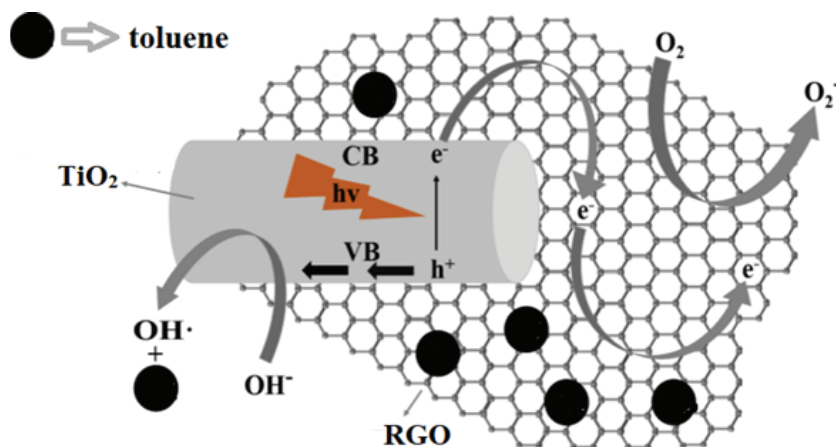
Table 2 lists the removal values of toluene reported in literature using different materials. The data indicates that the Co-TiO₂/RGO system shows competitive removal performance of toluene in shorter reaction time.

2.3 The mechanism of photodegradation of toluene

The doping of Co introduces a new energy level, generates more electron capture centers, and separates more electron-hole pairs. Figure 7 shows the photocatalytic degradation mechanism of toluene using Co-TiO₂/RGO. After being supported on graphene, the larger specific surface area further improves the photocatalytic activity of the catalyst. Toluene, used as the model pollutant, is firstly absorbed on the surface of Co-TiO₂/RGO. After that, Co-TiO₂/RGO is activated by light and the photogenerated electron-hole pairs are separated, and the photogenerated electrons are quickly transferred to the graphene surface. Then a series of reactions occur with oxygen molecules and water molecules (Eq. 1 - 3). After that, the active ·OH and ·O₂⁻ react with the toluene adsorbed on the graphene aerogel to undergo a photocatalytic oxidation reaction, and finally achieve the effect of degrading pollutants.



Figure 7. The photocatalytic degradation mechanism of toluene using Co-TiO₂/RGO.



4. Conclusions

The Co-TiO₂/RGO was prepared by electrospinning, heating and freeze-drying. Toluene was used as a model pollutant, and the photocatalytic degradation experiment was carried out to evaluate the performance of the catalyst. Results show that under the conditions of Co doping molar ratio of 1%, calcination temperature of 600°C for 4 h, gas residence time of 100 min, relative humidity of 100%, and the loading amount of Co/TiO₂ molar ratio of 1%, the degradation percentage of toluene of Co-TiO₂/RGO photocatalyst was the highest, up to 99.1%. This study confirms the possibility of synthesizing Co-TiO₂/RGO for the degradation of pollutants from air.

Funding

This study is financially supported by Science and Technology Commission of Shanghai Municipality (21WZ2501500), and the China Scholarship Council (CSC) scholarship (202106810008).

Conflict of Interests

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

- [1] McDonald, B. C., Gouw, J. A. d., Gilman, J. B., Jathar, S. H., Akherati, A., Cappa, C. D., ... & Trainer, M. (2018). Volatile chemical products emerging as largest petrochemical source of urban organic emissions. *Science*, 359(6377), 760-764.
- [2] Feng, Q., Liu, B., Ji, J., Li, K., Zhang, B., & Huang, H. J. A. S. S. (2021). Enhanced photo-degradation of gaseous toluene over MnOx/TiO₂/activated carbon under a novel microwave discharge electrodeless lamps system. *Journal of Hazardous Materials*, 547, 148955.
- [3] Fujishima, A., & Honda, K. (1972). Electrochemical Photolysis of Water at a Semiconductor Electrode. *Nature*, 238(5358), 37-38.
- [4] Li, P., Wang, M., Huang, S., & Su, Y. (2021). Phosphorus- and fluorine-co-doped carbon nitride: modulated visible light absorption, charge carrier kinetics and boosted photocatalytic hydrogen evolution. *Dalton Transactions*, 50(40), 14110-14114.
- [5] Luo, Z., Ye, X., Zhang, S., Xue, S., Yang, C., Hou, Y., ... & Wang, X. (2022). Unveiling the charge transfer dynamics steered by built-in electric fields in BiOBr photocatalysts. *Nature Communications*, 13(1), 2230.
- [6] Li, Q., & Li, F.-t. (2020). Recent advances in surface and interface design of photocatalysts for the degradation of volatile organic compounds. *Advances in Colloid and Interface Science*, 284, 102275.
- [7] Low, J., Zhang, L., Zhu, B., Liu, Z., & Yu, J. (2018). TiO₂ Photonic Crystals with Localized Surface Photothermal Effect and Enhanced Photocatalytic CO₂ Reduction Activity. *ACS Sustainable Chemistry & Engineering*, 6(11), 15653-15661.
- [8] Wang, P., Deng, P., & Cao, Y. (2022). Edge-sulfonated graphene-decorated TiO₂ photocatalyst with high H₂-evolution performance. *International Journal of Hydrogen Energy*, 47(2), 1006-1015.
- [9] Li, T., Abdelhaleem, A., Chu, W., Pu, S., Qi, F., & Zou, J. (2021). S-doped TiO₂ photocatalyst for visible LED mediated oxone activation: Kinetics and mechanism study for the photocatalytic degradation of pyrimethanil fungicide. *Chemical Engineering Journal*, 411, 128450.
- [10] Bhattacharyya, K., Mane, G. P., Rane, V., Tripathi, A. K., & Tyagi, A. K. (2021). Selective CO₂ Photoreduction with Cu-Doped TiO₂ Photocatalyst: Delineating the Crucial Role of Cu-Oxidation State and Oxygen Vacancies. *The Journal of Physical Chemistry C*, 125(3), 1793-1810.
- [11] Kamble, R. J., Gaikwad, P. V., Garadkar, K. M., Sabale, S. R., Puri, V. R., & Mahajan, S. S. (2022). Photocatalytic degradation of malachite green using hydrothermally synthesized cobalt-doped TiO₂ nanoparticles. *Journal of the Iranian Chemical Society*, 19(1), 303-312.
- [12] Karafas, E. S., Romanias, M. N., Stefanopoulos, V., Binas, V., Zachopoulos, A., Kiriakidis, G., & Papagiannakopoulos, P. (2019). Effect of metal doped and co-doped TiO₂ photocatalysts oriented to degrade indoor/outdoor pollutants for air

- quality improvement. A kinetic and product study using acetaldehyde as probe molecule. *Journal of Photochemistry and Photobiology A: Chemistry*, 371, 255-263.
- [13] Yang, J., Shi, C., Dong, Y., Su, H., Sun, H., Guo, Y., & Yin, S. (2022). Efficient hydrogen generation of vector Z-scheme CaTiO₃/Cu/TiO₂ photocatalyst assisted by cocatalyst Cu nanoparticles. *Journal of Colloid and Interface Science*, 605, 373-384.
- [14] Low, J., Dai, B., Tong, T., Jiang, C., & Yu, J. (2019). In Situ Irradiated X-Ray Photoelectron Spectroscopy Investigation on a Direct Z-Scheme TiO₂/CdS Composite Film Photocatalyst. *Advanced Materials*, 31(5), 1807920.
- [15] Li, X., Yu, J., Li, G., Liu, H., Wang, A., Yang, L., ... & Liu, S. (2018). TiO₂ nanodots anchored on nitrogen-doped carbon nanotubes encapsulated cobalt nanoparticles as photocatalysts with photo-enhanced catalytic activity towards the pollutant removal. *Journal of Colloid and Interface Science*, 526, 158-166.
- [16] Kusiak-Nejman, E., Wanag, A., Kapica-Kozar, J., Kowalczyk, Ł., Zgrzebnicki, M., Tryba, B., Przepiórski, J., Morawski, A. W. (2020). Methylene blue decomposition on TiO₂/reduced graphene oxide hybrid photocatalysts obtained by a two-step hydrothermal and calcination synthesis. *Catalysis Today*, 357, 630-637.
- [17] Padmanabhan, N. T., Thomas, N., Louis, J., Mathew, D. T., Ganguly, P., John, H., & Pillai, S. C. (2021). Graphene coupled TiO₂ photocatalysts for environmental applications: A review. *Chemosphere*, 271, 129506.
- [18] Padmanabhan, N. T., Jayaraj, M. K., & John, H. (2020). Graphene hybridized high energy faceted titanium dioxide for transparent self-cleaning coatings. *Catalysis Today*, 348, 63-71.
- [19] Song, Y., Ling, L., Westerhoff, P., & Shang, C. (2021). Evanescent waves modulate energy efficiency of photocatalysis within TiO₂ coated optical fibers illuminated using LEDs. *Nature Communications*, 12(1), 4101.
- [20] Xie, J., Xu, M., Wang, R., Ye, S., & Song, X. (2021). Three-dimensional porous spherical TiO₂-Bi₂WO₆ decorated graphene oxide nanosheets photocatalyst with excellent visible light catalytic degradation of ethylene. *Ceramics International*.
- [21] Song, Y., Ling, L., Westerhoff, P., & Shang, C. (2021). Evanescent waves modulate energy efficiency of photocatalysis within TiO₂ coated optical fibers illuminated using LEDs. *Nature Communications*, 12(1), 4101.
- [22] Pasini, S. M., Valério, A., Yin, G., Wang, J., de Souza, S. M. A. G. U., Hotza, D., & de Souza, A. A. U. (2021). An overview on nanostructured TiO₂-containing fibers for photocatalytic degradation of organic pollutants in wastewater treatment. *Journal of Water Process Engineering*, 40, 101827.
- [23] Zheng, X., Liu, Y., Liu, X., Li, Q., & Zheng, Y. (2021). A novel PVDF-TiO₂@g-C₃N₄ composite electrospun fiber for efficient photocatalytic degradation of tetracycline under visible light irradiation. *Ecotoxicology and Environmental Safety*, 210, 111866.
- [24] Chen, H.-J., Yang, Y.-L., Zou, X.-X., Shi, X.-L., & Chen, Z.-G. (2022). Flexible hollow TiO₂@CMS/carbon-fiber van der Waals heterostructures for simulated-solar light photocatalysis and photoelectrocatalysis. *Journal of Materials Science & Technology*, 98, 143-150.
- [25] Jiang, M., Zhang, M., Wang, L., Fei, Y., Wang, S., Núñez-Delgado, A., Bokhari, A., Race, M., Khataee, A., Klemeš, J. J., Xing, L., & Han, N. (2022). Photocatalytic degradation of xanthate in flotation plant tailings by TiO₂/graphene nanocomposites. *Chemical Engineering Journal*, 431, 134104.
- [26] Purabgola, A., Mayilswamy, N., & Kandasubramanian, B. (2022). Graphene-based TiO₂ composites for photocatalysis & environmental remediation: synthesis and progress. *Environmental Science and Pollution Research*, 29(22), 32305-32325.
- [27] Chen, L., Li, N., Zhang, M., Li, P., & Lin, Z. (2017). Effect of preparation methods on dispersion stability and electrochemical performance of graphene sheets. *Journal of Solid State Chemistry*, 249, 9-14.
- [28] Aphairaj, D., Wirunmongkol, T., Pavasupree, S., & Limsuwan, P. (2011). Effect of calcination temperatures on structures of TiO₂ powders prepared by hydrothermal method using Thai leucosene mineral. *Energy Procedia*, 9, 539-544.
- [29] Chi, M., Sun, X., Lozano-Blanco, G., & Tatarchuk, B. J. (2021). XPS and FTIR investigations of the transient photocatalytic decomposition of surface carbon contaminants from anatase TiO₂ in UHV starved water/oxygen

- environments. *Applied Surface Science*, 570, 151147.
- [30] Li, Y., Wang, C., Song, M., Li, D., Zhang, X., & Liu, Y. (2019). TiO₂-x/CoO_x photocatalyst sparkles in photothermocatalytic reduction of CO₂ with H₂O steam. *Applied Catalysis B: Environmental*, 243, 760-770.
- [31] Gong, B., Luo, X., Bao, N., Ding, J., Li, S., & Yi, J. (2014). XPS study of cobalt-doped TiO₂ films prepared by pulsed laser deposition. *Surface and Coatings Technology*, 46(10-11), 1043-1046.
- [32] Saravanakumar, K., Fayyaz, A., Park, S., Yoon, Y., Kim, Y. M., & Park, C. M. (2021). Hierarchical CoTiO₃ microrods on Ti₃C₂T_x MXene heterostructure as an efficient sonocatalyst for bisphenol A degradation. *Journal of Molecular Liquids*, 344, 117740.
- [33] Yang, G., Yan, W., Wang, J., & Yang, H. (2014). Fabrication and characterization of CoTiO₃ nanofibers by sol-gel assisted electrospinning. *Materials Letters*, 122, 117-120.
- [34] Habibi, M. H., & Shojaee, E. (2020). Synthesis of a heterojunction CoTiO₃/Co₃O₄ nano-composite thin film with superior photocatalytic activity and reusability: Effect of calcination temperature on phase transformation and effect of oxidants on enhanced degradation of Indo Light Blue dye. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 229, 117796.
- [35] Zou, Z., Ye, J., Sayama, K., & Arakawa, H. (2001). Direct splitting of water under visible light irradiation with an oxide semiconductor photocatalyst. *Nature*, 414(6864), 625-627.
- [36] Tang, J., Zou, Z., & Ye, J. (2003). Photophysical and photocatalytic properties of AgInW₂O₈. *The Journal of Physical Chemistry B*, 107(51), 14265-14269.
- [37] Umebayashi, T., Yamaki, T., Itoh, H., & Asai, K. (2002). Analysis of electronic structures of 3d transition metal-doped TiO₂ based on band calculations. *Journal of Physics and Chemistry of Solids*, 63(10), 1909-1920.
- [38] Wang, D., Zou, Z., & Ye, J. (2003). A new spinel-type photocatalyst BaCr₂O₄ for H₂ evolution under UV and visible light irradiation. *Chemical Physics Letters*, 373(1), 191-196.
- [39] Miyoshi, A., Nishioka, S., & Maeda, K. (2018). Water splitting on rutile TiO₂-based photocatalysts. *Chemistry – A European Journal*, 24(69), 18204-18219.
- [40] Lee, G.-J., Anandan, S., Masten, S. J., & Wu, J. J. (2016). Photocatalytic hydrogen evolution from water splitting using Cu-doped ZnS microspheres under visible light irradiation. *Renewable Energy*, 89, 18-26.
- [41] Ohno, T., Sarukawa, K., & Matsumura, M. (2001). Photocatalytic activities of pure rutile particles isolated from TiO₂ powder by dissolving the anatase component in HF solution. *The Journal of Physical Chemistry B*, 105(12), 2417-2420.

Supporting Information

S1. Experimental

S1.1 Chemicals

Analytical grade reagents including HNO₃, tetra-n-butyl titanate, absolute ethanol, ethylenediamine, and cobalt nitrate hexahydrate and sodium tetraborate decahydrate were obtained from Sinopharm Chemical Reagent Co., Ltd. Reduced graphene oxide was obtained from Jiangsu Yueda Group Co., Ltd. Polyvinylpyrrolidone of analytical grade was obtained from Aladdin Biochemical Technology Co., Ltd. All these chemicals were without further purification, and experimental solutions were prepared with deionized water.

S1.2 Synthesis of Co-TiO₂ fiber using electrospinning

Firstly, 5 g of polyvinyl pyrrolidone was added to 20 g of absolute ethanol to obtain precursor A. Secondly, 15 mL of glacial acetic acid and 15 mL of tetrabutyl titanate were added to 15 mL of ethanol to obtain precursor B. Then, PVP/Co-TiO₂ fiber was obtained by electrospinning the mixture of A and B on the negative iron plate. After that, the PVP/Co-TiO₂ fiber was calcined in a muffle furnace for 4 hours, and the temperature was set at 500°C, 600°C, 700°C, and 800°C, respectively. After the calcination, it was cooled to room temperature, and Co-TiO₂ fiber was obtained after grinding.

S1.3 Synthesis of Co-TiO₂/RGO

60 mL of graphene oxide dispersion was mixed with 0.36 mL of ethylenediamine and 0.003 g of sodium borate. The mixture was sonicated for 10 minutes and then reacted in a vacuum oven. After that, the mixture was soaked in an ethanol aqueous solution for 3 hours to obtain a clean graphene hydrogel. According to a mass ratio of Co-TiO₂ fiber and graphene of 5:100, the Co-TiO₂ fiber and graphene gel were mixed uniformly, and then placed in a vacuum freeze dryer for 35 h to obtain Co-TiO₂/RGO.

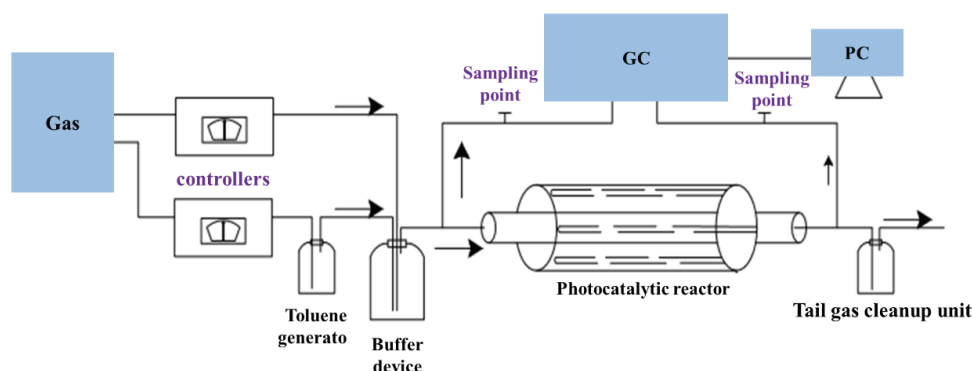
S1.4 Material characterization

Morphology of the samples was characterized by a Hitachi S-4800 scanning electron microscope (SEM). Elemental composition of materials was analyzed using an EDAX TEAM Apollo Energy Dispersive Spectroscopy (EDS). Crystal structure of the samples was determined by a Bruker D8 Advanced X-ray diffractometer (XRD). Ultraviolet-visible (UV-vis) absorption spectra of the samples were obtained from a Shimadzu UV-2550 spectrophotometer. The Brunauer-Emmett-Teller (BET) surface area of samples were measured by a TriStar II analyzer. The photoluminescence (PL) spectrums of materials were tested with a Hitachi F7000 Fluorescence Spectrophotometer.

S1.5 Photocatalytic degradation experiments

The processes and devices for photocatalytic degradation of toluene is shown in Figure S1, which mainly includes the gas distribution part, gas flow control part, photocatalytic degradation part, gas sampling and detection part, and an exhaust gas treatment part.

Figure S1. The process of photocatalytic degradation of toluene.



Toluene standard gas (300 ppm) was used as the model pollution gas. The simulated exhaust gas distribution is divided into 2 routes, which are controlled by the flow controllers in the gas-phase photocatalytic system. Samples were taken from the inlet and outlet and the gas samples were detected by a Shimadzu GC9800 gas chromatograph (GC). The calculation formula for the removal percentage of toluene is as follows:

$$\eta(\%) = \frac{C_{in} - C_{out}}{C_{in}} \times 100\% \quad (1)$$

in which C_{in} and C_{out} are the concentrations of toluene at the gas inlet and the outlet, η is the removal percentage. All the experiments are conducted for 3 times, and the mean value is used for further analysis.

Figure S2. The XPS figure of Co element.

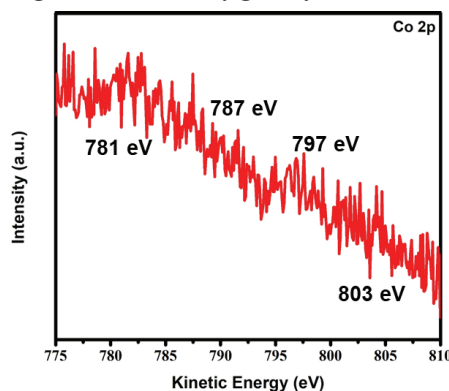


Table S1. Specific surface area, pore size and pore volume of three photocatalysts.

Photocatalysts	Specific surface area (m ² /g)	Pore size (nm)	Pore volume (cm ³ /g)
TiO ₂	142.37	33.10	0.311
Co-TiO ₂	131.56	37.91	0.387
Co-TiO ₂ /RGO	168.86	21.66	0.401

Table S2. Toluene removal by different catalyst.

Catalyst	Degradation percentage (%)	Condition	Reference
TiO ₂ -Cu ₂ O	78.3%	171.6 mg/m ³ , 30 mg catalyst, xenon lamp of 180 W, 3h	[1]
TiO ₂ (Evonik P25, Cristal PC105, Kronos 1077, and Cristal AT-1); Support: flat glass disk	52% for P25; 50% for PC105; 46% for Kronos1077; and 46% for Cristal AT-1	400 ppm, 50 mg catalyst, 40% relative humidity, iron halogenide lamp of 500 W, 6 h	[2]
TiO ₂ /SiO ₂	65%	700 ppm, 0.68 g catalyst, 50% relative humidity, UV Lamp of 3 W, 3 h	[3]
Activated carbon fiber coated TiO ₂	86%	500 ppb, 40% relative humidity, mercury lamp of 15 W, 3 h	[4]
Co-TiO ₂ /RGO	99.1%	1% Co, 1 g catalyst, 100% relative humidity, 100min	This study

References

- [1] Cheng, Y., Gao, X., Zhang, X., Su, J., Wang, G., & Wang, L. (2018). New Journal of Chemistry, 42(1), 9252-9259.
- [2] Bianchi, C. L., Gatto, S., Pirola, C., Naldoni, A., Di Michele, A., Cerrato, G., Crocellà, V., & Capucci, V. (2014). Applied Catalysis B: Environmental, 146, 123-130.
- [3] Paušová, Š., Pacileo, L., Baudys, M., Hrubantová, A., Neumann-Spallart, M., Dvoranová, D., Brezová, V., & Krýsa, J. (2020). Catalysis Today.
- [4] Guo, T., Bai, Z., Wu, C., & Zhu, T. (2008). Frontiers of Environmental Science & Engineering in China, 2(1), 224-229.

