ISSN 2616-5767 Vol.9, Issue 2: 55-62, DOI: 10.25236/AJETS.2026.090207

Carbon Footprint Manager: An AI-assisted Carbon Footprint Calculation System for Non-standard Products

Sean Gu[1], Chi Shing Lin[2]

[1]Beanstalk International Bilingual School Shunyi Campus, Beijing, China
[2]Shenzhen College of International Education, Shenzhen, China
mrsomeoneeeeee@gmail.com

Abstract: *This study addresses the low adoption of carbon footprint assessment for non-standard products (e.g., local foods, customized goods) due to the lack of standardized databases and accounting models. Against the backdrop of global low-carbon trends and corporate compliance needs, the research aims to overcome challenges such as high costs and poor scalability in traditional methods. An integrated intelligent system was developed, combining hardware (electronic scales, cameras) with AI technologies: Baidu OCR for text extraction, a SentenceTransformer-based vector database for querying carbon factors, and the Deepseek large language model for semantic reasoning and calculation. A Gradio-based interface enables end-to-end operation. Tests showed text recognition accuracy of 90.2% (net content 95.0%, ingredient list 92.5%). Carbon footprint estimates had an average absolute error of 4.7% (range: 1.2%–7.3%) against expert benchmarks. The system enables rapid, manual-modeling-free on-site assessment, generating standardized reports and tailored reduction strategies. This research fills a market gap by offering a universal estimation framework for non-standard products. It demonstrates the potential of combining OCR, vector databases, and LLMs for efficient carbon factor inference, while edge AI integration enhances practicality. Applicable to green procurement, supply chain management, and ESG reporting, the system empowers SMEs to conduct carbon accounting cost-effectively. It also supports the broader adoption of carbon labeling in food, retail, and manufacturing, driving green supply chain development and corporate low-carbon transition with significant social and commercial value.*

Keywords: *Non-standard products; Carbon footprint accounting; AI empowerment; OCR recognition; Large language model; Vector database*

1. Introduction

As your paper will be an important component in the journal, we highly recommend that all the authors follow this guideline to adjust the format of your paper so as to promise the highest reading experience.

1.1. Research Background

In the context of global climate change intensifying, carbon peak and carbon neutrality have become the core goals for all countries to promote green and low-carbon development. China has also clearly stated in the "14th Five-Year Plan" that it will strengthen carbon emission management and improve the carbon accounting and carbon trading system. As the main body of carbon emission reduction, enterprises urgently need to establish a product full life cycle carbon footprint assessment mechanism to meet policy compliance, green product certification, ESG evaluation, and consumers' demands for transparency of products' low-carbon attributes. Carbon footprint accounting has become an important foundation for enterprises to participate in the construction of green supply chains and market competition [1]. Currently, the mainstream carbon footprint assessment methods are all based on standardized production processes, uniform product formulas, and comprehensive carbon factor databases. This model demonstrates good applicability in the carbon accounting of standardized industrial products, but it has obvious limitations when dealing with non-standard products. In practical applications, local specialty foods and small-scale handicraft products have significant differences in ingredients, production processes, and raw material origins, and thus lack corresponding standard carbon factors; customized products and compound new products have frequent changes in components and raw material sources, making it difficult to directly apply a unified accounting model; at the same time, most small and medium-sized enterprises lack professional carbon emission assessment personnel and convenient accounting tools, resulting in problems such as long calculation cycles, high costs, and poor result accuracy in the carbon footprint accounting of non-standard products[2]. These pain points directly lead to the low actual implementation rate of carbon footprint assessment in the food, retail, and small manufacturing industries, seriously hindering the promotion of green procurement, supply chain low-carbon management, and carbon labels at the consumption end, and becoming an important factor restricting the low-carbon development of the entire society.

In current research, carbon footprint accounting techniques mainly focus on standardized products and key high-energy-consuming industries, such as chemicals, steel, and automotive manufacturing. There is relatively little research on general accounting methods for non-standard products. Some studies attempt to optimize the accounting process by establishing an industry carbon factor database, but due to the diversity and differences of non-standard products, the coverage and applicability of the database are difficult to meet actual needs; some studies have explored the application of machine learning in carbon footprint prediction, but they mostly rely on a large amount of standardized training data and cannot adapt to the personalized characteristics of non-standard products[3-4]. Therefore, developing an AI-assisted carbon footprint accounting system that is oriented towards non-standard products, has both universality, efficiency, and convenience, has become a key direction to address the current pain points in the carbon accounting industry[5-8].

1.2. Research question

This study addresses the practical needs for the carbon footprint accounting of non-standard products, and proposes the core research questions: How to establish an intelligent system that does not rely on standardized carbon factor databases and manual modeling, and can quickly and accurately calculate the carbon footprint of non-standard products throughout their entire life cycle? Specifically, it includes: How to efficiently collect multi-dimensional characteristic information of non-standard products? How to achieve intelligent inference and precise calculation of carbon factors for non-standard products? How to develop a carbon footprint accounting visualization system that is easy to operate and suitable for small and medium-sized enterprises? How to verify the identification accuracy and accounting precision of this system in practical applications?

1.3. Research objective

This research focuses on addressing the pain points in the carbon footprint calculation of non-standard products. Centering on the technical path of "combining soft and hard approaches and leveraging AI", the following research goals are set.

Develop a comprehensive life-cycle carbon footprint estimation framework for non-standard products, clearly defining the technical specifications for information collection, factor inference, calculation, and report output throughout the process.

Create an integrated information collection and intelligent calculation system to achieve automatic collection of product physical information and text information, as well as intelligent calculation of carbon footprints.

Realize non-standard product information extraction and carbon factor inference based on OCR recognition and large language models, enabling semantic understanding and carbon emission estimation without the need for manual modeling.

Develop a visual human-computer interaction interface that supports batch input of product information, real-time display of carbon footprint calculation results, and automatic generation of standardized assessment reports.

Verify the information recognition accuracy and carbon footprint calculation precision of the system through actual tests, ensuring that the system meets the practical application needs of small and medium-sized enterprises.

1.4. Research significance

This study has constructed a universal carbon footprint estimation framework for non-standard

ISSN 2616-5767 Vol.9, Issue 2: 55-62, DOI: 10.25236/AJETS.2026.090207

products, filling the research gap in the existing carbon accounting theory in the field of non-standard products; it has for the first time deeply integrated OCR recognition, vector databases, and large language models to apply in carbon footprint accounting, enriching the application theory and method system of AI technology in the low-carbon field; it has clarified the characteristic information collection norms and accounting processes for carbon footprint estimation of non-standard products, providing theoretical references and technical frameworks for subsequent related research. The carbon emission footprint management system realizes on-site rapid identification and estimation of carbon footprints for non-standard products, significantly reducing the accounting cost and shortening the accounting cycle, solving the problem of insufficient professional capabilities of small and medium-sized enterprises in carbon accounting; the system can generate standardized carbon footprint reports and implementable emission reduction optimization plans, providing direct support for enterprises' green procurement, supply chain low-carbon management, and ESG report preparation; the application of the system can promote the popularization of carbon label systems in the field of non-standard products, enhance consumers' green consumption awareness, accelerate the construction of green supply chains, and help the entire society achieve the goals of carbon peak and carbon neutrality.

2. Research Methodology

2.1. System architecture

This system adopts a three-layer architecture of "hardware acquisition - AI processing - report output" to achieve end-to-end carbon footprint estimation for non-standard products. In terms of hardware, the system is compatible with common consumer-level hardware devices, including cameras (resolution ≥ 720P), electronic scales (accuracy ≥ 0.1g), and ordinary computers (CPU ≥ i5, memory ≥ 8G). No professional high-end acquisition equipment is required, which lowers the usage threshold for small and medium-sized enterprises. The AI processing layer is developed based on the Python programming language, with the core development environment being Python 3.8 and above versions, and the operating system being compatible with Windows 10/11 and Linux Ubuntu 20.04. The core tools and libraries used in system development include: Gradio 4.0 and above versions, PIL/Pillow, Requests, OpenAI, Base64, Pandas, Faiss, NumPy, SentenceTransformer, Json, etc. The overall architecture of the system is shown in Figure 1.