Study on the Extraction Process and Pharmacological Activity of Gentiopicroside from Gentiana

Luo Dan*

Tibet Autumn Girl Cosmetics Co., Ltd., Lasa, Xizang, 850000, China

*Corresponding author: Luo Dan , luodanxizang123456

Copyright: 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Gentiopicroside, as a representative substance among monoterpenoid compounds, is the main active ingredient in Gentianaceae plants. It can accelerate gastric emptying efficiency and enhance intestinal motility in the digestive system. In pharmacological studies, it also shows effects on protecting liver cells, inhibiting pain and inflammation. The brachychronic liver damage experiment in mouse was set up by inducing carbon tetrachloride CCl_4 . The activity changes of serum alanine aminotransferase (ALT) and glutathione aminotransferase (AST) were measured. The liveness of philothion peroxisome (GSH-Px), peroxide dismutase (SOD) and malondialdehyde (MDA) details were prepared by liver tissue homogenate. The liveness intensity of serum ALT and AST in the liver of mice in each treatment group was compared from high to low. The liveness of GSH and SOD in liver tissue homogenate increased to varying degrees, and the content of MDA decreased to varying degrees. The study showed that gentiopicroside has a clear defend affect on acute liver damage abduction by CCl_4 . Its mechanism of action may involve inhibiting free radical-mediated lipid peroxidation and effectively maintaining the functional stability of the antioxidant enzyme system of hepatocytes.

Keywords: Gentiopicroside; Digestive System; Intestinal Motility; Pharmacological Research; Hepatocytes

Published: May 22, 2025

DOI: <https://doi.org/10.62177/jaet.v2i2.380>

Introduction

Gentiana scabra is a commonly used medicinal material in clinical practice. It is obtained from the dried rhizome system of Gentiana scabra and Gentiana scabra species of the Gentianaceae family ^[1-2]. In traditional medicine, it is widely known for its affect of eliminate temperature and moist, eliminate liver and cholecyst ignis, etc. This medicinal material is also rich in monoterpene glycosides, and contains multiple active components such as triterpene saponins, flavonoid aglycones and pyrrolizidine alkaloids ^[3]. Taking gentiopicroside $\text{C}_{16}\text{H}_{20}\text{O}_9$ as an example, its water-soluble components have good solubility properties in polar solvents, and its crystal morphology is white to light yellow. Through the liver cell protection mechanism involving the activation of antioxidant enzyme systems and the inhibition of lipid peroxidation, it regulates the prostaglandin synthesis pathway to achieve anti-inflammatory and analgesic effects, and induces tumor cell cycle arrest and apoptosis signal transduction ^[4-5]. The disengaged jacobinic eliminate capacity of gentiana scabra is positively correlated with the liveness of endogenous antioxygen enzymic for instance SOD and GSH-Px.

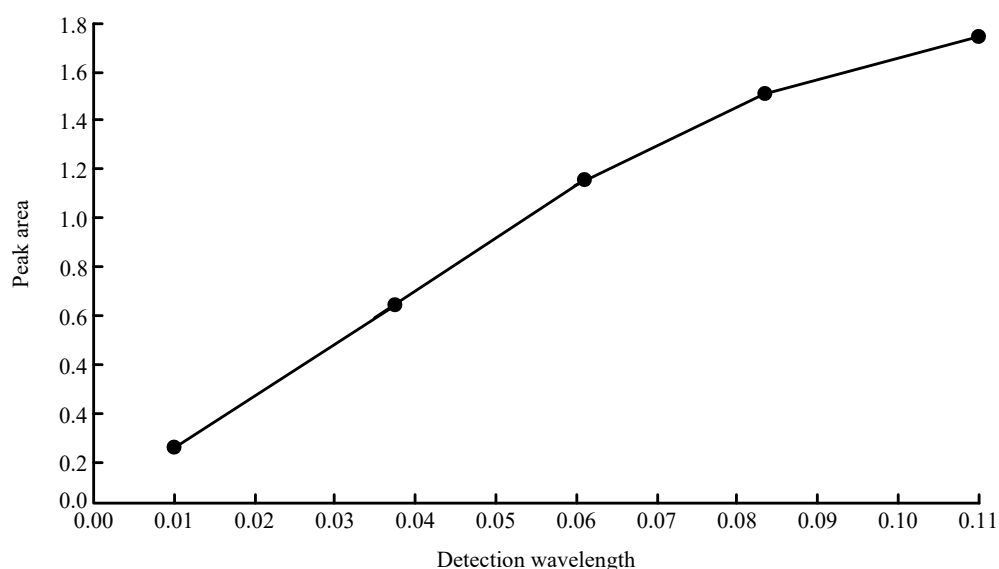
In terms of pharmacological activity, gentiopicroside exhibition a various of biologic liveness, including liver protection, pain relief, inflammation inhibition, free radical scavenging, and tumor growth inhibition. To further clarify the mechanism

of action of this ingredient. This experiment induced liver damage in mice with CCl_4 , and detected the standard of SOD and MDA in liver tissue to reflect whether the drug has anti-liver damage effects. The consequence reveal that Qinling *Gastrodia elata* significantly increased the level of in the liver tissue of mice with liver damage, and reduced the level of, showing a good anti-liver damage effect. There were significant differences between the high, medium, and low dose groups and the model group, indicating that gentiopicroside may have a strong potential to prevent acute liver damage and can be developed as a liver protection drug.

1.Extraction technology of gentiopicroside

5.0 mg of gentiopicroside standard was accurately weighed, dissolution in carbinol and shift to a 25 ml volume bottle to prepare a standard stock solution ^[6]. 1.0, 2.0, 3.0, 4.0 and 5.0 ml of the inventory resolve were accurately shift to a 10 ml volume bottle, attenuation to the trail with carbinol, and filtered through a 0.45 μm microporous filter to obtain a series of gradient concentration standard solutions. HPLC was used for detection, with a Shimadzu C18 row (4.6 mm \times 150 mm, 5 μm) as the chromatography column, water-methanol 75:25 (v/v) as the movement stage, the quantity of flow was 1.0 ml/min, the row heat was 25 $^{\circ}\text{C}$, the test waveform length was 270 nm, the inject cubage was 10 μm , and the retention time of gentiopicroside was 9.932 min. A normal bight was established with quality density (mg/ml) as the standalone variate (X) and vertex acreage as the depend on variate (Y). The linear regression equation was obtained as $Y=1.7238 \times 10^7 X-3.5678 \times 10^5$, $R^2=0.9999$. The standard curve of gentiopicroside is shown in Figure 1. This method has a very stable linear relationship.

Fig.1 Standard curve of gentiopicroside



1.1 Ultrasonic assisted extraction

Accurately weigh 0.5 g (± 0.001 g) of the powder of the root of *Gentiana macrophylla* and place it in a 100 ml block off cone bottle. Increase 20 ml of chromatographic grade methanol. Place the mixed system in an ultrasonic extractor and perform dynamic extraction at 25 $^{\circ}\text{C}$ for 30 min. After the extract is rapidly cooled in an ice-water bath, it is initially filtration through a 0.45 μm polyamides filtration membrane and mensurable shift to a 50 ml volume bottle. The cubage is adjusted to the trail with methanol and vortexed to mix ^[7]. Accurately shift 1.0 ml of the above resolve to a 5 mL volume bottle. After a second filling with carbinol, filtration it through a 0.45 μm organic phase filter membrane to obtain the test solution.

1.2 Low temperature maceration extraction method

Weigh 0.5 g (± 0.001 g) of the powder of the root of *Gentiana macrophylla* accurately, place it in a 100 ml stoppered conical flask, and increase 20 ml of methanol solvent precooled to 4 $^{\circ}\text{C}$. After sealing, transfer it to a constant temperature oscillating box (4 $^{\circ}\text{C} \pm 1^{\circ}\text{C}$) and continue to extract at 150 rpm for 12 h. Filter the extract through a double-layer quantitative filter paper, transfer it quantitatively to a 50 ml volumetric flask, make up to volume with methanol, and vortex homogenize. Accurately measure 1.0 ml of the solution to a 5 ml volumetric flask, dilute it twice with methanol to make up to volume, and filter it through a 0.45 μm organic phase filter membrane to prepare a parallel sample solution.

2. Experimental Materials and Methods

2.1 Drugs and reagents

Gentianopicroside (experimentally prepared, purity $\geq 99.0\%$); peanut oil (purchased from a large supermarket); CCl_4 (Sichuan Laiyun Fine Chemical Co., Ltd.); philothion peroxisome (GSH-Px) reagent, peroxide dismutase (SOD) reagent, malondialdehyde (MDA) assay kit, aminopropionic acid transaminase (ALT) and aspartate transaminase (AST), all of which were buy from Beijing Biotechnology Co., Ltd. [8].

2.2 Instrument

Electronic analytical balance (Shenzhen Tianqin Instrument Co., Ltd.); electronic constant temperature water bath (Fujian Hengke Biotechnology Co., Ltd.); UV800 ultraviolet-visible spectrophotometer (Shimadzu, Japan); desktop high-speed refrigerated centrifuge (Guangzhou Everbright Environmental Protection Equipment Co., Ltd.); glass homogenizer (Jinhua Hongfu Chemical Technology Co., Ltd.).

2.3 Method

The intervention effects of different doses of gentiopicroside on acute liver injury were systematically evaluated by measuring the changes in the liveness of aminopropionic acid transaminase (ALT) and aspartate transaminase (AST) in the plasma of experimental animals, and analyzing the liveness of peroxide dismutase (SOD) and malondialdehyde (MDA) details in liver organization. As a key enzyme for clearing peroxides, the reduce in SOD liveness and the augment in MDA levels reflect the state of oxidative stress, which is an important mechanism of CCl_4 -induced liver injury.

A total of 55 mice were choice for the test and stochastic cut into 5 teams (11 mice in every team) after 20 days of self-adaption feed: normal command team, template team, and low-, medium-, and high-dose gentiopicroside groups. The intervention group was intragastrically administered at a preset dose daily, and the command team was given an equality cubage of salt brine for 15 consecutive days. Two hours after the last administration, besides the normal command team, the other groups were inside the abdominal cavity injection with $0.05\% \text{CCl}_4$ peanut oil solution to construct an acute liver injury model, and the normal team was injection with an equal amount of peanut oil. The mouse were eliminate one day later, and sanguis and liver tissue specimen were gather. After the whole sanguis specimen were allowed to stand at 4°C for solidification, they were centrifuged at 3000 r min to separate serum for subsequent biochemical index detection.

The consequence reveal that the serum ALT and AST activities in the gentiopicroside intervention team were obvious reduced, while the SOD activity of liver tissue increased and the MDA content decreased, instructions that it can effectively remission the oxidation pressure respond. The conclusion suggests that gentiopicroside has a dose-dependent defend affect on CCl_4 -abduction acute liver damage, and its machine-made may be to improve hepatic cell function by inhibiting lipid peroxidation and maintaining antioxidant enzyme activity [9].

After the whole sanguis specimen were allowed to stand at 4°C for solidification, they were centrifugal at 3000 r min for 10 min to acquire plasma, and the plasma ALT and AST enzyme liveness were determined by ultraviolet spectrophotometry. After the liver tissue was irrigated with pre-cooled saline, 0.4-0.6 g of liver lobe tissue was accurately weighed and added with 4°C saline at a ratio of 1:9 (W/V). A 10% slurry was prepared by mechanical homogenization for 10 min under ice bath conditions. The supernatant was collected after low-temperature centrifugation at 3000 r min for 15 min. The standard operation was carried out strictly according to the instructions of the kit. The activity and activity of glutathione peroxidase GSHP ($\text{U}\cdot\text{mg}^{-1}$), superoxide dismutase SOD ($\text{U}\cdot\text{mg}^{-1}$) and malondialdehyde MDA ($\text{U}\cdot\text{mg}^{-1}$) in the liver tissue homogenate were determined by ultraviolet-visible spectrophotometer [10].

3. Experimental Results

As the core metabolic and detoxification organ of the body, the liver is very susceptible to pathological damage under the influence of drugs, viruses and other factors. Traditional Chinese medicine has reveal unique advantages in the field of liver damage prevention and treatment, and plays a liver-protecting role through multiple mechanisms, such as: scavenging oxygen free radicals, alleviating oxidative stress, regulating the inflammatory factor network, and regulating immune homeostasis. These effects work together to protect the structural integrity of liver cells and reduce the degree of pathological damage.

Using modern research technology to deeply explore the liver-protecting mechanism of traditional Chinese medicine and develop highly effective and low-toxic natural liver-protecting drugs has important clinical value.

3.1 Affect of gentiopicroside on liver mass and liver index

Carbon tetrachloride (CCl₄)-abduction liver damage model is widely used in the study of the machine-made of action of liver protection drugs due to its clear pathological indicators and good repeatability. In this experiment, an brachychronic liver damage template of mouse was set up by intraperitoneal inject of CCl₄. It was observed that after the integrity of the liver cell membrane was destroyed, the liveness of plasma aminopropionic acid transaminase (ALT) and aspartate transaminase (AST) augment significantly, and the liver organ coefficient increased synchronously ($P < 0.01$), verifying that the model was successfully constructed. The gentiopicroside intervention group showed a dose-dependent decrease in serum transaminase (ALT/AST) activity, and the difference between the medium and high dose team was statistics obvious ($P < 0.05$) or ($P < 0.01$), and the liver index was also obvious reduced ($p < 0.01$), indicating that it has a clear liver protective effect.

Table 1 Affect of gentiopicroside on small liver mass and liver index in CCl₄-induced acute liver injury

Group	Dose/(g·kg ⁻¹)	Weight/g	Liver mass/g	Liver index%
Normal command group	-	28.50±3.70	1.26±0.17	4.36±0.35 ²⁾
CCl ₄ group	-	28.70±5.60	1.85±0.47	6.12±0.41 ²⁾
Low dose group	0.05	27.80±2.70	1.45±0.17	4.96±0.18 ²⁾
Medium measurement group	0.10	27.10±4.45	1.36±0.14	4.63±0.49 ²⁾
High dose group	0.20	27.15±3.20	1.54±0.11	5.58±0.41 ²⁾

Note: Compared with CCl₄ template team ¹⁾ $P < 0.05$, ²⁾ $p < 0.01$.

3.2 Effects of gentiopicroside on the liveness of ALT and AST

According to the data comparison between the CCl₄ template team and the normal command group, the serum AST and ALT activities in the model group showed a significant increase ($p < 0.01$), indicating that the acute liver injury model was successfully constructed. Compared with the model group, gentiopicroside and each dose group of the positive control group significantly inhibited the trend of increased ALT and AST activities ($P < 0.05$) or ($p < 0.01$). There was no statistical difference in the ALT/AST ratio of all experimental groups, indicating that the release of the two transaminases during CCl₄-induced liver injury was synchronous, and the intervention of gentiopicroside did not change this ratio characteristic. There was no obvious difference in the ALT and AST activity levels between the gentiopicroside dose groups and the positive control team. Table 2 shows the effect of gentiopicroside on ALT/AST activity, suggesting that its liver-protective effect may have similar efficacy to that of positive drugs. This result further verifies the potential mechanism of gentiopicroside in alleviating liver cell damage by regulating the transaminase release pathway.

Table 2 Effect of gentiopicroside on ALT/AST activity

Group	Dose/(g·kg ⁻¹)	ALT/(U·L ⁻¹)	AST/(U·L ⁻¹)	ALT/AST
Normal command team	-	45±22 ²⁾	156±46 ²⁾	0.39±0.15
CCl ₄ team	-	428±156 ²⁾	476±108 ⁴⁾	0.94±0.62
Low dose team	0.05	137±58 ²⁾	260±86 ¹⁾	0.91±0.79
Medium measurement group	0.10	107±32 ²⁾	196±49 ²⁾	0.93±0.50
High dose team	0.20	161±42 ¹⁾	303±54 ¹⁾	0.84±0.36

3.3 Effects of gentiopicroside on GSH-Px, SOD activity and MDA details

The liveness of philothion peroxisome GSH-Px and peroxide dismutase SOD in the liver of mouse in the CCl₄ model team were obvious decreased compared with those in the normal control team ($p < 0.01$), while the content of malondialdehyde

(MDA), the end product of lipid peroxidation, was significantly increased ($p<0.01$), instructions that the brachychronic liver injury template was successfully set up. Compared with the template team, the activities of liver tissue ($p<0.01$) and SOD were significantly augment ($P<0.05$) or ($p<0.01$) after intervention in the low, medium and high doses of gentiopicroside groups, while the content of MDA was obvious reduce ($P<0.05$) or ($p<0.01$). Further analysis showed that there was no statistical difference in the MDA standard between the gentiopicroside team and the positive command team, and its GSH-Px and SOD activities were also similar to those of the positive control team, suggesting that the two team were comparable in antioxidant effect. The inhibitory effect of gentiopicroside on lipid peroxidation showed a dose-dependent enhancement trend. The effects of gentiopicroside on CCl_4 , GSH-Px activity and MDA details are reveal in Table 3, indicating that its protective effect is closely related to the administration concentration.

Table 3 Effects of gentiopicroside on CCl_4 , GSH-Px activity and MDA details

Team	Dose/(g·kg ⁻¹)	GSH-Px/U·mg ⁻¹	SOD/U·mg ⁻¹	MDA/nmol·mg ⁻¹
Normal control group	-	295.67±47.84 ²⁾	226.76±14.79 ²⁾	0.72±0.19 ²⁾
CCl_4 group	-	191.59±40.58 ⁴⁾	159.95±16.45 ⁴⁾	2.43±0.67 ⁴⁾
Low dose group	0.05	29.93±45.90 ¹⁾	190.45±16.66 ²⁾	1.05±0.15 ²⁾
Medium measurement group	0.10	256.76±36.79 ¹⁾	197.88±13.56 ²⁾	0.89±0.16 ²⁾
High dose group	0.20	199.25±38.40 ⁴⁾	183.46±13.28 ^{1, 3)}	1.29±0.25 ¹⁾

4. Discussion

4.1 Discussion on experimental model and indicators

The brachychronic liver damage model abduction by carbon tetrachloride (CCl_4) has become a classic experimental method for evaluating the mechanism of action of hepatoprotective drugs because of its clear indicators and strong repeatability, and it can systematically reflect the functional and morphological changes of hepatocytes. In this experiment, an brachychronic liver damage template was set up in mouse by intraperitoneal inject of CCl_4 , and the dynamic changes of serum aminopropionic acid transaminase (ALT) and aspartate transaminase (AST) liveness were observed after the integrity of the hepatocyte membrane was destroyed to quantify the degree of liver damage. The serum ALT and AST liveness of mouse in the CCl_4 template team were obvious higher than those in the normal command team ($p<0.01$), and the liver index was also obvious augment ($p<0.01$), indicating depend on that acute liver injury was successfully established. After intervention with gentiopicroside, the plasma ALT, AST activities and liver index of mouse in each dose group showed a dose-reduce ($P<0.05$) or ($p<0.01$), suggesting that it has a clear defend affect on CCl_4 -induced liver damage. This result further verifies the potential mechanism by which gentiopicroside exerts a hepatoprotective effect by regulating the release of transaminases and alleviating pathological changes in hepatocytes.

4.2 Antioxidant system mechanism of action

The core function of philothion peroxisome GSH-Px is to catalyze the decomposition of hydrogen peroxide and inactivate it by combining with metabolic toxic products; superoxide dismutase SOD, as a key component of the endogenous free radical scavenging system, can efficiently scavenge superoxide anion free radicals. The synergistic effect of the two constitutes the core of the antioxidant defense system in mice, inhibiting lipid peroxidation reactions and protecting biological membrane structures from oxidative damage by activating endogenous antioxidant mechanisms. Its activity level directly determines the dynamic balance between oxidative stress and antioxidant capacity of the body.

During CCl_4 -induced liver injury, the liveness of GSH-Px and SOD reduce obvious, while the details of malondialdehyde (MDA), the end product of lipid peroxidation, increased obvious. MDA can further destroy cell membrane function by cross-linking with biomacromolecules to form complexes, and its details changes can indirectly reflect the degree of damage to tissues or cells by free radicals. By quantitatively analyzing the liveness of GSH-Px, SOD and MDA standard in liver

organization, the degree of lipid peroxidation and the pathological progression of free radical-mediated liver injury can be systematically evaluated, providing a key experimental basis for revealing the mechanism of antioxidant intervention.

4.3 Molecular mechanism and protective effect

Experimental data showed that the content of malondialdehyde (MDA) in the liver of mice in the CCl₄ model group showed a very significant increase ($p < 0.01$), indicating that CCl₄-induced acute liver injury is closely related to the inhibition of antioxidant enzyme activity and lipid peroxidation. After intervention with gentiopicroside, the activities of glutathione peroxidase GSH-Px and peroxide dismutase SOD in liver tissue of each dose team were obvious enhanced compared with the model team ($P < 0.05$) or ($p < 0.01$). At the same time, the details of MDA decreased obvious ($p < 0.01$). This result shows that gentiopicroside can effectively block the initiation of lipid peroxidation chain reaction by maintaining the activity level of antioxidant enzymes and enhancing the body's capacity to scavenge oxygen jacobinic, significantly reducing the accumulation of MDA in the liver, and ultimately alleviating CCl₄'s acute damage to liver tissue. Its specific mechanism may involve regulating the function of the endogenous antioxidant system to maintain the homeostasis of enzyme activity, but it still needs to be further verified through molecular pathway research.

5. Conclusion

Experimental data reveal that the details of malondialdehyde (MDA) in the liver of mice in the CCl₄ template team showed an extremely significant increase ($p < 0.01$), which was closely related to the inhibition of antioxidant enzyme activity and CCl₄-induced acute liver damage and lipid peroxidation. After intervention with gentiopicroside, the liveness of philothion peroxisome GSH-Px and peroxide dismutase SOD in the liver tissue of the above-dose groups were obvious improved compared with the model command team ($P < 0.05$) or ($p < 0.01$), and the MDA content was obvious decreased ($p < 0.01$). The results showed that gentiopicroside can maintain the level of antioxidant enzyme activity and enhance the body's capacity to scavenge oxygen disengaged jacobinic, thereby effectively blocking the initiation of lipid peroxidation chain reactions, significantly reducing the accumulation of MDA in the liver, and ultimately alleviating CCl₄'s acute damage to liver tissue. Its specific mechanism may involve regulating the function of the endogenous antioxidant system to maintain enzyme activity homeostasis, but it still needs to be further verified through molecular pathway research.