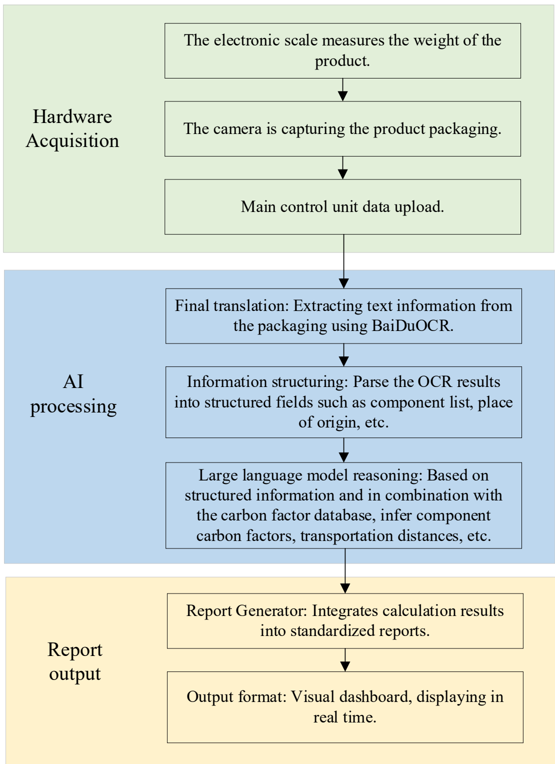

Figure 1: Overall Architecture Diagram of the System.

2.2. Physical information collection

The electronic scale is used to collect core physical information such as the net weight and packaging weight of the products, with an accuracy of up to 0.1g, meeting the weight collection requirements for various types of non-standard products (such as food, handicrafts, and custom small goods); cameras are used to collect image information such as product packaging, raw material labels, and production process records, supporting mainstream image formats such as JPG and PNG, providing a basis for subsequent text information extraction. Physical information collection supports real-time collection on-site, and the data can be directly synchronized to the system for subsequent processing.

2.3. Text information extraction

2.3.1. OCR Recognition

For the image information collected by the camera, the intelligent OCR interface provided by Baidu Intelligent Cloud is used to achieve the intelligent extraction of text information. It supports mixed recognition of Chinese and English, and can extract key text information such as product name, place of origin, ingredient list, net content, production process flow, transportation information, packaging materials, etc. The OCR recognition process is as follows: Firstly, the image data is encoded in Base64 format, and the Baidu OCR interface is called to complete text recognition. The recognition results are concatenated and preprocessed, invalid characters and redundant information are removed, and finally, the structured text content is output. If the call to the Baidu OCR interface fails, the system will automatically trigger local OCR recognition as a backup plan to ensure the stability of text information extraction.

2.3.2. Information structuring

The original text recognized by OCR needs to be converted into structured data. The system uses a large language model (DeepSeek-Reasoner) for semantic parsing. Through carefully designed prompts, the unstructured text is transformed into a standardized Markdown format. The prompts require the model to output tables and lists containing the following contents:

• Basic information: product name, place of origin

• Raw material data: raw material name, quantity, raw material place of origin

• Production process flow: process name, power consumption

• Transportation process: transportation segment, distance, vehicle type, loading rate

• Packaging information: material name, quantity

This step is implemented by the parse_product_info function, which combines the OCR results of five images and inputs them to the DeepSeek model to return structured product information.

2.4. Carbon Estimation Module

Carbon footprint estimation is the core module of this system. It integrates vector database carbon factor retrieval and intelligent reasoning technology of large language models to achieve precise calculation of carbon footprints for non-standard products throughout their entire life cycle (raw materials, production, transportation, packaging), without relying on standardized carbon factor databases or manual modelling.

2.4.1. Vector database construction

To achieve rapid and accurate query of carbon emission factors, this study has constructed a vector database based on SentenceTransformer and Faiss. The specific steps are as follows.

(1) Collect and organize the carbon factor data from mainstream carbon factor databases (Agribalyse, Ecoinvent, CLCD, China Packaging Association, National Logistics Standards, etc.), including various carbon emission factors such as raw materials, energy, transportation methods, and packaging materials;

(2) Use the SentenceTransformer model (m3e-base) to convert the text information related to carbon factors into high-dimensional vectors, with a vector dimension of 768;

(3) Utilize Faiss to build a vector index, associating the vector data with the structured carbon factor data (stored in pkl format) to achieve efficient retrieval of carbon factors;

ISSN 2616-5767 Vol.9, Issue 2: 55-62, DOI: 10.25236/AJETS.2026.090207

(4) Develop a vector database search function, supporting batch search based on the query words input by the user (such as "wheat carbon emission factor" "road transportation carbon emission coefficient"), and returning the carbon factor data with the highest similarity. The similarity calculation uses Euclidean distance to convert it into a similarity score to ensure the accuracy of the search results.

2.4.2. Carbon footprint calculation

The Deepseek-reasoner large language model is adopted as the core reasoning engine for carbon footprint calculation. Combined with the product information identified by OCR and the carbon factor data retrieved from the vector database, intelligent calculation of the entire life cycle carbon footprint is achieved. The specific process is as follows.