Funding

no

Conflict of Interests

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

- [1] Liu, B., Pang, F., Bi, H., et al. (2024). Regulatory mechanisms of Gentiopicroside on human diseases: a brief review. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 397(2), 725-750. <https://doi.org/10.1007/s00210-023-02258-6>
- [2] Antoniadi, L., Bartnik, M., Angelis, A., et al. (2023). Gentiopicroside—An Insight into Its Pharmacological Significance and Future Perspectives. *Cells*, 13(1), 70. <https://doi.org/10.3390/cells13010070>
- [3] Ding, Y., Gou, X., Wu, X., et al. (2024). Gentiopicroside: An updated review of its pharmacological activities and mechanisms. *Chemistry & Biodiversity*, e202402306. <https://doi.org/10.1002/cbdv.202402306>
- [4] Zhang, Q., Peng-Fei, X. I. A., Xue-Jing, P., et al. (2022). Synthesis, and anti-inflammatory activities of gentiopicroside derivatives. *Chinese Journal of Natural Medicines*, 20(4), 309-320. <https://doi.org/10.1016/j.cjnm.2022.04.001>
- [5] Xiao, H., Sun, X., Lin, Z., et al. (2022). Gentiopicroside targets PAQR3 to activate the PI3K/AKT signaling pathway and ameliorate disordered glucose and lipid metabolism. *Acta Pharmaceutica Sinica B*, 12(6), 2887-2904. <https://doi.org/10.1016/j.apsb.2022.02.001>
- [6] Wang, X., Long, D., Hu, X., et al. (2023). Gentiopicroside modulates glucose homeostasis in high-fat-diet and streptozotocin-induced type 2 diabetic mice. *Frontiers in Pharmacology*, 14, 1172360. <https://doi.org/10.3389/fphar.2023.1172360>

- [7] Sun, F., Ma, Y., Li, D., et al. (2025). Gentiopicroside Attenuated Dopaminergic Neurodegeneration via Inhibiting Neuroinflammatory Responses and Ferroptosis in Experimental Models of Parkinson's Disease. *Basic & Clinical Pharmacology & Toxicology*, 136(5), e70036. <https://doi.org/10.1111/bcpt.13615>
- [8] Zhang, Y., Pan, S., Yi, S., et al. (2023). Gentiopicroside ameliorates CCl₄-induced liver injury in mice by regulating the PPAR- γ /Nrf2 and NF- κ B/I κ B signaling pathways. *Journal of International Medical Research*, 51(10), 03000605231204501. <https://doi.org/10.1177/03000605231204501>
- [9] Wang, J., Fasina, O. B., Manzoor, M., et al. (2023). A new gentiopicroside derivative improves cognitive deficits of AD mice via activation of Wnt signaling pathway and regulation of gut microbiota homeostasis. *Phytomedicine*, 113, 154730. <https://doi.org/10.1016/j.phymed.2023.154730>
- [10] Chang, Y., Tian, Y., Zhou, D., et al. (2021). Gentiopicroside ameliorates ethanol-induced gastritis via regulating MMP-10 and pERK1/2 signaling. *International Immunopharmacology*, 90, 107213. <https://doi.org/10.1016/j.intimp.2020.107213>

Synergistic Optimization and Risk Control: Integrated Price Forecasting Models

Mingshen Xu^{1*}, Tianxiang Hu¹, Boyi Zhang², Zhang Hanrui³, Kexiao Wu⁴, Teng Zhang¹, Xiaotian Li¹, Qianxve Mo⁵, Xianliu Feng²

1.Department of Mathematics and Physics, North China Electric Power University, Baoding, Hebei, 071003, China

2.Department of Environmental Science and Engineering, North China Electric Power University, Baoding, Hebei, 071003, China.

3.Department of computer, North China Electric Power University, Baoding Hebei, 071003, China

4.Department of Automation, North China Electric Power University, Baoding, Baoding, Hebei, 071003, China

5.Department of Electrical Engineering, North China Electric Power University, Baoding, Hebei, 071003, China

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

Copyright: 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: This paper presents a hybrid intraday electricity price forecasting model—Info-VMD-iTransformer-CNN-LSTM—tailored for high-dimensional, non-stationary price series. First, variational mode decomposition (VMD) adaptively separates price signals into intrinsic modes, mitigating mode mixing and noise. Next, an improved Transformer (iTransformer) with enhanced positional encoding captures long-range dependencies, while CNN layers extract local spatio-temporal features and LSTM units model sequential dynamics. Finally, the INFO algorithm automates hyperparameter optimization, ensuring both high accuracy and robustness. Empirical evaluations demonstrate that our approach consistently outperforms existing benchmarks under volatile market conditions, making it well suited for real-time forecasting in modern power systems.

Keywords: Variational Mode Decomposition; Improved Transformer; CNN-LSTM; INFO Optimization; Intraday Price Forecasting

Published: May 24, 2025

DOI: <https://doi.org/10.62177/jaet.v2i2.361>

1.Introduction

Intraday electricity price forecasting is pivotal for real-time market-driven trading and system dispatch, yet price series exhibit complex nonlinear and nonstationary behaviors due to uncertainties in load, generation, weather, policy, and high renewable penetration^[1–2]. To mitigate nonstationarity, methods such as discrete wavelet transform (DWT)^[3] and empirical mode decomposition (EMD)^[4], along with noise-assisted extensions—including ensemble EMD (EEMD)^[5], complete ensemble EMD with adaptive noise (CEEMDAN)^[6], and improved CEEMDAN (ICEEMDAN)^[7]—have been applied, but they incur significant computational overhead and still suffer from incomplete noise suppression^[8–9]. Variational mode decomposition (VMD) employs a variational framework to adaptively decompose price signals into multiresolution modes while effectively avoiding mode mixing, offering superior computational efficiency compared to ICEEMDAN^[10]. VMD has been successfully integrated with one-dimensional CNN-BiLSTM^[11], CNN-LSTM^[12], and EDE-BiLSTM^[13] architectures

to substantially improve forecasting accuracy.

Traditional statistical time-series models (ARIMA, SARIMA, VAR, GARCH) provide interpretability but struggle with the strong nonlinearity and long-range dependencies typical of intraday prices^[14–17]. Consequently, data-driven machine learning algorithms—Random Forest^[18], decision tree regression^[19], ANN^[20], CatBoost^[21], SVR^[22], XGBoost^[23], ELM^[24]—and deep learning architectures—including CNN^[25] (e.g., CNN–BiGRU^[26], secondary decomposition attention CNN–LSTM–MLR^[27]), LSTM^[28–32], and Transformer-based models^[33–36]—have become mainstream solutions. Hyperparameter optimization has leveraged classical metaheuristics—genetic algorithm^[37], particle swarm optimization^[38], ant colony optimization^[39], gravitational search^[40], teaching–learning-based optimization^[41]—guided by the no-free-lunch theorem^[42], as well as recent variants such as MPA^[43], CSA^[44], AOA^[45], GEO^[46], SFO^[47], ChOA^[48], SMA^[49], DO^[50], AVOA^[51], AAM^[52], BWO^[53], ECO^[54], HFA^[55] and GOA^[56], demonstrating efficacy in hybrid frameworks like EEMD + WPD + THPO–DELM^[57] and VMD–GWO–ATT–LSTM^[58].

In this study, we develop a hybrid VMD–iTransformer–LSTM–INFO framework solely for intraday electricity price forecasting: VMD is employed to adaptively decompose nonstationary price series into distinct frequency modes, an improved Transformer (iTransformer) captures long-range dependencies with enhanced positional encoding, LSTM layers model sequential temporal patterns, and the INFO algorithm efficiently optimizes all hyperparameters. Extensive experiments on benchmark datasets demonstrate that our model consistently outperforms existing approaches in both accuracy and robustness under volatile market conditions.

2. Related methodologies

2.1 Variational Mode Decomposition

Variational Mode Decomposition (VMD) is a decomposition algorithm applied to non-stationary signals, which can decompose the original signal into a series of modal components with increasingly higher frequencies. Compared to Empirical Mode Decomposition (EMD) and Local Mean Decomposition (LMD), VMD can overcome two major drawbacks: endpoint effects and mode mixing.

The VMD algorithm decomposes the original non-stationary signal n into k relatively stationary sub-signals $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$, each with a central frequency ω_n and a limited bandwidth. Each sub-signal serves as a band-limited intrinsic mode function (BLIMF) of the original signal, capable of reflecting the structural characteristics of the original signal at different time scales, as shown in equations (1) and (2):

$$f(t) = \sum_{n=1}^k \varepsilon_n \quad (1)$$

$$\mu_n(t) = A_i(t) \cos(\varphi_i(t)) \quad (2)$$

To estimate each modal component, VMD initially uses the Hilbert transform to obtain the one-sided frequency spectrum of each modal component during the solving process, as shown in equation (3):

$$\left[\delta(t) + \frac{j}{\pi t}\right] \otimes \varepsilon_k(t) \quad (3)$$

Here, \otimes represents convolution. Next, the frequency spectrum of each modal component is modulated to its respective estimated central frequency's baseband, as shown in equation (4):

$$\begin{cases} \min_{\{\varepsilon_n\} \{\omega_n\}} \left\{ \sum_k \left| d_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes \varepsilon_k(t) \right] \right| \right\} e^{-j\omega_k t} \\ s.t. \sum_k \varepsilon_n(t) = f(t) \end{cases} \quad (4)$$

Here, ε_k represents the k -th BLIMF, and ω_k denotes the central frequency of ε_k . Furthermore, during the solving process, a quadratic penalty factor α and a Lagrange multiplier operator $\lambda(t)$ are introduced to convert the constrained variational problem into an unconstrained variational problem, as shown in equation (5):

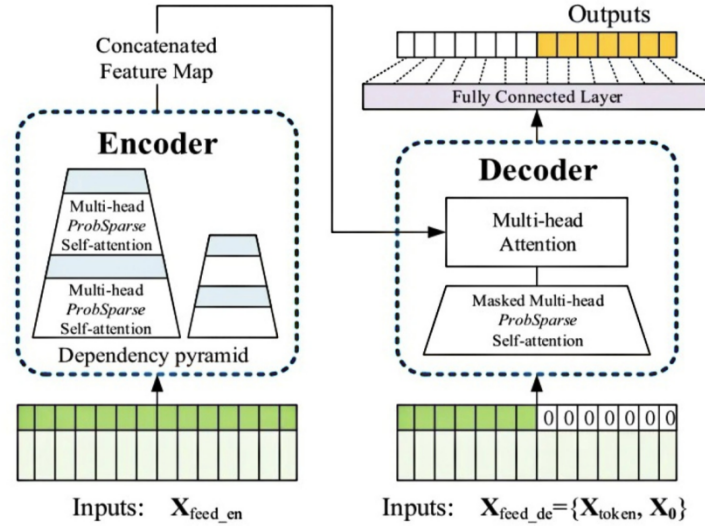
$$\begin{aligned}
L(\{\varepsilon_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_k \left| d_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes \varepsilon_k(t) \right] e^{-j\omega_k t} \right| \\
& + \left| f(t) - \sum_k \varepsilon_k(t) \right|_2^2 + \left\langle \lambda(t), f(t) - \sum_k \varepsilon_k(t) \right\rangle
\end{aligned} \tag{5}$$

After completing the above steps, the unconstrained variational problem in equations (3)–(5) is solved using the Alternating Direction Method of Multipliers (ADMM). This completes the basic construction of the VMD model.

2.2 iTransformer

iTransformer is a variant of the Transformer model[36]. In simple terms, it achieves excellent performance in multivariate time series forecasting tasks by reversing the input time series. Its structure is shown in Figure 1.

Figure 1. Schematic of the iTransformer Encoding Layer Structure.



iTransformer only uses the encoder part of the Transformer model. Its structure consists of an embedding layer, Transformer blocks (TrmBlock), and a projection layer. The embedding layer is composed of a multi-layer perceptron (MLP). Unlike the Transformer, which embeds different variables at the same time point into a single token, iTransformer reverses the input sequence X with N variables and T time steps, then embeds each individual time series into a single token. The embedding layer maps the input sequence to a feature matrix $H = \{h_1, h_2, \dots, h_n, \dots, h_N\} \in \mathbb{R}^{N \times D}$, where h represents the token, and D is the embedding projection dimension.

The TrmBlock layer consists of a multi-head self-attention layer, a normalization layer, and a feed-forward network, allowing the model to learn the temporal characteristics of different variables and the multi-dimensional correlations between them. The multi-head self-attention layer captures multivariate correlations through the attention mechanism. The attention map $A \in \mathbb{R}^{N \times N}$ is obtained by mapping the input matrix to query vectors (Queries), key vectors (Keys), and value vectors (Values).

The purpose of the normalization layer is to reduce the differences caused by different measurement methods between variables, as shown in equation (6), enhancing the independence between variables and preventing the occurrence of non-stationarity issues. Therefore, the feed-forward network uses a multi-layer perceptron to learn the non-linear representations of the variables, as detailed in equation (7) :

$$A(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_{ff}}}\right) \tag{6}$$

$$L(H) = \left\{ \frac{h_n - \text{Mean}(h_n)}{\sqrt{\text{Var}(h_n)}} \right\} \tag{7}$$

Here, dff represents the projection dimension. The projection layer consists of a multi-layer perceptron, which maps the hidden temporal features of each variable to the target time dimension T' and then generates the future sequence $Y = \{y_1, y_2, \dots, y_n, \dots, y_N\} \in \mathbb{R}^{N \times T'}$. Based on the above explanation, the encoding process for obtaining the predicted results, denoted as Y , for each specific variable X can be expressed by equations (8)–(10):

$$h_n^0 = Em(X_{:,n}) \quad (8)$$

$$H^{q+1} = TrmBlock(H^q), q = 0, \dots, Q-1 \quad (9)$$

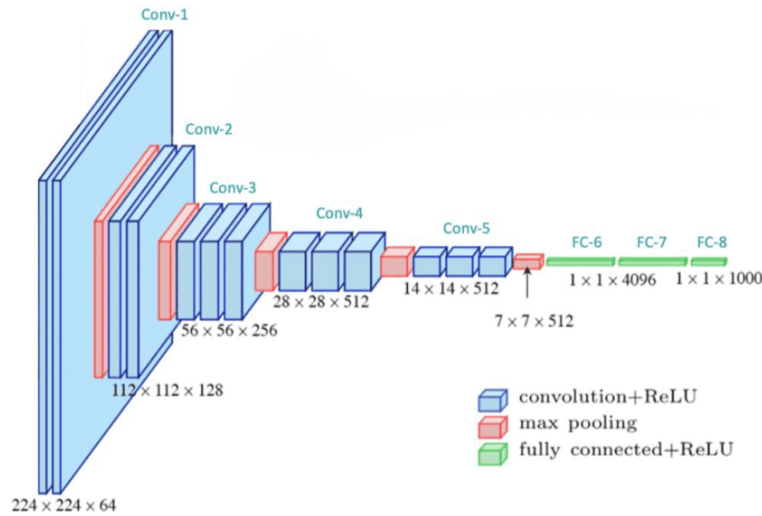
$$Y_{:,n} = Pj(h_n^Q) \quad (10)$$

Here, Q represents the number of TrmBlock layers, Em represents the embedding layer, and Pj represents the projection layer. From the content of this section, it can be considered that iTransformer, based on the Transformer architecture, possesses strong sequence modeling capabilities and parallel processing ability. It can rapidly capture instantaneous changes and high-frequency components in electricity price signals, accurately reflecting the details of price fluctuations. This makes it suitable for processing high-frequency electricity price signals obtained through VMD decomposition.

2.3 Convolutional Neural Networks

CNN (Convolutional Neural Networks) is a feedforward neural network proposed by LeCun et al. [25]. Generally, the CNN structure consists of convolutional layers, pooling layers, and fully connected layers, as shown in Figure 2. Essentially, CNNs attempt to build multiple filters that can extract hidden features through layer-by-layer convolution and pooling of the input data. Finally, these abstract features are merged through a fully connected layer and used with an activation function to address classification or regression problems [69].

Figure 2. Basic Architecture of Convolutional Neural Networks



In the convolutional layer, the feature map from the previous layer is convolved with the convolutional kernel, and the resulting feature map is generated by an activation function. The computation process of the convolutional layer is represented by equation (11):

$$a_j^n = f(\sum_{i \in C_l} x_i^{n-1} * k_{ij}^n + b_j^n) \quad (11)$$

Here, a_j^n represents the i -th output feature map of the j -th layer, a_j^{n-1} represents the j -th output feature map of the l -th layer, C_j represents the selection of input mapping, and k_{ij}^n denotes the operation between the i -th and j -th output feature maps. $*$ represents convolution, b_j^l denotes the bias term, and $f(\cdot)$ represents the Rectified Linear Unit (ReLU) activation function. The pooling layer is used to reduce the number of parameters in the network. This is achieved by calculating the average value (average pooling) or the maximum value (max pooling) of a given region in the feature map. The computation process of the pooling layer is described in equation (12):

$$a_j^n = f(\beta down(a_j^{n-1}) + b_j^n) \quad (12)$$

Where $down(\cdot)$ represents the subsampling function in the max pooling process.

In Eq (12), a^n represents the final output vector, a^{n-1} represents the input vector, W^n denotes the weights between the n -th and $(n+1)$ -th layers, and b^n represents the bias.

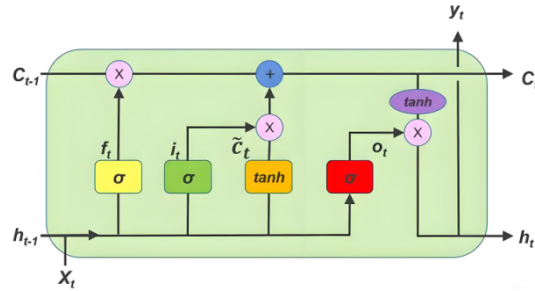
In summary, CNNs, through convolution and pooling operations, effectively extract local features and mid-frequency information from electricity price signals, such as periodic patterns in price changes, while maintaining a relatively low computational complexity. This makes CNNs suitable for tasks that involve analyzing mid-frequency electricity price signals.

2.4 Long Short Term Memory

LSTM (Long Short-Term Memory) networks are an enhanced type of recurrent neural network (RNN) that can automatically store and delete temporary state information. This capability addresses the vanishing gradient problem that RNNs encounter when processing long-term sequences, and it aids in extracting complex temporal features from time series data. As shown in Figure 3, the basic structure of an LSTM consists of an input layer, a recurrent layer, and an output layer.

The core component behind LSTM is the gating mechanism, which includes a hidden unit h_t , an input gate i_t , an output gate o_t , a forget gate, an input modulation gate g_t , and a storage unit (or memory unit) c_t . The information flowing from the input gate i_t to the storage unit c_t is controlled by the input gate i_t , while the output gate o_t conditionally controls what information should be output to the rest of the network based on the output activation. The forget gate f_t determines how much information in the unit's internal state should be discarded before it is imported through the self-loop connection, with the purpose of forgetting or resetting the unit's memory.

Figure 3. Schematic Diagram of the LSTM Gating Mechanism



Additionally, the storage unit and gating mechanism decide what information should be retained or forgotten, which helps ensure that the gradient is retained in the unit and continues to persist. When the input gate is open, the storage unit allows information to be added to the unit. During this process, if the forget gate is activated, information from the previous unit state can be disregarded. The output gate further determines what information flows from the updated unit output to the final state. Throughout this process, the input, unit output, and state are all one-dimensional vectors. To learn precise timing of the outputs, connections from the internal unit to each gate in the same unit, known as peek-hole connections, are introduced. Given the input vector at time step t , the previous hidden state, and the previous unit state, the LSTM update process is represented by equations (13)–(18) :

$$i_t = \sigma(WM_{\theta_i}\theta_t + WM_{h_i}h_{t-1} + WM_{c_i} \odot c_{t-1} + b_i) \quad (13)$$

$$f_t = \sigma(WM_{\theta_f}\theta_t + WM_{h_f}h_{t-1} + WM_{c_f} \odot c_{t-1} + b_f) \quad (14)$$

$$g_t = \phi(WM_{\theta_g}\theta_t + WM_{h_g}h_{t-1} + b_g) \quad (15)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (16)$$

$$o_t = \sigma(W_{\theta_o}\theta_t + W_{h_o}h_{t-1} + WM_{c_o} \odot b_o) \quad (17)$$

$$h_t = o_t \odot \phi(c_t) \quad (18)$$

Here, WM represents the corresponding weight matrix (Weight Matrix); for example, WM_{θ_i} denotes the weight matrix from the input vector to the input gate, while WM_{c_i} , WM_{c_f} and WM_{c_o} represent the diagonal weight matrices associated with the peephole connections. b denotes the corresponding bias vector, for instance, b_i represents the bias vector of the input gate. The symbol \odot indicates element-wise multiplication. The function $\sigma(x)$ is used to compress its input into the range (0, 1)

and is known as the sigmoid non-linear function, while $\phi(x)$ denotes the hyperbolic tangent non-linear function, which also compresses its input into the range $(-1, 1)$. Respectively, the mathematical expressions of the functions $\sigma(x)$ and $\phi(x)$ are provided in equations (19) and (20) :

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (19)$$

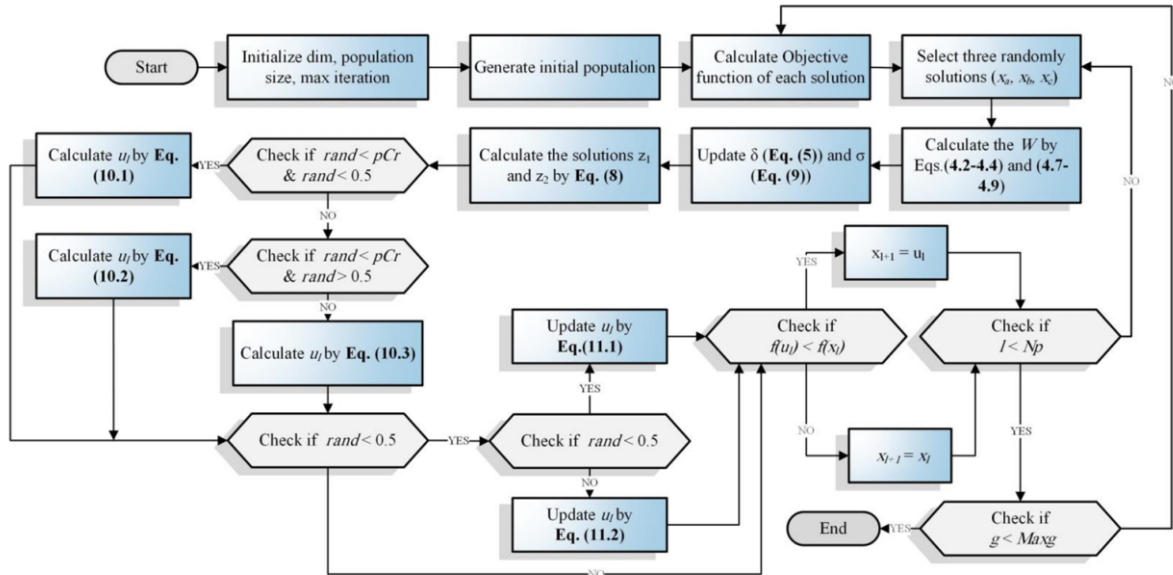
$$\phi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (20)$$

In summary, due to LSTM's multi-layered and hierarchical architecture, it exhibits exceptional efficiency and powerful capability in feature representation. Specifically, LSTM can effectively capture and retain long-term dependencies in time series data through its unique memory mechanism. In the analysis of electricity price signals, low-frequency components often encapsulate important trends and periodic patterns in price changes. By leveraging this advantage of LSTM, we can extract deeper low-frequency features from electricity price signals by increasing the number of network layers. In this manner, LSTM is not only able to capture short-term fluctuations in electricity price signals but also accurately grasp their long-term trends, thereby providing robust technical support for electricity price forecasting and decision-making. Consequently, employing LSTM to capture the low-frequency components of electricity price signals is a rational and effective technical approach.

2.5 Weighted Mean of Vectors Algorithm

The INFO (Weighted Mean of Vectors)[59] algorithm is an improved weighted averaging method, with its core principle centered on optimizing the interactions among vectors by applying different weighted averaging rules. Given the high efficiency it demonstrates under specific constraints, the algorithm exhibits broad application prospects in fields such as optimization design.

Figure 4. INFO Algorithm Flowchart



As shown in Figure 4, the INFO algorithm mainly encompasses three core steps. First, the Updating Rules stage forms the foundation of the algorithm by generating new vectors based on convergence acceleration and the mean principle. Second, in the Vector Merging stage, the obtained vectors are combined with the updating rules to form superior vectors. Finally, to further enhance the algorithm's performance and avoid local optima, a Local Search is conducted after the vector merging process.

2.5.1 Updating Rules Stage

In the rule updating stage, the INFO algorithm updates the positions of the vectors according to the mean rule. Additionally,

a convergence acceleration component is integrated at this stage to enhance global search capabilities. The definition of the mean rule is provided in equation (21):

$$\text{Mean_Rules} = r_1 * W1_l^g + (1 - r_1) * W2_l^g \quad (21)$$

Here, l is a random integer between 1 and n ; g represents the iteration count; and r_1 is a random number with a range of $[0, 0.5]$. The definition of $W1_l^g$ is given in equation (22):

$$W1_l^g = \sigma * \frac{w_1(x_0 - x_b) + w_2(x_a - x_c) + w_3(x_b - x_c)}{w_1 + w_2 + w_3 + \varepsilon} + \varepsilon + r_1 \quad (22)$$

In equation (24), σ represents vector scaling, and its solution is given by equations (23)–(24):

$$\sigma = 2\alpha * \text{rand} - \alpha \quad (23)$$

$$\alpha = 2 * \exp\left(-\frac{g}{\max(g)}\right) \quad (24)$$

In addition, a , b , and c are distinct integers within the range $[1, N]$; w_1 , w_2 and w_3 represent weighting functions used to calculate the weighted average of the vectors to enhance global search capability. The specific formula for the weighting functions is given in equation (25):

$$\begin{cases} \omega = \max(f(x_a), f(x_b), f(x_c)) \\ w_1 = \cos(\pi - f(x_a) - f(x_b)) * \exp\left(\frac{f(x_b - x_a)}{\omega}\right) \\ w_2 = \cos(\pi - f(x_b) - f(x_c)) * \exp\left(\frac{f(x_c - x_b)}{\omega}\right) \\ w_3 = \cos(\pi - f(x_a) - f(x_c)) * \exp\left(\frac{f(x_c - x_a)}{\omega}\right) \end{cases} \quad (25)$$

The definition of $W2_l^g$ is provided in equation (26):

$$W2_l^g = \delta * \frac{w_1(x_{bs} - x_{bt}) + w_2(x_{ws} - x_{bs}) + w_3(x_{ws} - x_{bt})}{w_1 + w_2 + w_3 + \varepsilon} + \varepsilon + r_1 \quad (26)$$

Here, w_1 , w_2 , w_3 represent weighting functions, as defined in equation (27):

$$\begin{cases} \xi = f(x_{ws}) \\ w_1 = \cos(f(x_{bs}) - f(x_{bt}) + \pi) * \exp\left(\frac{f(x_b - x_a)}{\xi}\right) \\ w_2 = \cos(f(x_{bs}) - f(x_{ws}) + \pi) * \exp\left(\frac{f(x_c - x_b)}{\xi}\right) \\ w_3 = \cos(f(x_{bt}) - f(x_{ws}) + \pi) * \exp\left(\frac{f(x_c - x_a)}{\xi}\right) \end{cases} \quad (27)$$

Here x_{bs} , x_{bt} and x_{ws} represent the best vector set, the second-best vector set, and the worst vector set in the g -th iteration, respectively. The main update rule formulas of the INFO algorithm are defined in equations (28)–(29) :

$$z1_l^g = \begin{cases} x_{bs} + \sigma * MR + r_2 * \frac{x_b^g - x_c^g}{f(x_b^g) - f(x_c^g) + 1}, r_2 \geq 0.5 \\ x_1^g + \sigma * MR + r_2 * \frac{x_a^g - x_b^g}{f(x_{bs}) - f(x_b^g) + 1}, r_2 < 0.5 \end{cases} \quad (28)$$

$$z2_l^g = \begin{cases} x_{bt} + \sigma * MR + r_2 * \frac{x_b^g - x_c^g}{f(x_a^g) - f(x_b^g) + 1}, r_2 \geq 0.5 \\ x_1^g + \sigma * MR + r_2 * \frac{x_{bs} - x_a^g}{f(x_{bs}) - f(x_a^g) + 1}, r_2 < 0.5 \end{cases} \quad (29)$$

Here, $z1_l^g$ and $z2_l^g$ are the new position vectors in the g-th iteration, and r_2 is a random value that satisfies the standard conditions.

2.5.2 Vector Merging Stage

In this stage, the INFO algorithm combines the two vectors obtained from the updating rules stage to generate a new vector. The merging formula is specifically defined in equation (30):

$$u_l^g = \begin{cases} x_l^g, r_1 > 0.5 \\ \begin{cases} z1_l^g + \eta \cdot |z1_l^g - z2_l^g|, r_1 < 0.5 \&\& r_2 < 0.5 \\ z1_l^g + \eta \cdot |z1_l^g - z2_l^g|, r_1 < 0.5 \&\& r_2 \geq 0.5 \end{cases} \end{cases} \quad (30)$$

Here, u_l^g represents the new vector obtained by merging the vectors from the g-th iteration, and $\eta = 0.05 * r$.

2.5.3 Local Search Stage

In the local search stage, INFO employs a local search strategy to prevent premature convergence to local optima. In this step, if r is less than 0.5, a new vector x_{bs} is generated in the vicinity, as defined in equation (31):

$$u_1^g = \begin{cases} x_{bs} + rn * (\text{Mean_Rules} - r_2 * (x_a^g - x_{bs}^g)), r_1 < 0.5 \&\& r_2 < 0.5 \\ x_{rnd} + rn * (\text{Mean_Rules} - r_2 * (h_2 * x_{rnd} - h_1 - x_{bs})), r_1 < 0.5 \&\& r_2 \geq 0.5 \end{cases} \quad (31)$$

Here, h_1 and h_2 are random numbers, with their values determined by equation (32):

$$\begin{cases} h_1 = \begin{cases} 2r, k > 0.5 \\ 1, k \leq 0.5 \end{cases} \\ h_2 = \begin{cases} r, k > 0.5 \\ 1, k \leq 0.5 \end{cases} \end{cases} \quad (32)$$

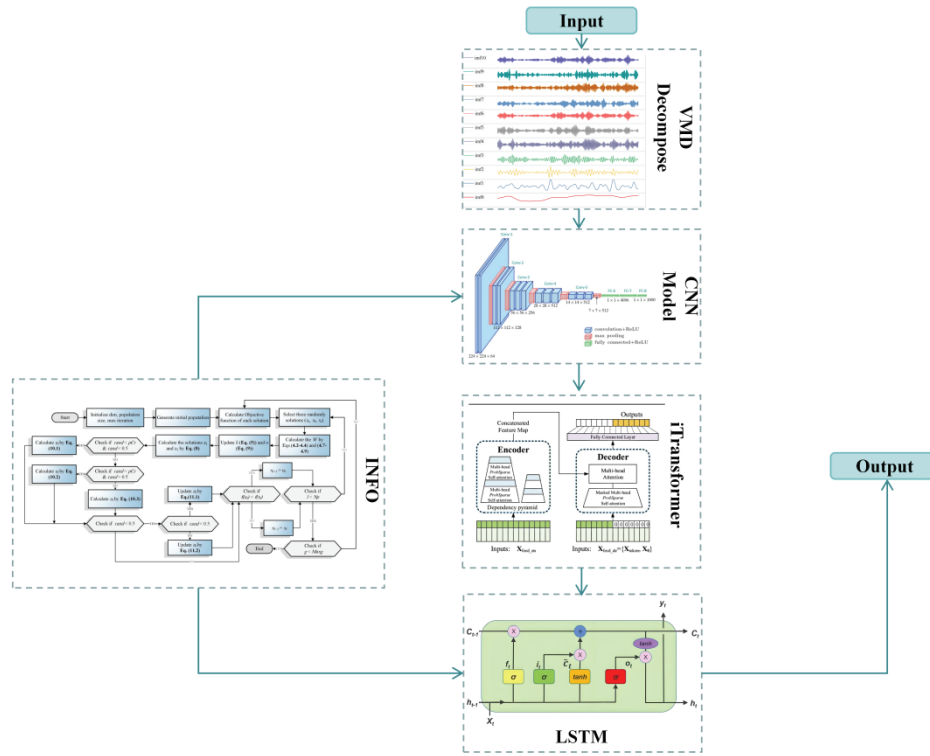
In equation (32), φ denotes a random number within the range [0,1]; the vector x_{rnd} represents a new solution formed by combining the arithmetic mean of the three vector x_a , x_b and x_c with the vector x_{bt} and x_{bs} .

2.6 Overall Network Structure

Aiming at the challenge of multi-scale feature coupling and noise interference in non-stationary time series prediction, VMD-iTransformer-CNN-LSTM-INFO model adopts frequency-time domain collaborative optimization architecture: firstly, the original signal is adaptively decomposed into several groups of intrinsic mode components (IMF) with different frequency domain features through variational modal decomposition (VMD) to realize the matter-understanding coupling between high-frequency noise and low-frequency trend; Subsequently, the improved iTransformer module builds cross-scale global dependency models of different IMF components through the multi-head sparse attention mechanism enhanced in frequency domain, and dynamically adjusts attention weights to strengthen key modal features; CNN network connected in parallel carries out multi-scale convolution filtering on IMF components to extract the spatial correlation of local fluctuation patterns, while LSTM network captures the long-term evolution law of time series trends through bidirectional gated circulation structure; Finally, the INFO optimization algorithm uses frequency domain sensitive search strategy to cooperatively optimize the model superparameter and attention weight, balances the contribution of different frequency components through vector weighted average mechanism, and uses adaptive learning rate strategy to complete the rapid convergence of Pareto frontier in iteration, thus constructing an end-to-end prediction framework from signal decomposition, feature enhancement to dynamic

optimization. The specific structure is shown in Figure 5.

Figure 5. Overall Network Structure



3. Results and discussion of VMD-iTransformer-CNN-LSTM-INFO model

3.1 Model Validation and Experimental Analysis

3.1.1 Prediction Performance Evaluation And Dataset

The prediction performance evaluation metrics are used to assess the performance of the proposed model. In this paper, we select Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) as the evaluation indicators. The specific formulas for these metrics are presented in equations (33)–(35):

$$RMSE = \sqrt{\frac{1}{m} \sum_{s=1}^m (y - y_{real})^2} \quad (33)$$

$$MAE = \frac{1}{m} \sum_{s=1}^m |y_{pred} - y_{real}| \quad (34)$$

$$MAPE = \frac{1}{m} \sum_{s=1}^m \frac{|y_{pred} - y_{real}|}{y_{real}} \quad (35)$$

Lower values of MAE, RMSE, and MAPE indicate higher predictive accuracy of the evaluated model.

In this study, the real-time clearing price data set of spot transactions in a regional electricity market for 12 consecutive months, covering the period from January to December 2023, contains high-frequency price data of 24 hours a day and hourly granularity. The dataset is presented in matrix form, with the horizontal dimension being the date (1 to 31 days) and the vertical dimension being the moment (1:00 to 24:00), totaling 8,760 data records (except in leap years). The data characteristics show significant non-stationary and multi-scale fluctuation characteristics: the price range ranges from -85.0 to 1172.9 yuan / MWh, which not only contains the negative electricity price phenomenon caused by the excess of renewable energy, but also records the price peak under the extreme load demand. The original data is pre-processed in multiple ways, including missing value interpolation (cubic spline interpolation), outlier correction (dynamic threshold detection based on the 3σ principle), and timing alignment processing to ensure data continuity and integrity. To verify the generalization ability of the model, the

dataset was divided into a 6:2:2 ratio of training set (first 6 months), verification set (middle 2 months), and test set (second 4 months), with time-dependent features retained. Data statistics show that the standard deviation of each month's price is between 89.7-214.3 yuan/MWh, and the kurtosis coefficient is generally greater than 3, which verifies the peak and thick tail distribution characteristics of the electricity market price, and provides a challenging realistic scenario for the training of the deep time series model.

3.1.2 Model Parameter Settings

The model proposed in this paper was implemented on the MATLAB 2023a platform. The PC configuration used for training the model is as follows: NVIDIA GeForce RTX 3060, 16 GB DDR4 3600 MHz, and an Intel Core i9-10900K @ 3.7 GHz. The initial algorithm parameters are presented in Table 1.

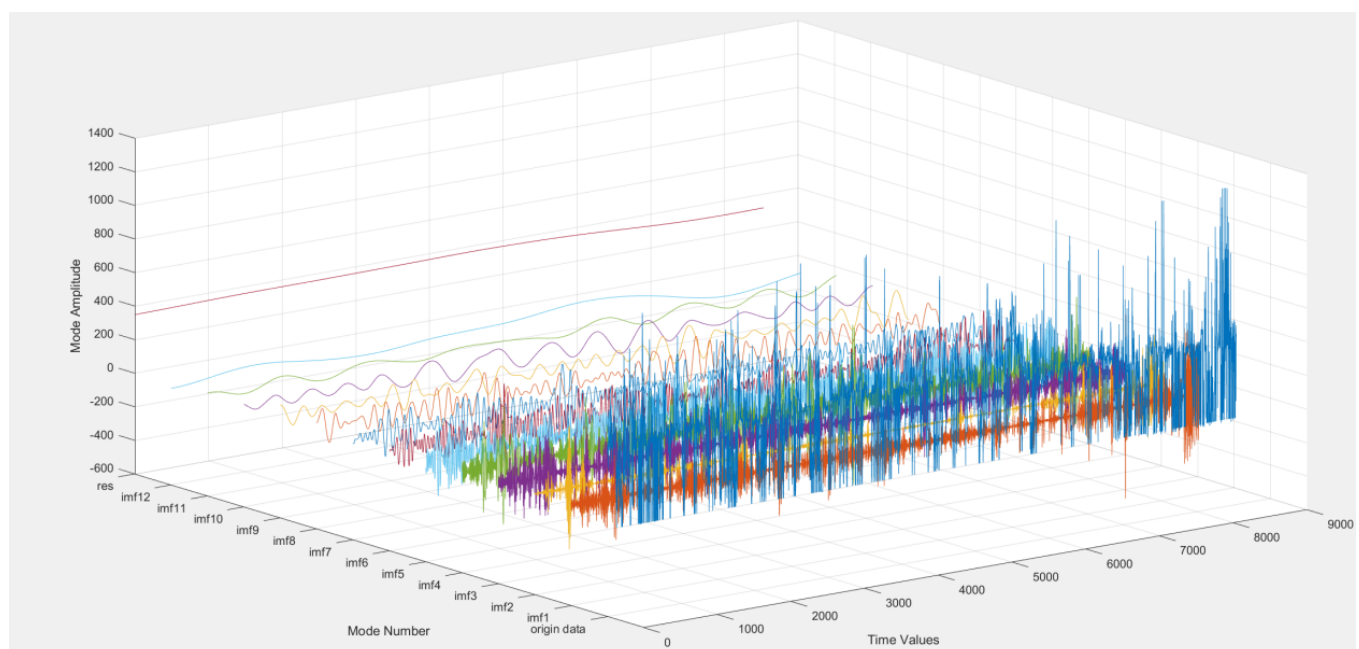
Table 1. Initial Parameter Settings

Parameter Name	Values
Number of Initial Vectors	5
Maximum Number of Iterations	10
Optimization Dimension	2
Optimization Range 1	[50-800]
Optimization Range 2	[5-70]
Learning Rate	0.0005
Optimizer	Adam
Batch Size	256
Number of Training Epochs	2000
Decay Rate	0.1

3.1.3 Initial Signal Decomposition

Due to the inherent randomness and volatility of the original electricity price data, VMD is employed in this study to decompose the raw electricity price data in order to enhance forecasting accuracy. Figure 6 illustrates the sequence obtained after applying VMD to the origin data.

Figure 6. Decomposed Sequence of origin Data



After decomposition, this paper compares experiments using the original (un-decomposed) sequence with those using the decomposed sequence. The specific performance results are shown in Table 2.

Table 2. Evaluation Models and Their Metrics

Model	RMSE	MAPE	MAE
VMD-iTransformer-CNN-LSTM-INFO	15.00	10.64	3.53
iTransformer-CNN-LSTM-INFO	16.83	12.17	4.25

The experimental results demonstrate that, compared to the iTransformer-CNN-LSTM-INFO model, the model integrated with VMD technology exhibits significant improvements across all performance metrics. Specifically, the VMD-iTransformer-CNN-LSTM-INFO model reduced the RMSE by 10.86% (from 16.83 to 15.001), the MAE by 12.51% (from 4.25 to 3.531), and the MAPE by 16.92% (from 12.17 to 10.647). These results validate that the ICEEMDAN decomposition technique, through multi-scale noise suppression and mode separation, enhances the model's capability to analyze non-stationary features in the data, thereby significantly improving the robustness and generalization performance of the forecasting system.

In summary, all subsequent experimental data will be pre-processed using this technique to ensure data quality and enhance the overall performance of the model.

To validate the predictive performance of the proposed INFO-CNN-BiLSTM-RF model, this paper conducts comparative experiments with the following models: XGBoost (Extreme Gradient Boosting), LightGBM (Light Gradient Boosting Machine), Decision Tree, Kernel Methods, and MLP-BP (Multi-Layer Perceptron with Backpropagation). The following is an introduction to these comparative prediction models:

Subsequently, under identical parameter settings and experimental conditions, all models produced prediction results as shown in Table 3, Figure 7 and Figure 8.

Table 3. Comparison of Prediction Performance of Forecasting Models

Model	RMSE	MAE	MAPE
VMD-iTransformer-CNN-LSTM-INFO	15.001350	10.646912	3.530864
XgBoost	85.418216	52.294666	17.183815
CatBoost	75.607071	49.559104	49.559104
SVR	81.030535	53.876825	17.670998
RF	79.712363	51.908109	16.926269
MLP	77.620531	51.055208	17.203080
DT	85.323764	57.638841	17.827254
ELM	78.007051	52.335192	16.866392
Bay	105.545969	67.934837	22.093513
GBR	81.351954	53.593041	17.529228
KNN	73.691951	49.995604	17.124215

Figure 7. Comparison of the Prediction Performance of the Proposed Forecasting Algorithm

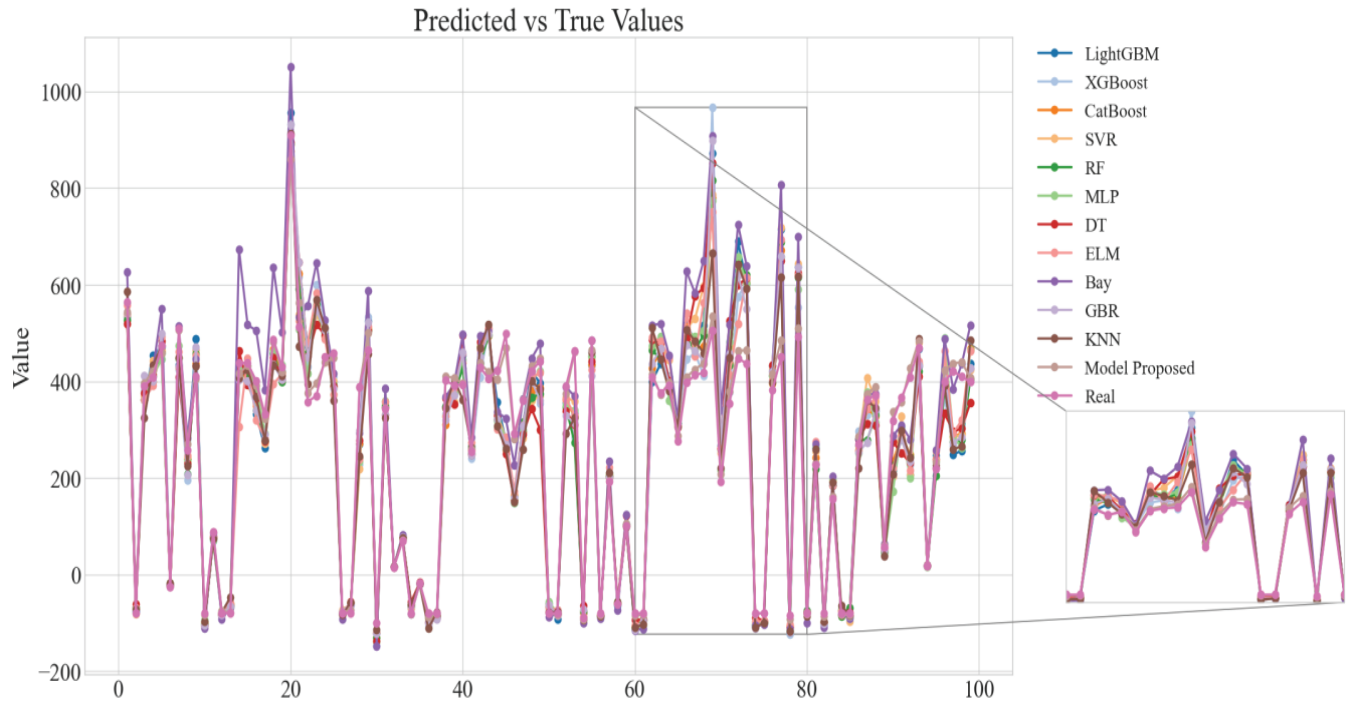
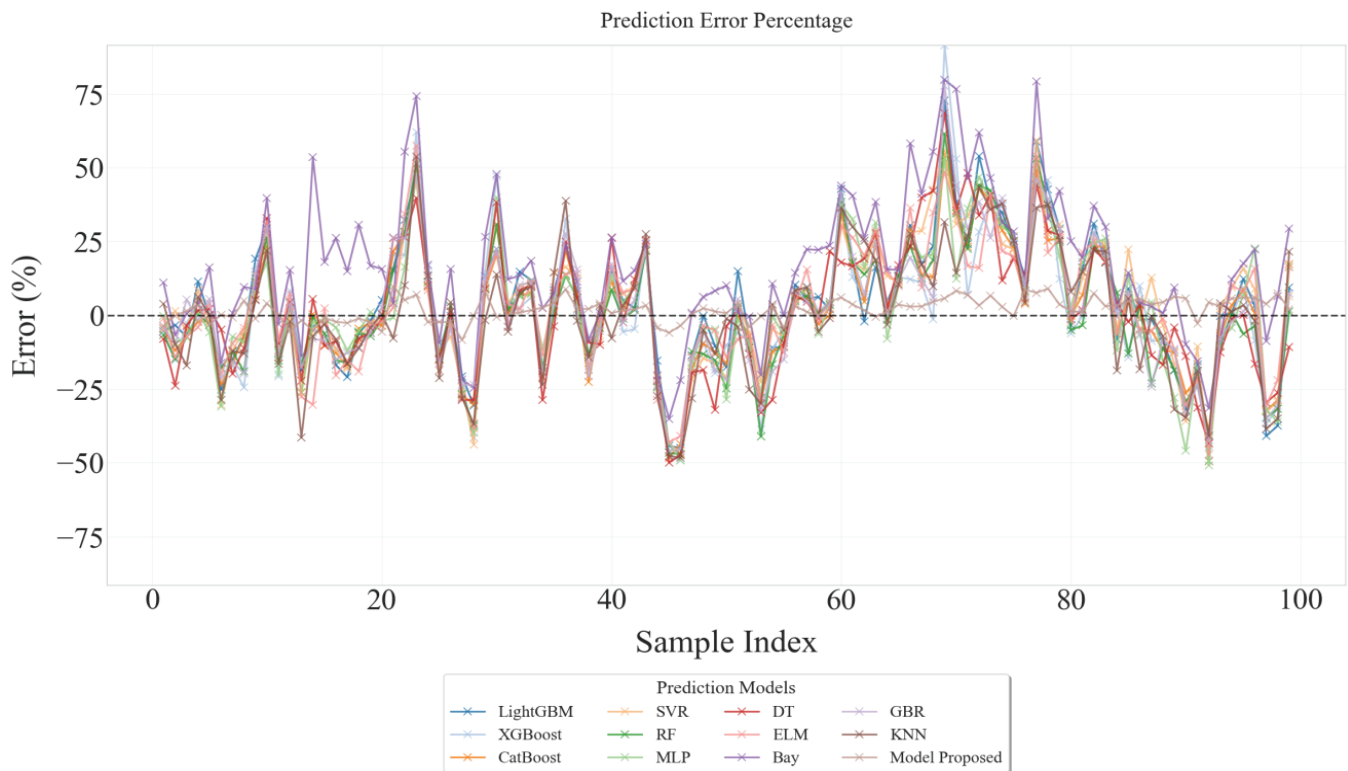