(1) The OCR-identified product information is preprocessed. The core accounting elements such as raw material data, production process flow, transportation process, and packaging information are extracted by the large language model and a structured product information table is generated;

(2) Based on the extracted accounting elements, the large language model automatically determines the type of carbon factor to be retrieved and generates a carbon factor query list, and calls the vector database for retrieval;

(3) The large language model, based on the retrieved carbon factor data, calculates the carbon emissions of the four stages (raw materials, production, transportation, and packaging) according to the preset life cycle accounting rules;

(4) The carbon emissions of the four stages are summarized to obtain the total carbon footprint of the product throughout its life cycle, and the proportion of carbon emissions in each stage is calculated.

The core accounting rules strictly follow industry standards. The calculation formulas for each stage are as follows.

Raw material stage:

$$E_{raw} = \sum(M_i \times C_{rawi})$$

where M_i represents the amount of the i-th raw material, and C_{rawi} represents the carbon emission factor of the i-th raw material. The priority of carbon factors is: vector database > Agribalyse > Ecoinvent > CLCD > industry average.

Production stage:

$$E_{prod} = \sum(E_j \times C_{ener})$$

where E_j represents the energy consumption of the j-th process, and C_{ener} represents the energy carbon emission factor (fixed at 0.485 kg CO_2e/kWh).

Transportation stage:

$$E_{trans} = \sum\frac{D_k \times W \times C_{transk}}{L_k}$$

where D_k represents the distance of the k-th transportation segment, W represents the weight of the product, C_{transk} represents the carbon emission coefficient of the k-th transportation method, and L_k represents the loading rate of the k-th transportation method. The priority of the accounting framework is: vector database > GLEC > national logistics standards.

Packaging stage:

$$E_{pack} = \sum(P_m \times C_{packm})$$

where P_m represents the amount of the m-th packaging material, and C_{packm} represents the carbon emission factor of the m-th packaging material. The data priority is: vector database > Ecoinvent > China Packaging Association.

Total life cycle quantity: $E_{total} = E_{raw} + E_{prod} + E_{trans} + E_{pack}$.

ISSN 2616-5767 Vol.9, Issue 2: 55-62, DOI: 10.25236/AJETS.2026.090207

2.4.3. Output feedback stage

The output feedback stage realizes the visual display of the accounting results and the generation of standardized reports, providing enterprises with clear carbon footprint accounting results and feasible emission reduction optimization plans. It specifically includes three modules: product information display, carbon footprint data display, and standardized diagnostic report generation.Product information display: The pre-processed product information is structuredly displayed in Markdown format, including basic product information (name, origin), raw material data, production process flow, transportation process, packaging information, etc., ensuring clear and traceable information. Carbon footprint data display: The carbon emissions at each stage, their proportions, and the total amount throughout the life cycle are displayed in Markdown table format, and detailed calculation details for each stage are provided, including calculation formulas, used carbon factors, data sources, etc., ensuring the transparency of the accounting results. Standardized diagnostic report generation: Based on the product information and carbon footprint accounting results, the large language model automatically generates a carbon footprint diagnostic report according to preset specifications. The report includes five main modules: core overview, stage decomposition, key emission hotspots, hierarchical optimization plans (P0: immediate implementation, P1: medium-term planning, P2: long-term collaboration), and emission reduction outlook. It also provides green certification recommendations and emission reduction target visualization (equivalent tree planting quantity), providing direct support for enterprises' low-carbon decision-making. The processing flow of the system software is shown in Figure 2.

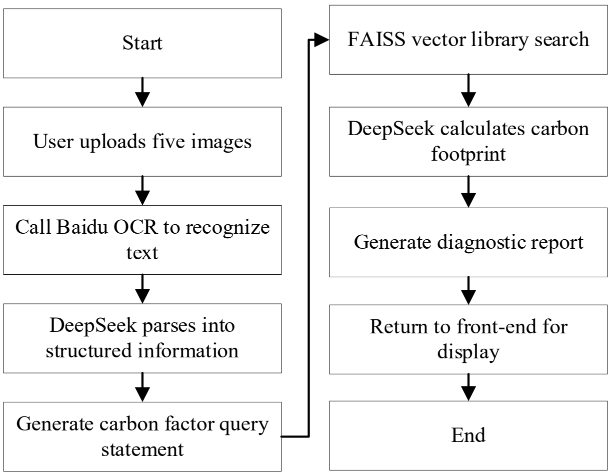

Figure 2: Flowchart of System Software Processing.