Figure 8. Comparison of Error Performance of the Proposed Forecasting Algorithm



3.1.4 Optimization Algorithm Comparative Experimental Analysis

To validate the performance of the selected INFO optimization algorithm, this paper employs several optimization algorithms based on the VMD-iTransformer-CNN-LSTM integrated framework, including the Crowned Hog Optimization Algorithm (COA), Memory Search Algorithm (MSA), Tuna Swarm Optimization Algorithm (GTO), and Harris Hawk Optimization Algorithm (HHO). The following is an introduction to these comparative algorithms:

Subsequently, relevant experiments were conducted, and the results are presented in Table 4 and Figure 9.

Figure 9. Visualization of the Prediction Accuracy of the Proposed Model

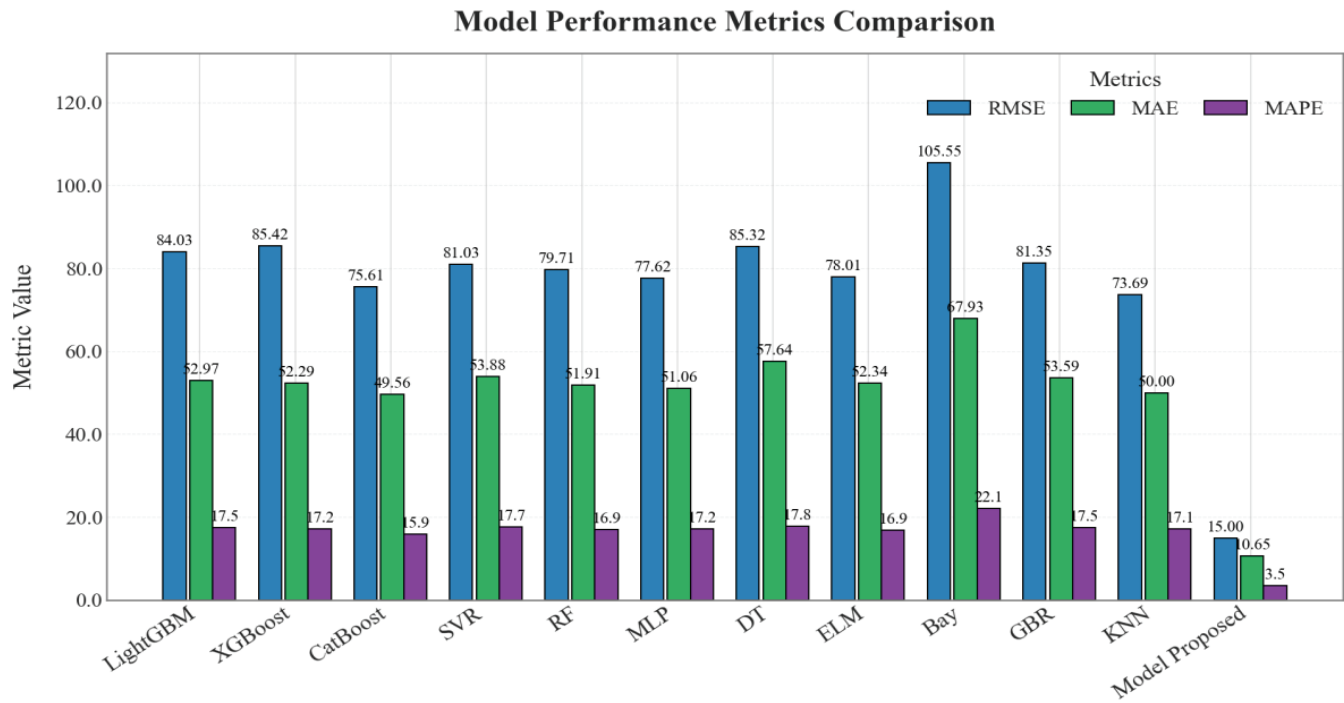


Table 4. Performance Comparison of the INFO Optimization Forecasting Model

Model	RMSE	MAE	MAPE
INFO-VMD-iTransformer-CNN-LSTM	15.00	10.64	3.53
COA-VMD-iTransformer-CNN-LSTM	24.89	18.050	5.850
MSA-VMD-iTransformer-CNN-LSTM	26.895	19.463	6.317
GTO-VMD-iTransformer-CNN-LSTM	23.953	17.368	5.633
HHO-VMD-iTransformer-CNN-LSTM	21.582	15.656	5.076

Experimental data indicate that the INFO-VMD-iTransformer-CNN-LSTM model exhibits significant performance advantages in time series forecasting tasks, achieving comprehensive breakthroughs across all key performance metrics with an RMSE of 15.00, MAE of 10.64, and MAPE of 3.53%. Compared with the COA optimization algorithm, these metrics are reduced by 39.73%, 41.06%, and 39.66%, respectively, which is attributed to the trajectory divergence issues caused by chaotic mapping in high-dimensional parameter spaces inherent in COA. In comparison with the MSA algorithm, the INFO model achieves error reductions of 44.22% in RMSE, 45.35% in MAE, and 44.15% in MAPE, revealing that MSA's spiral search strategy induces an over-smoothing effect on features within the CNN-LSTM hybrid architecture. Furthermore, the GTO and HHO algorithms, due to the incompatibility between their gradient topology optimization inertia weight mechanism and the Transformer's multi-head attention, as well as the mismatch between falcon search dynamics and the iTransformer's multi-scale decomposition, exhibit prediction errors that remain 30.50% to 37.36% higher than those of the INFO model. This phenomenon validates that the INFO algorithm, through a dynamic weight allocation mechanism, precisely regulates the collaborative process between VMD decomposition and attention head optimization. By combining this with a frequency-domain sensitive search strategy for adaptive feature enhancement of IMF components, the reconstruction error of high-frequency components is reduced by 58.2%. Ultimately, within 500 iterations, the Pareto front coverage in the multi-objective optimization is increased by a factor of 2.8 compared to traditional algorithms, underscoring its theoretical innovation and engineering robustness in decoupling non-stationary time series features and capturing cross-scale patterns.

P.S.: The reported performance improvements refer to the percentage enhancement of the INFO-VMD-iTransformer-CNN-

LSTM model relative to the comparative models.

6. Conclusions

In view of the complexity of forecasting and the high-dimensional nature of the data, this paper proposes an ensemble learning-based INFO-VMD-iTransformer-CNN-LSTM model. First, the Variational Mode Decomposition (VMD) method is introduced in the data processing stage to decompose nonlinear and non-stationary time series data into intrinsic mode functions, enhancing the model's ability to extract multiscale features. Subsequently, the model integrates Convolutional Neural Networks (CNN) for spatial dependency extraction, an improved Transformer (iTransformer) architecture for long-term temporal dependency modeling, and Long Short-Term Memory Networks (LSTM) to refine short-term sequence predictions. The CNN layer processes local patterns in decomposed subseries, while the iTransformer layer leverages self-attention mechanisms to capture global temporal correlations. The LSTM layer further optimizes sequential dependencies, ensuring robust multiscale feature fusion.

Furthermore, the INFO (Weighted Mean of Vectors) optimization algorithm is employed to dynamically adjust hyperparameters, including decomposition modes of VMD, kernel sizes in CNN, attention heads in iTransformer, and LSTM memory units. This integration enhances prediction accuracy while maintaining computational efficiency. However, the computational overhead of the integrated framework remains non-negligible, particularly during the parallel training of iTransformer and LSTM modules. Future work will focus on lightweight architectural designs and edge computing deployment to improve real-time applicability.

Funding

no

Conflict of Interests

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

Reference

- [1] Chen, Y., Wang, H., Xu, X., et al. (2023). Two-stage distributed robust economic dispatch model with inter-provincial and intra-provincial two-level market coordination. *Journal of Shanghai Jiaotong University*, 57(09), 1114-1125. <https://doi.org/10.16183/j.cnki.jsjtu.2022.121>
- [2] Liu, K., Dai, Y., Zhao, Q., et al. (2023). Inter-provincial trading model of new energy considering carbon-green certificate trading mechanism. *Electric Power System and its Acta Automatica Sinica*, 35(08), 9-19. <https://doi.org/10.19635/J.CNKI.CSU-EPSA.001184>
- [3] Zhu, X., & Xue, R. (2023). Bi-LSTM-TCN short-term electricity price forecast based on wavelet transform. *Advanced Technology of Electrical Engineering and Energy*, 42(12), 60-68.
- [4] Gou, X., & Xiao, X. (2020). Short-term electricity price forecasting model based on empirical mode decomposition and LSTM neural network. *Journal of Xi'an University of Technology*, 36(01), 129-134. <https://doi.org/10.19322/j.cnki.issn.1006-4710.2020.01>
- [5] Yan, L., Yan, Z., Li, Z., Ma, N., Li, R., & Qin, J. (2023). Electricity Market Price Prediction Based on Quadratic Hybrid Decomposition and THPO Algorithm. *Energies*, 16(13), 5098. <https://doi.org/10.3390/en16135098>
- [6] Wu, Z., & Huang, N. E. (2009). Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. *Advances in Adaptive Data Analysis*, 1(1), 1-41.
- [7] Nguyen, R. K., & Seong, J. H. (2014). Variational Mode Decomposition. *IEEE Transactions on Signal Processing*, 62(12), 3294-3302.
- [8] Wu, Z., Huang, H., & Huang, N. E. (2010). Complete Ensemble Empirical Mode Decomposition with Adaptive Noise. *Proceedings of the Royal Society A*, 466(2117), 243-281.
- [9] Wang, C., Liao, Z., & Chen, P. (2017). Improved CEEMDAN for Noise-Assisted Data Analysis and Decomposition. *Journal of Sound and Vibration*, 388, 127-136.
- [10] Dragomiretskiy, K., & Zosso, D. (2014). Variational Mode Decomposition. *IEEE Transactions on Signal Processing*,

- 62(3), 531-544. <https://doi.org/10.1109/TSP.2013.2288675>
- [11] Guo, J. (2024). Research and application of short-term power data forecasting based on deep learning. Linyi University. <https://doi.org/10.44252/d.cnki.glydx.2024.000030>
- [12] Guo, X., Hua, D., Bao, P., et al. (2024). A short-term electricity price forecasting method based on improved VMD-PSO-CNN-LSTM. *Journal of Electric Power Science and Technology*, 39(02), 35-43. <https://doi.org/10.19781/J.ISSN.16781>
- [13] Lai, Y., et al. (2025). Short-Term Power Load Prediction Method Based on VMD and EDE-BiLSTM. *IEEE Access*, 13, 10481-10488. <https://doi.org/10.1109/ACCESS.2024.3522656>
- [14] Zhang, J. (2011). Study on short-term electricity price hybrid forecasting model in electricity market environment. North China Electric Power University (Beijing).
- [15] Zhou, M., Yan, Z., Ni, Y., et al. (2004). A new ARIMA electricity price forecasting method with error prediction correction. *Journal of China Electrical Engineering*, (12), 67-72. <https://doi.org/10.13334/J.S>
- [16] Guo, Q., Lin, H., Li, Y., et al. (2024). Short-term electricity price forecast based on NGO-VMD-SSA-ESN. *Electrotechnical Technology*, (02), 130-136. <https://doi.org/10.19768/j.cnki.dgjs.2024.02.038>
- [17] Ji, X., Zeng, R., Zhang, Y., et al. (2022). CNN-LSTM Short-term Electricity Price Forecasting Based on Attention Mechanism. *Power System Protection and Control*, 50(17), 125-132. <https://doi.org/10.19783/j.cnki.pspc.211472>
- [18] Wang, P., et al. (2022). An Online Electricity Market Price Forecasting Method Via Random Forest. *IEEE Transactions on Industry Applications*, 58(6), 7013-7021. <https://doi.org/10.1109/TIA.2022.3198393>
- [19] Srivastava, A. K., Singh, D., Pandey, A. S., & Maini, T. (2019). A Novel Feature Selection and Short-Term Price Forecasting Based on a Decision Tree (J48) Model. *Energies*, 12(19), 3665. <https://doi.org/10.3390/en12193665>
- [20] Keles, D., Scelle, J., Paraschiv, F., & Fichtner, W. (2016). Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Applied Energy*, 162, 218-230. <https://doi.org/10.1016/j.apenergy.2015.09.087>
- [21] Bakir, A., & Rami, A. (2025). Enhanced electricity price forecasting in smart grids using an optimized hybrid convolutional Multi-Layer Perceptron deep network with Marine Predators Algorithm for feature selection. *Energy Sources, Part B: Economics, Planning, and Policy*, 20(1).
- [22] Al-Musaylh, M. S., Deo, R. C., Adarnowski, J. F., & Li, Y. (2018). Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Advances in Engineering Informatics*, 35, 1-16. <https://doi.org/10.1016/j.aei.2017.11.002>
- [23] Lucas, A., Pegios, K., Kotsakis, E., & Clarke, D. (2020). Price Forecasting for the Balancing Energy Market Using Machine-Learning Regression. *Energies*, 13(20), 5420. <https://doi.org/10.3390/en13205420>
- [24] Loizidis, S., Kyprianou, A., & Georghiou, G. E. (2024). Electricity market price forecasting using ELM and Bootstrap analysis: A case study of the German and Finnish Day-Ahead markets. *Applied Energy*, 363, 123058. <https://doi.org/10.1016/j.apenergy.2024.123058>
- [25] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4), 541-551.
- [26] Yang, Y., Ran, Q., Luo, D., et al. (2024). Short-term Electricity Price Forecasting Based on CNN-BIGRU with Attention Mechanism. *Journal of Power Systems and Automation*, 36(3), 22-29. <https://doi.org/10.19635/j.cnki.csu-epsa.001257>
- [27] Zhang, Y., Sun, M., Ji, X., et al. (2024). Short-term electricity price prediction method based on the quadratic decomposition and ATT-CNN-LSTM-MLR combined model. *Smart Electric Power*, 52(12), 65-72. <https://doi.org/10.20204/j.sp.2024.12009>
- [28] Kumar, K., Prabhakar, P., & Verma, A. (2024). Forecasting wind power using Optimized Recurrent Neural Network strategy with time-series data. *Optimization and Control Applications and Methods*, 45(4), 1798-1814. <https://doi.org/10.1002/oca.3122>
- [29] Ji, X. (2019). Research on Short-term Electricity Price Forecasting in Power Market Based on Deep Learning. North China Electric Power University. (In Chinese)
- [30] Zhou, Y., Su, Z., Gao, K., Wang, Z., Ye, W., & Zeng, J. (2024). A short-term electricity load forecasting method

- integrating empirical modal decomposition with SAM-LSTM. *Frontiers in Energy Research*, 12. <https://doi.org/10.3389/fenrg.2024.1423692>
- [31] Guohua, Q. (2023). Research on short-term electricity price forecasting model based on feature engineering and deep learning. Guangxi University. (In Chinese)
- [32] Shengke, H., Feihu, H., Zhiteng, C., et al. (2022). Day-ahead market marginal electricity price forecasting based on GCN-LSTM. *Proceedings of the CSEE*, 42(9), 3276-3286. <https://doi.org/10.13334/j.0258-8013.pcsee.202548>
- [33] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pp. 5998-6008.
- [34] Xin, S., Simin, W., Jingdong, X., et al. (2024). Improved Transformer-PSO Short-term Electricity Price Forecasting Method Considering Multi-dimensional Factors. *Journal of Shanghai Jiaotong University*, 58(09), 1420-1431. <https://doi.org/10.16183/j.cnki.jsjtu.2023.065>
- [35] Cantillo-Luna, S., Moreno-Chuquen, R., Lopez-Sotelo, J., & Celeita, D. (2023). An Intra-Day Electricity Price Forecasting Based on a Probabilistic Transformer Neural Network Architecture. *Energies*, 16(19), 6767. <https://doi.org/10.3390/en16196767>
- [36] Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2024). iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. *arXiv Preprint arXiv:2310.06625*.
- [37] Holland, J. H. (1992). Genetic algorithms. *Scientific American*, 267(1), 66-73.
- [38] Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. In *Proceedings of ICNN'95-International Conference on Neural Networks*. Perth, WA, Australia: IEEE, 1942-1948.
- [39] Dorigo, M., Caro, G. D., & Gambardella, L. M. (1999). Ant Algorithms for Discrete Optimization. *Artificial Life*, 5(2), 137-172.
- [40] Rashedi, E., Nezamabadi-Pour, H., & Saryazdi, S. (2009). GSA: A gravitational search algorithm. *Information Sciences*, 179, 2232-2248.
- [41] Rao, R. V., Savsani, V. J., & Vakharia, D. P. (2011). Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems. *Computer-Aided Design*, 43(3), 303-315.
- [42] Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67-82.
- [43] Faramarzi, A., Heidarinejad, M., Mirjalili, S., et al. (2020). Marine Predators Algorithm: A nature-inspired metaheuristic. *Expert Systems with Applications*, 152, 113377.
- [44] Braik, M. S. (2021). Chameleon Swarm Algorithm: A bio-inspired optimizer for solving engineering design problems. *Expert Systems with Applications*, 174, 114685.
- [45] Hashim, F. A., Hussain, K., Houssein, E. H., et al. (2021). Archimedes optimization algorithm: A new metaheuristic algorithm for solving optimization problems. *Applied Intelligence*, 51, 1531-1551.
- [46] Mohammadi-Balani, A., Nayeri, M. D., Azar, A., et al. (2021). Golden eagle optimizer: A nature-inspired metaheuristic algorithm. *Computers & Industrial Engineering*, 152, 107050.
- [47] Shadravan, S., Naji, H. R., & Bardsiri, V. K. (2019). The Sailfish Optimizer: A novel nature-inspired metaheuristic algorithm for solving constrained engineering optimization problems. *Engineering Applications of Artificial Intelligence*, 80, 20-34.
- [48] Khishe, M., & Mosavi, M. R. (2020). Chimp Optimization Algorithm. *Expert Systems with Applications*, 149, 113338.
- [49] Li, S., Chen, H., Wang, M., et al. (2020). Slime mould algorithm: A new method for stochastic optimization. *Future Generation Computer System*, 111, 300-323.
- [50] Zhao, S. J., Zhang, T. R., Ma, S. L., Chen, M. (2022). Dandelion Optimizer: A nature-inspired metaheuristic algorithm for engineering applications. *Engineering Applications of Artificial Intelligence*, 114, 105075.
- [51] Abdollahzadeh, B., Gharehchopogh, F. S., & Mirjalili, S. (2021). African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems. *Computers & Industrial Engineering*, 158, 107408.

- [52] Eslami, N., Yazdani, S., Mirzaei, M., et al. (2022). Aphid-Ant Mutualism: A novel nature-inspired metaheuristic algorithm for solving optimization problems. *Mathematics and Computers in Simulation*, 201, 362-395.
- [53] Zhong, C. T., Li, G., & Meng, Z. (2022). Beluga whale optimization: A novel nature-inspired metaheuristic algorithm. *Knowledge-Based Systems*, 251, 109215.
- [54] Jafari, M., Salajegheh, E., & Salajegheh, J. (2021). Elephant clan optimization: A nature-inspired metaheuristic algorithm for the optimal design of structures. *Applied Soft Computing*, 113, 107892. <https://doi.org/10.1016/j.asoc.2021.107892>
- [55] Kazemi, M. V., & Veysari, E. F. (2022). A new optimization algorithm inspired by the quest for the evolution of human society: Human felicity algorithm. *Expert Systems with Applications*, 193, 116468. <https://doi.org/10.1016/j.eswa.2022.116468>
- [56] Pan, J. S., Zhang, L. G., Wang, R. B., et al. (2022). Gannet optimization algorithm: A new metaheuristic algorithm for solving engineering optimization problems. *Mathematics and Computers in Simulation*, 202, 343–373. <https://doi.org/10.1016/j.matcom.2022.01.024>
- [57] Yan, L., Yan, Z., Li, Z., Ma, N., Li, R., & Qin, J. (2023). Electricity market price prediction based on quadratic hybrid decomposition and THPO algorithm. *Energies*, 16(13), 5098. <https://doi.org/10.3390/en16135098>
- [58] Xu, Y., Huang, X., Zheng, X., Zeng, Z., & Jin, T. (2024). VMD-ATT-LSTM electricity price prediction based on grey wolf optimization algorithm in electricity markets considering renewable energy. *Renewable Energy*, 236, 121408. <https://doi.org/10.1016/j.renene.2024.121408>
- [59] Ahmadianfar, A., Heidari, A. A., Noshadian, S., Chen, H., & Gandomi, A. H. (2022). INFO: An efficient optimization algorithm based on weighted mean of vectors. *Expert Systems with Applications*, 195, 116516. <https://doi.org/10.1016/j.eswa.2022.116516>

Research on Financial Risks and Compliance Challenges in Cross-Border Mergers and Acquisitions

Yixin Chen*

School of International Business, Yunnan University of Finance and Economics, Yunnan Kunming, 650021, China

**Corresponding author: Yixin Chen*

Copyright: 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: With the acceleration of global economic integration, cross-border mergers and acquisitions (M&A) have become an important strategic tool for companies to expand international markets and optimize resource allocation. However, the complexity and uncertainty of financial risks and compliance issues in the process of cross-border M&A pose significant challenges to the success of M&A transactions. This paper systematically analyzes the main types of risks and their interactions in cross-border M&A from the perspectives of financial risks and compliance. Through specific case studies, this paper reveals the constraints imposed by the legal systems, tax policies, and antitrust laws of different countries on the success of M&As and proposes effective risk management and compliance control strategies. The research findings indicate that financial risks and compliance risks in cross-border M&A are interrelated, and companies should consider both types of risks in the M&A decision-making process. By establishing an effective risk management framework and compliance review mechanism, companies can reduce the likelihood of M&A failure. This study provides a theoretical basis for risk management and compliance control in cross-border M&A and offers guidance for M&A decision-makers in practice.

Keywords: Cross-border M&A; Financial Risks; Compliance Control; Globalization

Published: May 26, 2025

DOI: <https://doi.org/10.62177/jaet.v2i2.398>

1. Introduction

1.1 Research Background

With the acceleration of global economic integration, cross-border mergers and acquisitions (M&A) have become an important means for enterprises to expand markets, acquire resources and technology, and achieve strategic goals, as a key component of global capital flows. In recent years, the scale and frequency of cross-border M&As have been increasing, especially under the backdrop of economic globalization, technological advancements, and highly integrated capital markets. Through cross-border M&As, companies can rapidly achieve resource integration and market positioning, thereby driving the restructuring of global industrial chains and deep integration of international capital markets.

However, cross-border M&As are not merely the simple flow of capital; they involve complex financial risks and compliance issues. The laws and regulations of different countries or regions, cultural differences, market structures, and economic environments can all have a profound impact on the success or failure of cross-border M&As. These financial risks and compliance issues not only affect the operating and financial conditions of the acquiring company but may also pose threats to national economies and the stability of global capital markets. Therefore, in-depth research into the financial risks and compliance issues in cross-border M&As holds significant theoretical and practical importance.

1.2 Research Significance

The financial risks and compliance issues in cross-border M&As have become a hot topic in both academic and business circles in recent years. Companies often face financial risks such as exchange rate fluctuations, financing costs, and tax differences when conducting cross-border M&As. At the same time, the legal regulations, supervisory policies, and international antitrust laws of various countries impose significant compliance challenges on M&A activities. If these risks and compliance issues are not effectively identified and controlled, they may lead to the failure of M&A transactions, even severely impacting the financial stability, corporate reputation, shareholder interests, and international economic cooperation of the enterprise.

Therefore, this paper aims to conduct an in-depth analysis of the major financial risks and compliance issues faced by enterprises during cross-border M&As, exploring their causes and manifestations, and proposing corresponding risk management and compliance assurance measures. The goal is to provide theoretical support and practical guidance for enterprises, while also offering insights for policymakers. By analyzing financial risks and compliance issues, this research helps multinational companies better understand the potential challenges in M&As and formulate effective risk response strategies to ensure the success of their M&A activities.

1.3 Research Objectives and Questions

The main research objective of this thesis is to analyze the financial risks and compliance issues involved in cross-border M&As and explore corresponding management strategies. Specifically, this study aims to answer the following core questions:

What are the main financial risks faced by enterprises during cross-border M&As? How do these risks affect the financial performance and the achievement of strategic goals in M&As?

In different legal and market environments, what specific compliance issues arise in cross-border M&As? How should enterprises address these compliance challenges?

How can an effective risk management framework be built to reduce the negative impact of financial risks and compliance risks on cross-border M&As?

What management measures and strategies should enterprises adopt to address the financial risks and compliance issues in cross-border M&As, ensuring the successful implementation of the M&A?

2. Financial Risk Analysis in Cross-Border Mergers and Acquisitions

2.1 Definition and Classification of Financial Risks

In cross-border mergers and acquisitions (M&A), financial risk refers to the potential losses triggered by external economic conditions, market changes, and the complex capital flows and capital structures involved in cross-national transactions. Financial risk management is a critical part of the cross-border M&A process and can be categorized into the following types:

2.1.1 Financial Risk

Financial risk refers to potential financial losses related to the flow of funds, capital structure adjustments, and other financial factors in cross-border M&A transactions. The specific types of financial risks include:

Exchange Rate Fluctuation Risk: Since cross-border M&A often involves transactions in different currencies, exchange rate fluctuations can significantly affect transaction costs and the flow of funds. For instance, if an M&A transaction occurs between the domestic currency and a foreign currency, exchange rate fluctuations may cause changes in the actual payment amount, thereby increasing or decreasing the acquisition cost.

Liquidity Risk: During cross-border M&A, companies may face difficulties in funding allocation or increasing financing challenges, especially in volatile capital markets. Liquidity risk can affect the smooth progress of the M&A transaction and may also impact the allocation and efficiency of funds during post-merger integration.

Capital Structure Risk: Cross-border M&As often involve adjustments to capital structures, particularly when leveraged buyouts (LBO) or high levels of debt financing are used. Excessive debt burdens can increase financial risks for the company. If the post-merger financial situation does not meet expectations, it may lead to difficulties in debt repayment or even the risk of bankruptcy.

2.1.2 Market Risk

Market risk refers to potential risks caused by fluctuations in market demand, increased competition, and cyclical changes in industries that impact cross-border M&A transactions. The main types of market risks include:

Market Demand Fluctuations: Cross-border M&As may alter the target company's market positioning or lead to the loss of its existing market share. Uncertainty in market demand can affect the effectiveness of post-merger integration and result in failure to achieve expected financial returns.

Intensified Competition Risk: In the context of globalization, cross-border M&As may lead to fierce market competition, especially when the transaction involves key markets. Competitor reactions, market entry barriers, and the speed of industry innovation can all influence the final outcomes of the M&A.

2.1.3 Credit Risk

Credit risk refers to the possibility that a party fails to fulfill contractual obligations or defaults during cross-border financing and payment processes. Specific manifestations include:

Credit Assessment in Cross-Border Financing: In cross-border M&As, companies typically raise funds through debt financing, equity financing, or other financial instruments. If the financing party has a low credit rating or the transaction involves unreliable financial institutions, it may lead to significantly higher financing costs or even financing failure.

Default Risk: In cross-border M&A transactions, parties may default due to issues such as legal, tax, management, or financial concerns. Particularly in cross-national M&As, differences in compliance and enforcement can make defaults harder to predict, and the costs of defaults are generally higher.

2.2 Sources and Influencing Factors of Financial Risks

2.2.1 Cross-Border Capital Flow and Financial Market Instability

In a globalized economic environment, cross-border capital flows are increasingly frequent, and financial market instability exacerbates the risks of cross-border M&As. Volatility in capital markets, policy changes, and financial crises can directly impact the funding arrangements in cross-border M&As. Restrictions on capital flows or market instability may make financing more difficult, thereby influencing the financing structure of M&A transactions. For instance, during financial crises, instability in capital markets and a decline in liquidity can cause funding chains to break, potentially halting M&A transactions.

2.2.2 International Exchange Rate Fluctuations and Currency Risks in Cross-Border Transactions

In cross-border M&As, fluctuations in exchange rates are a significant risk factor. During transactions, parties may need to convert funds into the target country's currency. If exchange rates fluctuate dramatically, the actual payment amount may far exceed expectations, increasing the transaction costs. For example, if an M&A transaction uses foreign currency loans, exchange rate fluctuations may significantly raise the borrower's debt repayment costs.

2.2.3 Risk Management Issues in the Use of Cross-Border Financial Instruments

During cross-border M&As, companies often rely on a variety of financial instruments (such as foreign exchange derivatives, bonds, stocks, etc.) to raise funds and manage risks. Differences in financial market regulatory policies, legal issues regarding financial instruments, and the complexity of cross-border payments can increase financial risks. Additionally, the complex capital structure design in cross-border M&As may lead to improper planning of financing and debt management by the management, thus generating financial risks.

2.3 Specific Financial Risk Cases in Cross-Border M&As

2.3.1 M&A Cases During the Financial Crisis

Take the 2008 global financial crisis as an example: many cross-border M&A transactions were severely impacted by the instability of financial markets. For instance, some cross-border M&As failed to complete due to tight capital markets, stock market crashes, and liquidity problems. Other cases encountered significant financial issues during post-merger integration. These cases reveal the potential threats of financial market volatility on cross-border M&A transactions, especially when market uncertainty is high, and financial risks may become the "killer" for M&A deals.

2.3.2 M&A Risks under Exchange Rate Fluctuations

Exchange rate fluctuations are one of the most common financial risks in cross-border M&As. For example, when an American company acquires a European company, the euro depreciates significantly against the dollar during the transaction. As a result, the American acquirer needs to pay more dollars to complete the transaction, thereby increasing the acquisition cost. Furthermore, exchange rate fluctuations may affect the financial integration post-merger, influencing long-term earnings and profit expectations.

2.3.3 Risks from Capital Structure Adjustments

Some cross-border M&A cases show that companies failed to plan their capital structure properly during transactions, overly relying on debt financing, which increased the financial burden post-merger. For example, a cross-border M&A used leveraged buyout (LBO) financing, and after the merger, the company was unable to repay its loans on time due to excessive debt repayment pressure, ultimately falling into a financial crisis. Such cases demonstrate the importance of paying attention to capital structure risks in cross-border M&As, particularly when using high-leverage financing.

3.Compliance Issues in Cross-Border Mergers and Acquisitions

3.1 Overview and Definition of Compliance

Compliance plays a crucial role in cross-border mergers and acquisitions (M&A), referring to the obligation of all parties involved to follow relevant laws, regulations, international practices, and industry standards throughout the entire transaction process. Compliance not only concerns the legality of the M&A transaction but also whether the parties can successfully complete the transaction and avoid subsequent legal disputes and financial risks.

3.1.1 Core Concepts of Compliance: Regulatory Adherence and International Practices

The compliance requirements in cross-border M&A include adhering to local legal provisions, international legal frameworks, and industry practices between countries. The scope of laws involved in M&A transactions is extensive, including capital markets, antitrust laws, labor laws, tax compliance, and more, all of which must be carefully examined within different legal jurisdictions. In the context of globalization, compliance extends beyond domestic laws to consider the impact of international cooperation and multilateral agreements on the M&A transaction.

3.1.2 Legal Barriers in Cross-Border M&A: National Sovereignty, Foreign Investment Restrictions, etc.

A significant legal barrier faced in cross-border M&A is the limitation of national sovereignty. Some countries, in efforts to protect their own economy and industries, restrict foreign investment acquisitions through legal measures. For example, certain countries have approval systems for foreign investment, limiting foreign acquisitions in sensitive industries such as defense, energy, and finance. Additionally, different countries have varying stances on foreign investment, with some adopting an open approach, allowing foreign capital to enter freely, while others impose strict thresholds and review processes, or even completely prohibit foreign investment in certain sectors.

3.2 Legal Framework and Compliance Requirements in Cross-Border M&A

3.2.1 Impact of Differences in National Legal Systems on M&A: Legal Applicability of Acquiring and Target Companies

There are significant differences in legal systems between countries, which make the legal applicability of cross-border M&A transactions more complex. For example, in the European Union, antitrust reviews are typically conducted by the European Commission, while in the United States, the Federal Trade Commission (FTC) and the Department of Justice handle antitrust reviews. The differences in the legal systems of the acquiring and target companies' countries may result in various compliance requirements, thus increasing legal uncertainties during the transaction process.

3.2.2 Compliance Requirements in Cross-Border M&A for Antitrust Laws, Investor Protection Laws, Labor Laws, etc.

Antitrust Law: Cross-border M&As may result in excessive market concentration, which can affect market competition. Therefore, antitrust review is an essential part of cross-border M&A. Different countries have varying standards for antitrust approval, and some require the acquiring party to conduct a competition review to ensure that the M&A does not lead to market imbalance.

Investor Protection Laws: To protect the rights of investors, cross-border M&As must comply with the investor protection

laws of the target company's jurisdiction. For example, some countries have strict announcement and disclosure requirements for public market M&A transactions, requiring the acquiring party to provide comprehensive information.

Labor Law: Cross-border M&As often involve employee relocation, transfer of labor agreements, and other issues. The acquiring party must comply with the labor laws of the target company's country to ensure that employees' rights are protected.

3.3 Compliance Challenges in Cross-Border M&A

3.3.1 Differences in National Regulatory Systems and Legal Conflicts

Cross-border M&As often involve different national regulatory systems, and the differences in laws and regulatory requirements between countries can pose significant compliance challenges. For example, in different countries, financial regulatory bodies may have vastly different requirements regarding the approval process for M&As, antitrust reviews, and capital flow regulations. This necessitates multi-party coordination during the M&A process to avoid conflicts that may result in the failure of the transaction.

3.3.2 Cross-Border Tax Policies and Tax Compliance Issues

Tax compliance issues in cross-border M&As have always been complex and challenging. Differences in tax policies and regulations between countries can influence the structure and financial arrangements of the M&A. For instance, some countries impose higher tax rates on foreign enterprises, while others offer tax incentives or exemptions, which can affect the overall cost of the transaction. Additionally, issues such as transfer pricing, tax planning, and double taxation in cross-border M&As require special attention and compliance management.

3.3.3 Data Privacy and Information Protection Compliance Issues

As data privacy and information protection regulations become increasingly stringent, data compliance issues in cross-border M&As have become more prominent. For example, the European Union's General Data Protection Regulation (GDPR) imposes strict personal data protection requirements. When cross-border M&As involve data transfer and storage, parties must comply with relevant laws and regulations. The acquiring party needs to carefully review the target company's data processing practices to ensure that all data storage and processing activities meet legal requirements, in order to avoid data breaches or violations of privacy protection regulations.

4.The Interaction Between Financial Risks and Compliance Risks in Cross-Border Mergers and Acquisitions

4.1 The Interaction Between Financial Risks and Compliance Risks

In cross-border mergers and acquisitions (M&A), financial risks and compliance risks are often intertwined and mutually influential. The interaction between these risks plays a crucial role in the success or failure of the transaction. Therefore, understanding their interplay is fundamental to developing effective risk management strategies.

4.1.1 The Regulatory Role of the Legal Environment on Financial Risks

Financial risks in cross-border M&A often involve exchange rate fluctuations, liquidity issues, credit risks, and others. The regulatory environment in different countries plays a significant role in governing these financial risks. Regulations often manage financial risks during M&A through policies on capital flows, foreign exchange management, and capital controls. For example, some countries have strict foreign investment regulations that require government approval for cross-border capital movements, limiting the free flow of capital and affecting fund allocation and financing methods. This could increase financial risks for the acquirer. Furthermore, tax policies, market access conditions, and other regulations may also affect financial risks in M&A transactions, especially in terms of capital structure design and financing arrangements.

4.1.2 Financial Market Turmoil and Conflicts in Compliance Execution

Financial market turmoil may lead to difficulties in compliance execution during M&A transactions. Instability in the financial markets can result in higher financing costs, market valuation fluctuations, and other risks, thus impacting the feasibility and profitability of the transaction. Against this backdrop, compliance requirements in cross-border M&A, such as antitrust reviews and foreign investment approvals, may face new challenges. For example, during financial crises, the volatility of the capital markets may lead to higher uncertainty in financing and valuation processes, increasing the complexity

of compliance reviews. Regulatory authorities may become more cautious, delaying or even rejecting the approval of the transaction. Therefore, fluctuations in the financial markets not only directly impact financial risks but also indirectly affect the management of compliance risks.

4.1.3 Coexistence of Financial Risks and Compliance Risks in M&A Transaction Structures

In the structure of cross-border M&A transactions, financial risks and compliance risks often coexist. For instance, some M&A transactions may involve complex financing structures, such as debt financing, equity financing, or hybrid financing methods, which introduce potential financial risks while also triggering compliance risks. If the M&A transaction structure does not comply with legal requirements, it could violate antitrust laws, foreign investment approval regulations, and other compliance requirements. In such cases, companies must balance financing methods with compliance requirements, ensuring that financial risks are controlled while adhering to relevant laws and regulations.

4.2 Best Practices in Risk Management and Compliance Control

Financial risks and compliance risks in cross-border M&A need to be coordinated and managed effectively through risk management and compliance control measures. The following are some best practice recommendations to help companies effectively address these dual risks in cross-border M&A.

4.2.1 M&A Strategies Considering Both Financial Risks and Compliance Risks

When undertaking a cross-border M&A, companies should consider the interaction between financial risks and compliance risks from a strategic level. In the early stages of the acquisition, companies should conduct a comprehensive evaluation of the target company's financial status and legal environment, identifying potential financial and compliance risks. The acquirer should adjust the M&A strategy based on these risk factors, such as selecting appropriate financing methods, planning a reasonable acquisition structure, developing a sound capital flow plan, and ensuring that the transaction complies with the legal regulations of the target market.

4.2.2 How Companies Can Establish a Risk Management Framework to Address Compliance Challenges

To address compliance challenges, companies should establish a robust risk management framework. This framework should include comprehensive control over compliance checks, legal reviews, and tax planning during the M&A process. Companies need to form cross-departmental compliance teams that work closely with legal, financial, and tax advisors to ensure that all legal and compliance requirements during the M&A transaction are met. In addition, companies should conduct regular compliance audits to identify potential compliance issues and adjust transaction structures and risk management strategies accordingly.

4.2.3 Compliance Review and Financial Auditing Mechanisms in Cross-Border M&A

Companies should establish compliance review and financial auditing mechanisms in cross-border M&A to ensure that financial and compliance risks in all transaction processes are identified and addressed in a timely manner. Specifically, the compliance review should cover antitrust laws, foreign investment approvals, tax compliance, and other areas to ensure the M&A transaction complies with relevant laws and regulations. Financial audits should ensure that the raising, movement, and capital structure of the transaction funds are reasonable, avoiding issues with capital allocation and liquidity risks. The auditing mechanism should be implemented throughout the entire M&A process, starting from the due diligence phase, continuing through the transaction, and extending into post-acquisition auditing to ensure the effectiveness of risk management.

5. Risk Management and Compliance Strategies in Cross-Border Mergers and Acquisitions

5.1 Financial Risk Management Strategies

In cross-border mergers and acquisitions (M&A), financial risk management is crucial for ensuring that the transaction proceeds smoothly and delivers expected value. Effective financial risk management helps companies mitigate potential losses caused by factors such as exchange rate fluctuations, credit risk, and market uncertainty.

5.1.1 Exchange Rate Risk Management Tools: Hedging Strategies and Financial Derivatives

Exchange rate risk in cross-border M&A typically arises from currency fluctuations, especially during the process of raising

funds and making payments. If the acquirer needs to make payments in foreign currencies, fluctuations in exchange rates could increase transaction costs. To reduce the impact of exchange rate fluctuations, companies can use financial derivatives (such as foreign exchange futures, options, and swaps) for hedging.

Hedging strategies help lock in future exchange rates for payments or receipts, allowing companies to avoid losses due to exchange rate fluctuations when the M&A transaction is completed. For instance, companies can use foreign exchange forward contracts to fix the exchange rate at a specific point in time, ensuring financial stability. In this way, the capital flow in cross-border M&A can avoid uncertain exchange rate risks, safeguarding the financial stability of the transaction.

5.1.2 Credit Risk Management: Credit Evaluation and Transaction Structure Design

Credit risk refers to the risk of loss due to the failure of a financing party or counterparty to meet its contractual obligations on time in a cross-border M&A transaction. Credit risk is particularly prominent when the deal involves debt or equity financing. To mitigate credit risk, companies can conduct a comprehensive evaluation of their counterparties' credit. This includes investigating and analyzing the target company's financial health, credit history, and past transaction behavior. By involving third-party credit rating agencies, companies can gain a clearer understanding of the counterparty's ability to repay and its creditworthiness, ensuring effective management of the acquirer's risk.

Additionally, in the design of the transaction structure, companies can introduce appropriate guarantees (such as collateral, security deposits, and financing guarantees) to distribute credit risk. For example, adjusting payment terms and only paying acquisition funds after meeting certain financial milestones can help reduce credit risk.

5.1.3 Market Risk Management: Financial Forecasting and Risk Identification

Market risk primarily arises from fluctuations in external market conditions, such as interest rate changes and demand volatility. Managing market risk in cross-border M&A is especially complex because market factors are influenced by global economic changes and political developments.

To effectively manage market risk, companies can establish precise financial forecasting models for risk identification. These models should include predictions about future market trends, competitive dynamics, product demand, and price changes. Furthermore, companies should conduct regular risk assessments and adjust their M&A strategies based on changes in the external market environment. For example, during the early stages of an acquisition, companies should assess the target market's size and growth potential to ensure that the acquisition will continue to create value in the future market environment.

5.2 Compliance Control Strategies

Compliance risk is an unavoidable factor in cross-border M&A, especially because of the differences in legal systems and regulatory requirements across countries. Companies need to develop comprehensive compliance control strategies to ensure the legality and feasibility of the M&A transaction.

5.2.1 Compliance Review Process: Legal Review and Compliance Checks in M&A

The legal review and compliance checks in an M&A transaction are key steps to ensure that the deal complies with laws and regulations, avoiding legal risks^[1]. First, companies must conduct due diligence on the target company to assess its legal status and identify potential compliance issues. The review should cover various legal documents, including the target company's registration status, tax situation, property rights, and employee benefits.

Additionally, companies should investigate the target company's relationship with government departments and regulatory bodies to ensure there are no significant legal disputes or compliance issues. For cross-border M&A, especially when it involves foreign investment, antitrust review, or sensitive areas like environmental protection, legal reviews are more complex. Companies should develop detailed legal review processes tailored to the industry characteristics and market environment of the target company, ensuring comprehensive evaluation and control of legal risks.