2.5. Front-end interactive interface

A lightweight and user-friendly visual human-computer interaction interface is developed using Gradio. It is compatible with ordinary computers and browsers, and does not require professional software operation skills. It meets the usage needs of small and medium-sized enterprises. The interface adopts a left-right panel layout. The left side is the data input area, which supports batch upload and real-time preview of product and various stage (raw materials, production, transportation, packaging) images. The right side is the analysis result area, which displays product information, carbon footprint data, and complete diagnostic reports in tabular form. The system provides a "Start Analysis" one-click operation. After clicking, it automatically completes information collection, OCR recognition, carbon factor retrieval, carbon footprint calculation, and report generation. The entire process does not require manual intervention, and the operation is convenient.

2.6. System Testing Methodology

To verify the practical application performance of the system, this study selected typical non-standard products such as food, handicrafts, and custom-made small goods as test samples. The system was tested from two dimensions: information recognition accuracy and carbon footprint calculation accuracy. Firstly, the packaging of the test products was recognized using OCR, and the total number of key information items recognized, the number of correctly recognized items, and the overall recognition accuracy were calculated. At the same time, the recognition accuracy of core information such as net content, ingredient list, and production date was separately counted, and the text information extraction ability of the system

ISSN 2616-5767 Vol.9, Issue 2: 55-62, DOI: 10.25236/AJETS.2026.090207

was analyzed. Then, the carbon footprint estimation results of the system were compared with the benchmark values of manual assessment by industry experts. The average absolute error rate, maximum error rate, and minimum error rate were calculated, and the error rate calculation formula is:

$$Error = \left| \left(V_{sys} - V_{exp} \right) / V_{exp} \right| \times 100\%$$

where V_{sys} is the system estimation value, V_{exp} is the expert assessment benchmark value, and the calculation of the system's calculation accuracy was verified.

3. Research results

3.1. OCR Recognition Performance

An OCR recognition test was conducted on the packaging of typical non-standard products. The system successfully identified 87 key information items, covering core information for carbon accounting of non-standard products such as product name, place of origin, net content, ingredient list, production date, packaging material, and transportation information. Among them, 80 items were correctly identified, and the overall recognition accuracy rate of the system was 90.2%. The recognition accuracy rates of each core information item are shown in Table 1.

Table 1: Statistics of OCR Recognition Accuracy.

Type	Number of items to be identified	umber of correctly identified items	Accuracy
Ingredient list	27	25	92.5%
Origin information	12	11	91.7%
Production date	13	11	86.9%
Net content	20	19	95.0%
Total	**87**	**80**	**90.2%**

From the test results, it can be seen that the system has the highest accuracy rate in identifying the net content, reaching 95.0%. The accuracy rates for identifying the ingredient list and product name are also both above 90%, indicating that the system has a strong ability to extract core accounting information for non-standard products. The accuracy rate of identifying the production date is relatively low (86.9%), mainly due to the small font size and concealed position of the production date on some non-standard products, as well as the presence of handwritten annotations, which increases the difficulty of OCR recognition. However, overall, it still meets the actual accounting requirements.

3.2. Accuracy of carbon estimation

The system's estimated carbon footprint results for the test samples were compared with the benchmark values provided by industry experts. The comparison trend between the system's estimates and the experts' assessment benchmarks is shown in Figure 3.

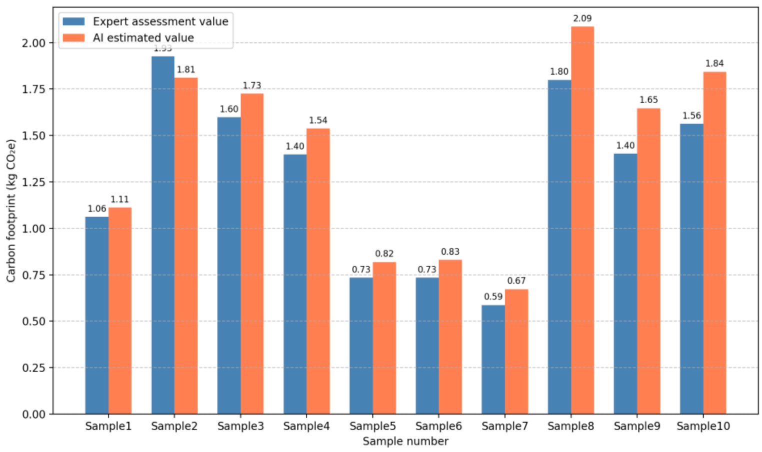

Figure 3: Bar Chart Comparing AI Estimation Results with Expert Evaluation Results.