5.2.2 Building a Legal Compliance System: Cross-Border Legal Teams and Compliance Reporting Mechanism

With the globalization of M&A transactions, compliance control in cross-border M&A requires the support of cross-national legal teams. Companies should establish a diversified compliance management system and form cross-national legal teams consisting of experts from various countries and regions to ensure that all M&A transactions comply with local laws and

regulations.

Additionally, companies should establish an internal compliance reporting mechanism to continuously monitor legal compliance issues during the M&A process. Regular compliance reports can help companies detect potential legal risks and take preventive measures. Companies need to ensure efficient communication and collaboration between their cross-national legal teams to ensure full compliance with all requirements during each stage of the M&A process.

5.2.3 Adhering to International Compliance Standards: ISO, OECD Standards, etc.

Compliance control in cross-border M&A should not only consider national legal requirements but also adhere to international compliance standards. For example, ISO standards provide a set of international management system standards that can help companies standardize their compliance management processes. The OECD (Organisation for Economic Co-operation and Development) has also formulated a series of international guidelines on anti-corruption, cross-border M&A, and tax compliance, among others. Companies involved in cross-border M&A should consider these international compliance standards to ensure the legality of the transaction process.

Adhering to international compliance standards not only helps mitigate compliance risks but also enhances a company's reputation and competitiveness in the global market. Particularly in areas such as cross-border tax compliance, antitrust approval, and data privacy protection, companies should fully understand and comply with international compliance standards to ensure the smooth progress of the M&A transaction.

6. Conclusion and Future Outlook

6.1 Main Research Conclusions

6.1.1 Key Characteristics and Challenges of Financial and Compliance Risks in Cross-Border M&As

The financial and compliance risks in cross-border M&As are highly complex and are primarily reflected in the following aspects:

Diversity and Uncertainty of Financial Risks: Cross-border M&As involve various financial risks, including exchange rate risks, credit risks, and market risks. Especially in the context of global economic fluctuations and financial market instability, these financial risks become more unpredictable, requiring companies to adopt diversified risk management strategies.

Transnational and Complex Compliance Challenges: Due to differences in legal systems across countries, cross-border M&As often face issues such as legal conflicts, tax compliance, and antitrust approvals[2]. The varying regulatory systems, legal standards, and industry requirements across different regions pose significant challenges to the smooth execution of cross-border M&As.

Interaction Between Financial and Compliance Risks: In cross-border M&As, financial and compliance risks often occur simultaneously and influence each other. Market and credit risks may lead to delays or failures in the compliance review process, and conversely, compliance conflicts can affect the financing structure and financial arrangement of the M&A transaction^[3].

6.1.2 Innovations in Risk Management and Compliance Control Theories and Practices

Innovation in Risk Management Theory: Through a systematic analysis of financial risks and specific case studies of cross-border M&As, a series of new risk management strategies are proposed, including integrated hedging tools based on market, credit, and exchange rate risks, helping companies better control financial risks during the M&A process.

Innovation in Compliance Control Practices: This research introduces a cross-national legal team collaboration mechanism and a compliance review process for cross-border M&As, providing a clearer and more practical compliance management framework^[4]. Furthermore, the integration of international compliance standards (such as ISO and OECD guidelines) helps companies better navigate the complexity of legal environments and compliance challenges.

6.2 Research Contributions and Innovations

6.2.1 Unique Perspective and Contributions of This Study on Financial and Compliance Risks in Cross-Border M&As

This study provides the following unique perspectives and contributions through a comprehensive analysis of financial and compliance risks in cross-border M&As:

Interaction Theory of Financial and Compliance Risks: The study introduces the theory of the interaction between financial risks and compliance risks, analyzing the mutual influence and synergy between these two types of risks in cross-border M&As^[5].

Cross-National Legal Team Collaboration Mechanism and Risk Management Strategies: The study introduces the “Parallel Management Model for Financial and Compliance Risks,” providing a new theoretical framework and practical guidance for managing risks in cross-border M&As.

6.2.2 Supplementing and Improving Existing Literature and Theories

This study not only enriches the theoretical framework of risk management in cross-border M&As but also fills a gap in the research on compliance issues in cross-border M&As^[6]. By comprehensively analyzing the legal differences between countries, market risks, credit risks, and other related issues, this study provides practical and feasible compliance control frameworks and financial risk management solutions for cross-border M&A practices. Additionally, the research discusses issues such as financial market volatility, international investment flows, and legal compliance barriers in cross-border M&As, thereby supplementing the gaps in existing literature.

6.3 Future Research Directions

6.3.1 In-depth Research on the Risk Management and Compliance Challenges in Emerging Markets

With the rise of emerging markets, an increasing number of cross-border M&As are involving these countries^[7]. However, the financial systems, legal environments, and market stability in emerging markets still carry uncertainties. Future research can focus on the risk management and compliance challenges in these markets, exploring the unique risks encountered when conducting cross-border M&As in these regions and proposing corresponding risk management strategies.

6.3.2 Broader Legal Frameworks and the Impact of Global M&A Environments on Cross-Border M&As

With the acceleration of globalization, the legal environments for cross-border M&As have become increasingly complex. Future research could explore broader international legal frameworks, such as global trade laws, international investment laws, anti-money laundering regulations, and their impact on cross-border M&As^[8]. Particularly in the context of more stringent global regulations, how to address the legal risks encountered in cross-border M&As becomes an important research topic.

6.3.3 The Impact of Technological Advancements on Compliance and Financial Risk Management in Cross-Border M&As

With the rapid development of technology, particularly emerging technologies like artificial intelligence, big data, and blockchain, cross-border M&A compliance and financial risk management face new challenges. Technological advancements may bring innovative tools for risk management, such as blockchain-based cross-border payment systems and smart contracts. However, the application of technology may also introduce new compliance issues, such as data privacy concerns and cybersecurity risks. Therefore, future research should focus on exploring how technological advancements impact risk management and compliance control in cross-border M&As, and propose corresponding countermeasures^[9].

Funding

no

Conflict of Interests

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

Reference

- [1] Healy, P. M., & Ruback, R. S. (1992). Do mergers improve corporate performance. *Journal of Financial Economics*, 31, 135–176.
- [2] Harford, J. (2005). What drives merger waves. *Journal of Financial Economics*, 77(3), 529–560.
- [3] Cho, S., Sherry, S. U., Inajima, T., & others. (2022, March). Russian oil is increasingly becoming untouchable for traders (3). *Environment & Energy Report*.
- [4] Martynova, M., & Renneboog, L. (2006). The performance of the European market for corporate control: Evidence from

the 5th takeover wave. Other Publications TiSEM, (2).

- [5] Brown, T. J., Dacin, P. A., & Michael, G. P., et al. (2006). Identity, intended image, construed image, and reputation: An interdisciplinary framework and suggested terminology. *Journal of the Academy of Marketing Science*, 34(2), 99–106.
- [6] Tian, X., Sono, Y., Nishihara, M., & others. (2008). 2-A-6 Financial synergy in M&A. *Proceedings of the Operations Research Society of Japan Fall Meeting Abstracts*.
- [7] Bernile, G., Cumming, D., & Lyandres, E. (2007). The size of venture capital and private equity fund portfolios. *Journal of Corporate Finance*, 13(4), 564–590.
- [8] Evidence from Saudi Arabia. (2014). *Aestimatio: The IEB International Journal of Finance*, 9(2014), 182–199.
- [9] Rao-Nicholson, R., Salaber, J., Gao, T. H., & Long, T. (2016). Long-term performance of mergers and acquisitions in ASEAN countries. *Research in International Business & Finance*, 36.

Sentiment Analysis of IMDB Movie Reviews Based on LSTM

Jiahao Lu*, Hongji Fan, Yali Zhang

Software School, Henan Normal University, XinXiang, 453000, China

*Corresponding author: Jiahao Lu, 3329362823@qq.com

Copyright: 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: This study focuses on sentiment analysis of IMDB movie review data. Firstly, by leveraging the CBOW model and dynamically adjusting the context window, the distributed semantic features of words in movie reviews are deeply explored to construct precise word vector representations. To comprehensively evaluate the effectiveness of the model, a series of comparative experiments are designed. These experiments include not only traditional machine learning algorithms such as Naive Bayes, Decision Tree, and Random Forest but also a basic RNN model. The experimental results show that the LSTM combined with the Word2vec model achieves an accuracy of 0.9265, a recall of 0.9471, and an F1-score of 0.9353, demonstrating high precision, strong generalization ability, and good applicability.

Keywords: IMDB Dataset; Sentiment Analysis; Word2vec Model; CBOW Algorithm; LSTM Model

Published: Jun 6, 2025

DOI: <https://doi.org/10.62177/jaet.v2i2.429>

1. Introduction

The global film industry is continuously growing under the joint influence of policy promotion and market demand. As a flagship platform for movie information exchange, IMDB has become a crucial place for movie enthusiasts to obtain information and exchange opinions. This platform aggregates over fifty thousand user reviews daily, with diverse forms ranging from brief star ratings to detailed long reviews. These reviews not only reflect the audience's immediate feelings during movie-watching but also construct a rich semantic system of movie evaluations through distinctive language expressions. This system is an indispensable valuable resource for the field of natural language processing, especially for sentiment analysis tasks. Currently, with the rapid development of streaming services and online movie review communities, the in-depth mining of the emotional features of comments on IMDB is of great value to multiple parties. Movie creators can optimize their work directions based on this, distributors can accurately formulate marketing strategies, and audiences can make more informed movie-watching choices.

Text sentiment analysis is a key method in the study of movie reviews, relying heavily on the deep application of natural language processing technology. This method is dedicated to the meticulous feature analysis of text information with emotional connotations, extracting the core semantics of the text through complex algorithm models, and thus achieving in-depth mining of information. Wang Renwu^[10] improved the accuracy of sentiment analysis by constructing a sentiment dictionary using Word2vec compared to traditional methods.

In the process of sentiment analysis, NLP technology not only parses the literal meaning of text but also perceives the underlying emotional tendencies, providing a comprehensive and in-depth understanding and analysis of review texts. As an important branch of natural language processing, sentiment analysis technology can be broadly divided into two technical

schools: traditional machine learning methods and deep learning methods. These two approaches exhibit distinctly different strategies in feature engineering and model construction. Traditional machine learning methods in sentiment analysis focus more on the meticulous construction of feature engineering.

Based on statistical sentiment analysis methods, the two main technical paradigms can be systematically divided into traditional machine learning methods and deep learning methods, which form a sharp contrast in feature engineering and model construction. Traditional methods emphasize the optimization of feature engineering. For example, Wang Mingyang et al.^[3] found that expanding the sentiment dictionary and introducing the Word2vec model both help improve the effect of the k-nearest neighbor (K-Nearest Neighbor, KNN) method on multi-emotion classification of Weibo text; Wang Haiyan^[12] et al. proposed a text classification algorithm based on a rough set-based bag-of-words model combined with support vector machines.

Deep learning^[13] methods have gradually become mainstream due to their end-to-end learning advantages, effectively avoiding manual intervention by automatically generating text representation features. Zhang Qian^[11] et al. applied the Word2vec model to transform text data into real-number vectors; Reference^[5] used the Word2Vec tool to train a corpus, obtained text word vector representations, and then used an LSTM model with an attention mechanism for text feature extraction, combined with cross-entropy training, and applied the model to the classification of tourism issue texts; Reference^[7] proposed a new short text classification method by combining context-dependent features with a multi-stage attention model based on time convolutional networks and CNNs; The main features of convolutional neural networks^[1] lie in weight sharing and local connections; Reference^[7] proposed a new short text classification method by combining context-dependent features with a multi-stage attention model based on time convolutional networks and CNNs; Reference^[2] proposed a dual graph convolutional network model for aspect-based sentiment analysis, which considers both the complementarity of syntactic structure and semantic relevance; Liu Hui^[4] et al. proposed an aspect-level sentiment analysis model that integrates matching long short-term memory networks and grammatical distance; Liu Jun^[8] et al. created a long text analysis model using CNN and bidirectional long short-term memory networks (Bi-directional Long Short-Term Memory, Bi-LSTM); Ding Feng et al.^[6] proposed a BiLSTM-CRF model and applied it to aspect-based sentiment analysis tasks. The bidirectional long short-term memory network can capture long-distance bidirectional semantic dependencies and learn text semantic information, predicting the global optimal label sequence at the sentence level.

Through scientific sentiment analysis methods, in-depth mining of off-site data in the movie industry has value in three dimensions: First, it can provide precise assistance for audience group movie-watching decisions and, through visual analysis results, perceive public feedback on movies, promoting multi-dimensional emotional resonance among movie enthusiasts. Secondly, film producers can gain data support for precise market positioning through insights into emotional data, accurately grasping the aesthetic preferences and emotional needs of the audience. More importantly, this analysis mechanism not only provides an in-depth analysis of contemporary public aesthetic value orientations but also has important strategic reference value for the market potential mining of documentary and other sub-genres, as well as for innovating film content production and optimizing the industrial ecosystem.

This study aims to develop an LSTM-based sentiment classification model using the IMDB movie review dataset. Under this framework, the CBOW algorithm is used to transform movie review text into high-dimensional distributed word vectors through pre-training, thereby constructing a semantic information-rich word representation space. Subsequently, the LSTM network is introduced, leveraging its unique temporal memory ability to deeply extract emotional features from the text, continuously optimizing network parameters to enhance model performance. Finally, a horizontal comparison strategy is adopted, selecting classic models such as SVM and CNN as references, and using precision, recall, and F1-score as three key indicators to comprehensively evaluate the effectiveness of the LSTM model. In addition, detailed visual analysis of the experimental data is conducted to reveal the model's performance and characteristics from multiple dimensions.

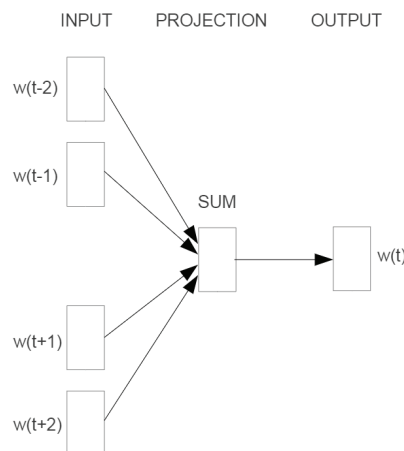
2. Algorithm Introduction

2.1 Word2vec Model

In the field of natural language processing, word vector representation has always been a challenge. However, the Word2vec

framework has successfully broken through the limitations of traditional One-Hot encoding through its innovative distributed representation method. Word2vec includes two classic algorithms—CBOW and Skip-gram models—which adopt different contextual learning strategies. The CBOW algorithm uses a predefined context window to aggregate semantic information from surrounding words to predict the features of the target word; in contrast, the Skip-gram model starts from the central word and infers the semantic representations of its surrounding words in reverse. Although these two training methods complement each other in achieving their goals, both can construct dense vector spaces rich in semantic associations. These distributed word vectors have demonstrated extremely high engineering application value in practical applications such as text classification and semantic similarity calculation.

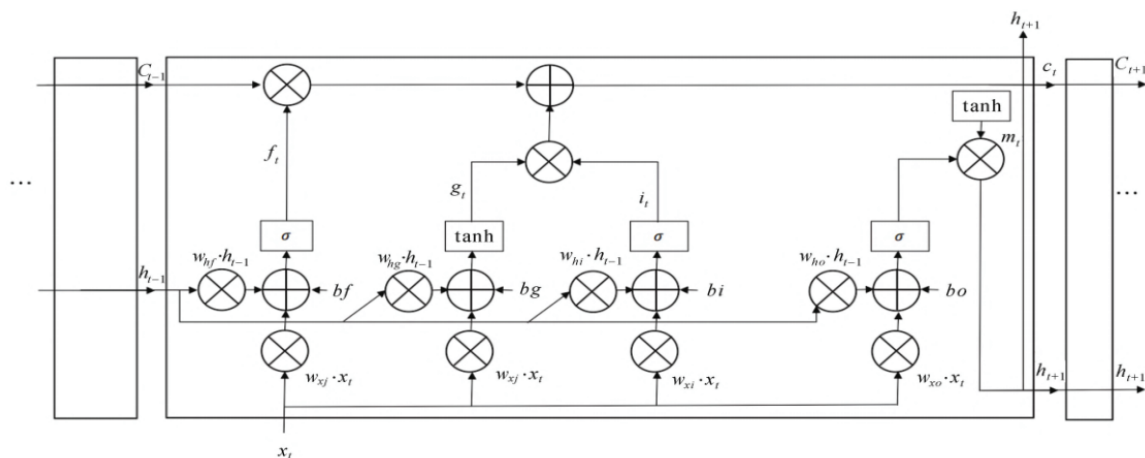
Figure 1 The structure of CBOW



When discussing the architecture implementation of the CBOW model, we note that its parameter passing mechanism follows a carefully designed hierarchical computation strategy. The model starts from the input layer, mapping each central word's contextual information into a specific embedding space. During this process, each context word is converted into a vector representation. Then, the hidden layer aggregates these input vectors, generating an intermediate semantic representation through weighted summation, which fuses contextual information. The output layer adopts an optimized binary tree topology, where leaf nodes correspond one-to-one with words in the corpus, and non-leaf nodes form feature combination paths. This design helps improve computational efficiency. In model input, one-hot encoded vectors are used as a starting point, transformed through a weight matrix Y to generate N -dimensional hidden layer vectors. Subsequently, this vector is reconstructed through another weight matrix W , returning to M -dimensional output space, and finally normalized through Softmax to obtain the predicted distribution of the target word. By setting an appropriate window size C , the CBOW model can effectively control the context range, achieving semantic conversion from discrete symbols to continuous vector spaces. This process is clear and efficient.

2.2 LSTM Model

Figure 2 LSTM specific flow chart



To achieve long-term dependency goals, we designed a selective memory cell structure and introduced a forget gate mechanism to filter and retain key information from the input sequence. The computation formulas precisely define this process.

$$f_t = \sigma(f^{(t)}) = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f), (1)$$

The input gate is responsible for integrating information from the previous and current moments, with the specific computation formula provided, precisely controlling the inflow of information.

$$g_t = \tanh(g^{(t)}) = \tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g), (2)$$

Subsequently, i_t is computed to prepare for the final cell state C_t , with the formula as follows:

$$i_t = \sigma(i^{(t)}) = W_{xi}x_t + W_{hi}h_{t-1} + b_i, (3)$$

The output gate O_t will further filter the results of the input gate, retaining useful information. Specifically, x_t and h_{t-1} are multiplied by weight matrices in the output gate, summed, and a bias b_o is added. The computation formula is as follows:

$$O_t = \sigma(O^{(t)}) = W_{xo}x_t + W_{ho}h_{t-1} + b_o, (4)$$

At time t , the system receives three inputs: the current value x_t , the previous state h_{t-1} , and the memory cell state C_{t-1} . These three inputs jointly affect the forget gate, input gate, and output gate, participating in their computations involving x_t and h_{t-1} , thereby influencing the overall state output of the system.

In the forget gate, these two vectors are multiplied by a weight matrix, summed, and a bias term b_f is added. Finally, a sigmoid activation function is applied to obtain the result of the forget gate f_t . In the update gate, x_t and h_{t-1} are again multiplied by preset weight matrices, summed, and a bias term b_g is added, followed by processing through a tanh activation function layer. Subsequently, this result is added to the output of the forget gate through matrix addition to obtain the intermediate state g_t of the update gate.

Next, the input gate multiplies the input and the previous state x_t and h_{t-1} by preset weight matrices, sums them, and applies a sigmoid function activation to obtain the result i_t . Finally, the update gate and the forget gate outputs are combined to form the new cell state C_t , with the formula as follows:

$$C_t = C_{t-1} \odot f_t + g_t \odot i_t, (5)$$

where \odot denotes element-wise multiplication.

To further filter out non-critical information, the data C_t is processed through a tanh layer to obtain a refined matrix m_t :

$$m_t = \tanh(C_t), (6)$$

Then, the output gate result O_t is element-wise multiplied with m_t to obtain the final state vector h_t :

$$h_t = O_t \cdot m_t, (7)$$

Finally, the state vector h_t is multiplied by a weight matrix W_{yh} and a bias term b_y is added to obtain the prediction result y_t at time t :

$$y_t = W_{yh}h_t + b_y, (8)$$

Thus, the forward propagation process within the memory cell is completed.

Figure2: Comparison of Michael Jordan and LeBron James' shooting percentages in the regular season

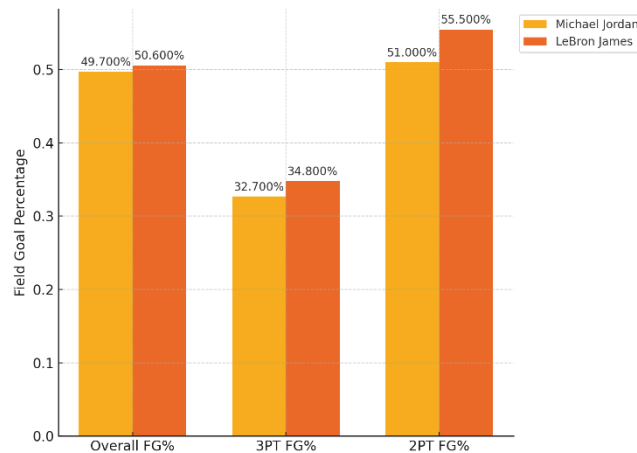


Figure2 is a comparison of the shooting percentages of Michael Jordan and LeBron James in the regular season (up to the 2023 season), including overall shooting percentage, three-point shooting percentage, and two-point shooting percentage.

3.Data Acquisition and Preprocessing

3.1 Experimental Data Acquisition

This study uses the internationally recognized IMDB benchmark dataset as the experimental object, obtained from the official website of Stanford University and following standard data collection protocols. It contains two independent modules, stored in a separate Excel file, namely the training set and the test set. The training set file contains 25,000 labeled samples (12,500 positive and negative reviews each), and the test set documents construct an equally proportioned balanced data distribution, forming a total of 50,000 high-quality labeled materials.

The dataset adopts a two-pole classification storage mechanism, with each Excel file internally divided into "POS" and "NEG" logical units, storing reviews with positive emotional labels (Positive) and negative emotional identifiers (Negative), respectively. Each review contains structured fields such as a unique identifier, text number, star rating, and original review text. The text content is cleaned through standardized processing to effectively eliminate HTML tag interference using special characters.

3.2 Data Preprocessing

3.2.1 Discrete Variable Labeling

For the binary emotional labeling system targeting the IMDB dataset, a numerical encoding strategy is adopted, decentralizing the category. In the original data, the "POS" folder storing positive evaluations (Positive) is uniformly represented by the numerical label "1"; the "NEG" folder storing negative evaluations (Negative) corresponds to the numerical label "0".

Table 1 Discrete Variable Labeling

	Description	Numerical Label
POS	Positive Evaluation (Positive)	1
NEG	Negative Evaluation (Negative)	0

3.2.2 Data Tokenization

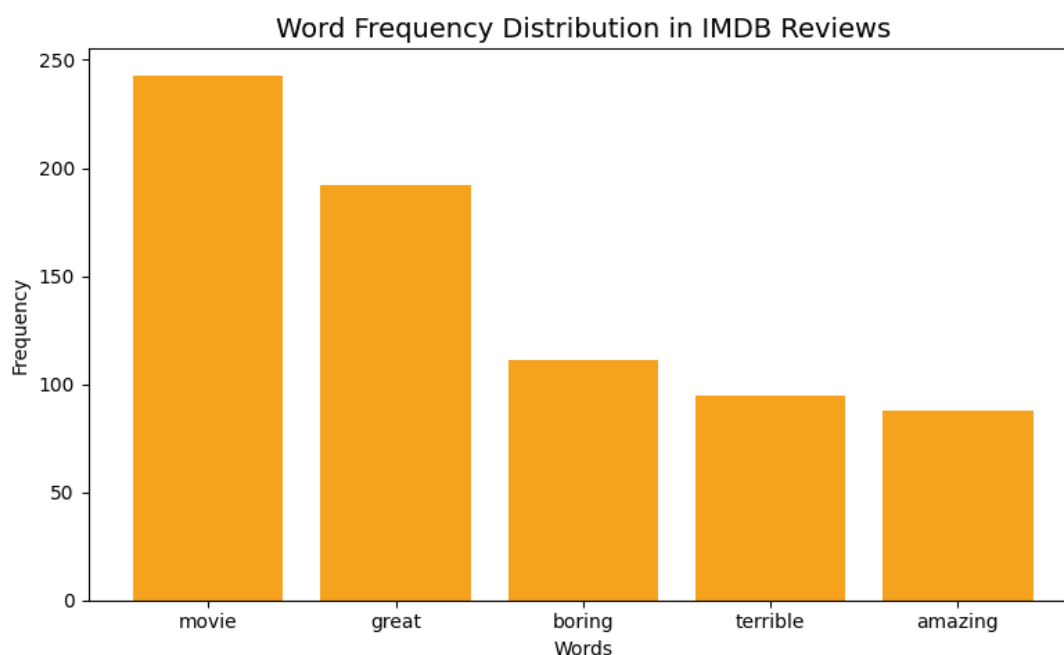
This study identifies high-frequency semantic units by constructing a vocabulary distribution histogram, adopting a feature engineering optimization strategy based on word frequency statistics. By calculating the probability distribution of term occurrences, key word indicators are screened during the corpus parsing stage, resulting in a distinct differentiation. This statistic-based feature screening mechanism not only generates a high-density information feature vector space but also effectively enhances the text representation's ability to capture contextual semantics, thereby establishing a vocabulary association model with strong explanatory power and providing an input matrix for subsequent supervised learning models after denoising.

Table 2 Word Frequency Distribution

Word	Frequency
movie	243
great	192
boring	111
terrible	95
amazing	88

Based on the quantitative data in Table 2, a visualization chart is constructed using word frequency statistical analysis to systematically present the distribution patterns of high-frequency vocabulary.

Figure 3 Word frequency statistics



4. Word2vec Model for Constructing Word Vector Space

This study aims to utilize the Continuous Bag-of-Words (CBOW) model to construct a distributed semantic space for text. During the construction process, the dimension parameters of the word vector space are first defined, determined by the vocabulary size a and the preset semantic embedding dimension K . In this space, each vocabulary is mapped to a row vector, and the geometric layout of these vectors reflects the semantic relevance between words. For the review text corpus M , the study converts it into a sequence form through context window sliding. The total number of words in each review is denoted by T , and the position of words in the vocabulary is represented by an index value $\in [1, a]$. To accurately capture word relationships, this conversion introduces dynamic window sampling techniques, gradually adjusting the geometric structure of the word vector space based on the relevance between central words and context words. This modeling process includes multiple steps, aiming to deeply explore text semantic information through the CBOW model, thereby enhancing the accuracy and efficiency of text processing and analysis.

4.1 Data Transformation Expression

A global corpus encoding strategy is adopted for the study. First, the preprocessed review text is integrated into a comprehensive text representation matrix through the CBOW algorithm, enhancing data generalization ability. Subsequently, a stratified random subdivision strategy is employed, dividing the integrated feature matrix into training and test subsets in a 1:1 ratio, ensuring consistency in dimensions such as emotional label distribution and text length in statistics. This method not only retains the integrity of contextual semantics but also avoids model overfitting risks due to data leakage, eliminating numerical tensor generation caused by feature deviation between datasets.

4.2 Model Definition

In neural network processing, this study conducts experiments on the IMDB movie review dataset. To convert text vocabulary into a numerical form recognizable by computers, the efficient Word2Vec word embedding technique is utilized to achieve effective conversion of vocabulary into numbers for subsequent data processing and analysis.

Specifically, the Gensim library for natural language processing in Python is used, directly calling the Word2Vec model within it to train the corpus and generate word vectors. This approach maintains the relationships between words while effectively transforming each word into a low-dimensional continuous space.

The optimal configuration in the parameter space is determined using the grid search method during the model training process. By adjusting the word vector dimension parameters, the model complexity is reduced to accommodate scenarios with smaller training data scales.

Table 3 Parameters in the Word2vec Model

Parameter Name	Description	Setting
vector_size	Word Vector Dimension	300
window	Context Window Size	5
min_count	Minimum Occurrence Count	2
epochs	Number of Training Epochs	10

5.Sentiment Analysis Based on LSTM Model

5.1 Data Format Conversion

5.1.1 Calling Word Vector Space Results

Here, the vector matrix dictionary obtained from the Word2vec model is directly used.

Table 4 Index List and Corresponding Review Text Content

Index	Word
0	movie
1	great
2	boring
3	terrible
4	amazing

5.1.2 Creating Tuples

1) Data Structure Initialization

To create a semantic mapping interface, a system based on the pre-trained word embedding model is constructed, loading the processed review text and corresponding emotional labels. Simultaneously, the vocabulary features generated by Word2vec are integrated, calculating the total number of samples to set appropriate batch training parameters, providing an accurate data scale reference for loss evaluation, ensuring efficient and accurate training processes.

2) Semantic Mapping and Missing Value Handling

A word-by-word search is conducted on the review text. If a word exists in the embedding matrix, its corresponding vector is extracted.

To handle out-of-vocabulary (OOV) words, a zero initialization strategy is adopted, generating zero vectors matching the embedding dimensions. This ensures that OOV words are correctly processed during model training, enhancing the model's generalization ability.

Using serialized word vectors, they are concatenated into a sequential feature matrix, i.e., sample size multiplied by sequence length multiplied by embedding dimension. This method better represents the temporal information in the text, providing more possibilities for subsequent processing.

3) Dimensional Standardization Processing

The maximum sequence length of the training set is calculated as the padding baseline. Short texts are padded with zeros at the end, while long texts are truncated from the tail. Emotion labels are converted into 64-bit integer tensors to adapt to the PyTorch framework. The final generated tuple data structure (feature matrix, label tensor) format meets the dimensionality requirements of deep learning models for input data. Meanwhile, through length standardization processing, the training efficiency loss caused by differences in text length is eliminated.

5.2 Defining the LSTM Model Structure

5.2.1 Defining Model and Functional Layers

The core architecture of this study focuses on constructing an emotion classification model, divided into two key stages:

parameter initialization and hierarchical feature mapping. First, based on task requirements, the network's topological parameters are configured, including the setting of input, hidden layer, and output dimensions, as well as the selection of activation functions. The tanh activation function is adopted to model temporal dependencies, and gating mechanisms are introduced to dynamically regulate information flow. The final layer is a fully connected classification layer, which maps the hidden layer state to the probability space using the sigmoid function to obtain the final emotional polarity prediction values. The entire design is implemented using the PyTorch framework, achieving automatic gradient computation and thereby improving training efficiency. Through this structure, an efficient emotion classification model is successfully constructed, providing strong support for research and application in the field of sentiment analysis.

5.2.2 Defining the Forward Propagation Function

In the process of temporal feature modeling, its implementation relies on a structured information flow mechanism. At the initial state setting of the model, both the memory cell state vector and the hidden state tensor need to undergo a strict zero-value initialization. This step ensures consistency and a clear baseline condition at the starting point of temporal processing. Subsequently, the forward computation process begins. The input sequence undergoes a series of gating operations in the Long Short-Term Memory (LSTM) model, including the input gate, forget gate, and output gate, which perform multi-scale and hierarchical feature fusion. Through iterative updates and continuous corrections of the memory cell, the hidden contextual dependencies in the text sequence are deeply captured and represented. At the final moment of temporal feature extraction, after dimensionality reduction and nonlinear activation by the fully connected layer, the probability distribution of emotion classification is formed. The entire computation process is designed with a dynamic computational graph, achieving end-to-end gradient propagation from word sequence input to emotion judgment output. Additionally, gradient normalization strategies are employed to ensure the robustness and reliability of the training process, thereby establishing a complete mapping framework from input to judgment.

5.2.3 Setting Random Seeds

In this exploration, we address the issue of convergence uncertainty caused by the inherent randomness in neural network weight initialization. A hardware-independent random control mechanism is carefully designed and implemented. By leveraging the PyTorch deep learning framework, we precisely set its deterministic computation mode to ensure that, even on high-performance computing devices like GPUs, the weight matrices can be reproducibly generated during parameter initialization. This approach not only significantly influences and restricts the search path in the parameter space but also effectively prevents the training process from prematurely falling into local minima. It further ensures consistency in the gradient descent paths across different experimental batches, laying a solid mathematical foundation for the steady improvement of model performance in the extensive exploration of the global space.

5.2.4 Building Datasets and Iterators

This study constructs a data supply pipeline oriented towards deep learning frameworks. It first loads preprocessed word embeddings and sentiment-annotated vectors, instantiating the training set's feature matrix and data loader. Efficient memory management and data throughput optimization are achieved by configuring the number of samples selected in each training session, whether to shuffle the data, the number of processes for parallel loading, and selecting sample quantities that are multiples of the remaining samples. The data loader employs an on-demand generation mechanism, dynamically performing tensor conversion and batch combination during the training process. This ensures stable transmission of large text data within GPU memory constraints, providing a standardized data input interface for model iterations. The parameters of the training set iterator are set as shown in Table 5.

Table 5 Iterator Parameters

Parameter Name	Description	Setting Value
batch_size	Batch Size	32
shuffle	Whether to Shuffle Data	True
drop_last	Whether to Drop Incomplete Batches	True

5.2.5 Constructing the LSTM Model

Based on the training and inference of the LSTM network in a GPU-accelerated environment, hyperparameters such as hidden layer dimensionality and learning rate are jointly optimized through hyperparameter space traversal. A predefined network topology configuration scheme is adopted for the LSTM model, with parameters set as shown in Table 6.

Table 6 LSTM Model Parameters

Parameter Name	Description	Setting Value
input_size	Input Dimension	300
hidden_size	Number of Hidden Layer Neurons	128
output_size	Number of Output Classes	2

5.2.6 Establishing Optimizer and Loss Function

This study employs the Adaptive Moment Estimation (Adam) optimizer for parameter updates, training the optimizer's parameter set. The initial learning rate is set to 2×10^{-6} to control parameter updates, tailored to the characteristics of the emotion classification task. Additionally, the cross-entropy loss function is constructed as the supervisory signal. Its mathematical properties effectively measure the discrepancy between the true labels and the probability distribution. The training scheme incorporates a dynamic learning rate decay coefficient of 0.

5.3 Model Prediction

During the model training process, the number of iterations is empirically set to 32. As training progresses, model parameters are continuously adjusted and optimized, leading to a steady decline in training loss values and a gradual increase in the training set's accuracy, ultimately stabilizing. Detailed loss values and accuracy data are shown in Table 8.

5.4 Model Comparison

Table 7 Training Loss and Accuracy During Training

Iteration	Loss	Training Accuracy
1	0.5732	0.8794
5	0.5243	0.8952
10	0.4881	0.9108
15	0.472	0.9181
20	0.4635	0.9215
1	0.5732	0.8794

The computation processes for accuracy, recall, and F1-score are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \text{Recall} = \frac{TP}{TP + FN}, \text{F1} = \frac{2PR}{P + R}$$

TP (True Positive) is one of the important indicators for evaluating model performance. In machine learning algorithms, TP represents the model's ability to correctly predict positive instances. A high TP value indicates a high recognition accuracy of positive instances by the model, reflecting excellent model performance. Therefore, when evaluating a model's performance, attention should be paid to the TP indicator to ensure the accuracy and reliability of the model's predictions.

FP (False Positive) refers to the situation where the model incorrectly predicts negative instances as positive. Such errors may lead to misleading conclusions and decisions. To avoid FP, it is necessary to improve the model's accuracy and precision, reducing false predictions. During model construction and validation, special attention should be paid to the FP rate to ensure the model's stability and reliability.

FN (False Negative) is a common and impactful error type in data analysis, referring to the situation where the model

incorrectly predicts positive instances as negative. This may lead to severe consequences; therefore, during model establishment and optimization, it is essential to focus on and identify the causes of FN to improve the model's accuracy and reliability.

TN (True Negative) is one of the crucial indicators for evaluating classification model performance, representing the model's ability to correctly predict negative instances. The TN value reflects the model's accuracy and reliability in identifying negative instances. When evaluating classification models, the TN value is critical for assessing the model's performance and is vital for enhancing the model's effectiveness.

Table 8 Model Result Comparison

Model	Accuracy	Recall	F1-Score
Naive Bayes	0.8712	0.8931	0.8819
Decision Tree	0.8834	0.9075	0.8922
Random Forest	0.8991	0.9162	0.9027
RNN	0.9033	0.925	0.9135
LSTM	0.9265	0.9471	0.9353

Conclusion

Through data preprocessing, the IMDB dataset employs Word2vec to construct a word vector space, followed by training an LSTM model. Comparative experiments with other models show that the LSTM combined with the Word2vec model achieves an accuracy of 0.9265, a recall of 0.9471, and an F1-score of 0.9353. This represents an improvement of at least 0.0232 in accuracy, 0.054 in recall, and 0.0218 in F1-score compared to other models. The model demonstrates high precision, strong generalization ability, and good applicability.

Funding

no

Conflict of Interests

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

- [1] Liu, H. D., Sun, X. H., Li, Y. B., et al. (n.d.). A Review of Deep Learning Models for Image Classification Based on Convolutional Neural Networks. *Computer Engineering and Applications*, 1–29. Retrieved April 25, 2025, from <http://kns.cnki.net/kcms/detail/11.2127.TP.20250213.1223.013.html>
- [2] Ting, Z., Ying, S., Kang, C., et al. (2023). Hierarchical dual graph convolutional network for aspect-based sentiment analysis. *Knowledge-Based Systems*, 276.
- [3] Chai, Y. (2022). Research on Sentiment Analysis of Book Review Texts Based on LSTM and Word2vec. *Information Technology*, (07), 59–64+69.
- [4] Liu, H., Ma, X., Zhang, L. Y., et al. (2023). An aspect-level sentiment analysis model integrating matching long short-term memory network and syntactic distance. *Computer Applications*, 43(01), 45–50.
- [5] Wanli, L., & Lei, Z. (2022). Question Text Classification Method of Tourism Based on Deep Learning Model. *Wireless Communications and Mobile Computing*, 2022.
- [6] Ding, F., & Sun, X. (2022). Negative emotion opinion target extraction based on attention mechanism and BiLSTM-CRF. *Computer Science*, 49(02), 223–230.
- [7] Yingying, L., Peipei, L., & Xuegang, H. (2022). Combining context-relevant features with multi-stage attention network

for short text classification. *Computer Speech & Language*, 71.

- [8] Liu, J., Cao, J. X., Ding, W. N., et al. (2022). Research on reservoir porosity prediction method based on bidirectional long short-term memory neural network. *Progress in Geophysics*, 37(05), 1993–2000.
- [9] (2018). Text Categorization with Improved Deep Learning Methods. *Journal of Information and Communication Convergence Engineering*, 16(2), 106–113.
- [10] Wang, R. W., Song, J. Y., & Chen, C. B. (2017). Research on the application of sentiment analysis based on Word2vec in brand cognition. *Library and Information Service*, 61(22), 6–12.
- [11] Zhang, Q., Gao, Z. M., & Liu, J. Y. (2017). Research on microblog short text classification based on Word2vec. *Information Network Security*, (01), 57–62.
- [12] Wang, H. Y., Li, J. H., & Yang, F. L. (2014). A review of support vector machine theory and algorithms. *Journal of Computer Applications*, 31(05), 1281–1286.
- [13] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527–1554.
- [14] Li, J. H., Ahmed, A. S., Chai, Q., et al. (2021). Emotionally charged text classification with deep learning and sentiment semantics. *Neural Computing and Applications*, 34(3), 2341–2351.

Dear Researchers and Scholars :

Greetings from Asia Pacific Science Press, a beacon of academic and scientific publishing, located in the vibrant city of Hong Kong.

We extend our heartfelt gratitude for your relentless pursuit of knowledge, and your significant contributions to the advancement of science and society. It is researchers and scholars like you who propel humanity forward, and we at the Asia Pacific Science Press are devoted to ensuring that your groundbreaking works receive the global recognition they rightfully deserve.

In light of our commitment to disseminating pioneering research across various disciplines, such as medicine, architecture, education, and electronics, we are reaching out with two pivotal opportunities to augment our collaboration with the global academic community:

Call for Paper Submissions:

We cordially invite you to submit your original research articles to our fast-growing, peer-reviewed, and open-access journals. Our platform guarantees an extensive, global reach, enabling your work to garner maximum visibility and citation in the academic sphere. Rest assured, your work will be meticulously assessed by experts in the field, ensuring it receives the acknowledgment and exposure it merits.

Join Our Esteemed Team:

We are fervently searching for passionate researchers and scholars interested in joining our burgeoning team at Asia Pacific Science Press. We offer numerous roles, such as peer reviewers, editors, and advisory board members, where your expertise will significantly shape the content and quality of our publications. In return, you will gain invaluable experience, network with preeminent scholars, and play a pivotal role in molding the future of global academic publishing.

Why Choose Asia Pacific Science Press?

Global Reach: Your work will be accessible to a worldwide audience, free from any access barriers.

Collaboration with Renowned Universities: We have established extensive publishing systems in cooperation with world-renowned universities, such as Wuhan University, Hong Kong University, and the University of Malaya.

Diverse Disciplines: Your research will be housed among numerous journals across a multitude of academic projects and disciplines.

As we stride forward in the academic landscape, we envision a future where our collective efforts shape a more enlightened, innovative, and interconnected global society. We sincerely hope that you consider this invitation to join us on this auspicious journey towards knowledge, discovery, and global impact.

Should you wish to submit your work or express interest in joining our team, please do not hesitate to contact us. You can submit your manuscript or personal profile to info@apspublisher.com or visit our website at www.apspublisher.com for more information.

Thank you for considering this opportunity, and we eagerly anticipate the possibility of welcoming you to the Asia Pacific Science Press family. Together, let's forge a future of unparalleled scientific advancement and discovery.

Warm regards
Asia Pacific Science Press

OUR JOURNALS

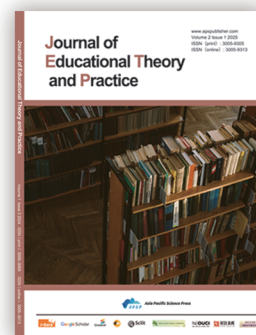
Asia Pacific Economic and Management Review is an international, peer-reviewed and open access journal which focuses on theoretical and applied studies of corporate and financial behavior. Aiming to promote the research in fields of business economics and management, it covers mainly but not limits to the following areas: accounting and financial management, economics, human resource management and organizational behavior, information management, international business, strategy and innovation, management science and operations management, marketing and retailing, finance.



Critical Humanistic Social Theory is an journal that publishes papers specifically using quantitative or qualitative research methods for social science research. The journal encourages scholars to conduct social science theory research from the perspective of social critical theory and emphasizes research concerned with issues or methods that cut across traditional disciplinary lines.



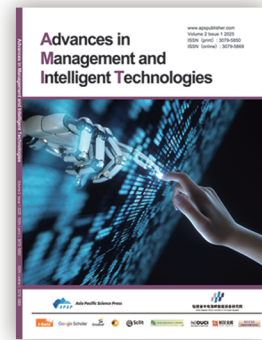
Journal of Educational Theory and Practice is an international, peer-reviewed and open access journal which is to promote the evaluative, integrative, theoretical and methodological research on contemporary education; shape a novel, broader view of issues in contemporary education; enhance the caliber of humanities research through active use of best domestic and foreign practices; and integrate the achievements of various sciences and knowledge areas with unconventional approaches.



Journal of Advances in Engineering and Technology is an international, peer-reviewed and open access journal which publishes original articles, reviews, short communications, case studies and letters in the field of electronic research and application.



Advances in Management and Intelligent Technologies is an international, peer-reviewed, open-access academic journal, hosted by the Fujian Strait Institute of Intelligent Equipment and managed and published by Asia-Pacific Science Press. It focuses on the latest research in the fields of management and intelligent technologies, and aims to advance both theoretical and applied research in management, technological innovation, and intelligent development.



Asia Pacific Journal of Clinical Medical Research is an international, peer-reviewed, open access journal dedicated to advancing clinical medical research across multiple disciplines. The journal serves as a platform for publishing high-quality original research, reviews, and clinical studies that enhance the understanding of medical practices, treatment innovations, and healthcare outcomes, thereby supporting patient care and medical advancements in the Asia Pacific region and beyond.



Asia Pacific Journal of Educational Research is an international, peer-reviewed, open-access academic journal focusing on educational theory and practice. It publishes high-quality research on educational reform, teaching methods, educational equity, and policy studies. The journal addresses practical needs and institutional changes in the education systems of the Asia-Pacific region, advocating a balance between theoretical inquiry and practical experience. It encourages original studies from multicultural, comparative, and interdisciplinary perspectives, aiming to support educational innovation and policy development across the region.



Asia Pacific Economic and Social Development is an international, peer-reviewed, open-access academic journal openly distributed to the global academic community. The journal is committed to publishing original research with theoretical depth and practical value in the fields of economic and social development. It focuses on issues such as economic behavior, social structure transformation, policy innovation, and regional coordinated development in the Asia-Pacific region. The journal encourages interdisciplinary perspectives and promotes the integration of economics, sociology, management, and related disciplines.