From the test results, it can be seen that the carbon footprint calculation accuracy of the system is quite high. The average absolute error rate is 4.7%, which is lower than the industry-acceptable error threshold of 10%. The maximum error rate is 7.3%, mainly occurring in customized handmade product samples with complex raw material types and diverse transportation processes. The minimum error rate is 1.2%, occurring in local food samples with simple raw materials and processes. From the comparison trend chart, it can be observed that the system's estimated results have a high degree of fit with the expert evaluation benchmark values, with a consistent overall trend and no significant deviation, indicating that the system can accurately calculate the carbon footprint of non-standard products.

4. Conclusions

This study addresses the lack of standards, high costs, and poor generalizability in non-standard product carbon footprint accounting. By adopting a "software-hardware-AI" integrated approach, it establishes a lifecycle carbon footprint estimation framework and develops the Carbon Emission Footprint Manager—an AI-assisted system enabling automatic data collection, intelligent calculation, and standardized reporting.

Key conclusions include: (1) A three-stage technical framework ("information collection—carbon estimation—output feedback") fills the gap in universal accounting methods for non-standard products. (2) The integrated collection system, combining electronic scales, cameras, and Baidu OCR, achieves 90.2% overall recognition accuracy, with core information (net content, ingredient list) exceeding 90%. (3) The estimation module, integrating vector databases and Deepseek LLM, enables rapid carbon factor retrieval and intelligent calculation without manual modeling, achieving a low average error rate of 4.7%. (4) The Gradio-based interface reduces single-product processing time to ≤3 minutes, improving efficiency by over 90% and lowering the barrier for SMEs. (5) The system automatically generates comprehensive diagnostic reports, offering actionable emission reduction recommendations.

This research pioneers a general architecture for non-standard product carbon accounting, enriching AI applications in the low-carbon field. The system empowers SMEs in food, retail, and manufacturing, promoting carbon labeling adoption and green supply chains. With strong performance and scalability, it holds significant potential for advancing carbon peak and neutrality goals through continued optimization and broader application.

References

[1] Zhu Liyang. Accelerating the Construction of Waste Recycling System to Assist in Achieving the Goals of Carbon Peak and Carbon Neutrality - Policy Interpretation No. 4 [J]. Resource Recycling, 2024, (02): 56-57.

[2] Shu Yinbiao, Zhao Liang, Yi Jun, et al. Methodological System and Technology Research on Product Carbon Footprint Accounting [J/OL]. Proceedings of the CSEE, 1-12 [2026-02-26]. https://link.cnki.net/urlid/11.2107.tm.20260121.1826.008.

[3] Wilting C H ,Veld ' I D .An overview of accounting approaches for national greenhouse gas emissions; Comparing Consumption and Production Footprints in the European Union [J]. Economic Systems Research, 2025, 37(3): 402-426.

[4] Xie Y, Zhou B, Wang Z, et al. Industrial Carbon Footprint (ICF) Calculation Approach Based on Bayesian Cross-Validation Improved Cyclic Stacking [J]. Sustainability, 2023, 15(19).

[5] Zhang Y, Yao S, Chen Y. Developing a carbon footprint calculation method for product life cycle based on low-carbon design: A case study of the STAGE Bluetooth speaker [J]. PLoS one, 2025, 20(8): e0327576.

[6] Feng J , Jiang Y , Xu J ,et al.Photonic-assisted division calculation of microwave signal envelopes and its applications[J].Optics Express, 2024, 32(24):43692.DOI:10.1364/oe.542473.

[7] Wang X C , Li W K , Zhang Y T .Deep Neural Network-Assisted Efficient Calculation on CO2 Pulsed Discharges Under Martian Pressure for Various Pulse Rise Rates[J].IEEE Transactions on Plasma Science, 2024(5):52.DOI:10.1109/TPS.2024.3404074.

[8] Zhang Y , Liu X , Wang W .Theoretical Calculation Assisted by Machine Learning Accelerate Optimal Electrocatalyst Finding for Hydrogen Evolution Reaction[J].ChemElectroChem, 2024, 11(13). DOI:10.1002/celc.202400084.