Automatic Classification System of Chinese Medicinal Materials Based on Visual Perception and Feature Extraction

Jia Weng[1], Wanying Zhang[2], Chun Yang[2,*]
[1]*School of Artificial Intelligent, Guangzhou Huashang College, Guangzhou, Guangdong, China*
[2]*School of Accounting, Guangzhou Huashang College, Guangzhou, Guangdong, China*
**Corresponding Author*

Abstract: Chinese medicinal herbs are diverse and complex in form, with some varieties being similar, which can easily lead to confusion and misidentification, affecting the quality of herbs and clinical safety. Traditional identification methods rely heavily on expert experience, making them subjective, inefficient, and unable to meet modern demands. This paper proposes an automatic classification system for Chinese medicinal herbs based on convolutional neural networks, integrating modules for image acquisition, preprocessing, feature extraction, and classification recognition. By constructing a variety of Chinese medicinal herb image datasets and employing optimization strategies such as transfer learning, feature fusion, and data augmentation, the accuracy, generalization ability, and robustness of the system have been significantly improved. Experimental results show that the system achieves an accuracy rate of 95.5% in classification tasks, demonstrating broad application potential for areas such as herb quality inspection, market regulation, and education and training.

Keywords: Chinese Herbal Medicine Identification; Image Classification; Deep Learning; Convolutional Neural Network; Visual Perception

1. Introduction

1.1 Research Background

With the implementation of policies and regulations such as the "Traditional Chinese Medicine Law," the TCM industry has entered a stage of rapid development. However, issues like substandard herbs, confusion and substitution, and adulteration continue to occur frequently, posing serious threats to the healthy development of the industry and medication safety. As public health awareness increases and TCM culture becomes more widespread, domestic and international market demand continues to grow. The efficacy of different herbs varies significantly, and accurate classification is essential for ensuring safety. In recent years, the state has issued documents such as the "Plan for the Construction of the Quality Assurance System for Chinese Herbal Medicines" and the "Guidance on Strengthening the Quality Supervision of Traditional Chinese Medicine," emphasizing the need to strengthen quality supervision, improve traceability systems, and promote standardization, informatization, and intelligence in the production and circulation of Chinese herbal medicines. Therefore, there is an urgent need for an objective, precise, and efficient technical means to ensure the quality and safety of Chinese herbal medicines.

1.2 Limitations of Traditional Identification Methods

Traditional methods for identifying Chinese medicinal herbs mainly rely on manual observation and experience, which are subjectively biased, inefficient, and difficult to standardize. When faced with the complex diversity and similar appearances of these herbs, manual identification not only requires extensive professional knowledge and long-term accumulated experience but also carries the risk of human error. This is especially true in primary healthcare facilities, where there is a lack of high-level experts and specialized equipment, limiting the promotion and popularization of traditional methods. As the market for Chinese medicinal herbs continues to expand, traditional identification methods can no longer meet the growing demands for quality testing and supervision.

1.3 Progress in the Application of Deep Learning in Image Recognition

In recent years, with the improvement of computing power and the development of big data, deep learning has made breakthrough progress in the field of computer vision, becoming a mainstream technology for tasks such as image recognition, object detection, and image segmentation [1]. In particular, models represented by convolutional neural networks have repeatedly broken performance records in public image classification competitions like ImageNet and CIFAR, driving the rapid development of image recognition technology.

In the fields of medical imaging, crop disease identification, and industrial defect detection, deep learning has demonstrated superior performance over traditional methods [2]. Specifically, in the image recognition of traditional Chinese medicine herbs, research has explored using deep learning architectures such as CNNs, Residual Networks (ResNet), Dense Connections (DenseNet), and attention mechanisms to classify herbs, achieving high recognition accuracy. In the field of TCM image recognition, Tian et al. constructed an image dataset NB-TCM-CHM dedicated to herb fruit classification and conducted deep learning classification experiments based on it, providing important benchmark resources for model evaluation. Gao et al. further proposed the TCM2Vec model, which constructs independent feature spaces to achieve deep representation of TCM formulas and image features, expanding the research dimensions of deep learning in TCM image semantic expression. However, due to the numerous varieties, complex appearance, and subtle differences among some herbs, existing methods still face issues such as decreased classification accuracy and insufficient generalization ability when dealing with similar herbs, complex backgrounds, and changes in lighting [3]. In addition, there are few large-scale image data sets available in the field of Chinese medicinal materials, which further limits the training effect and promotion and application of deep learning models.

1.4 Progress in the Application of Deep Learning in Image Recognition

In response to the aforementioned issues, this paper proposes and develops an automatic classification system for traditional Chinese medicinal materials based on visual perception and feature extraction [4]. The system centers around a deep convolutional neural network, integrating modules such as image preprocessing, convolutional feature extraction, pooling dimensionality reduction, and fully connected classification, forming a complete process from data input to classification output. By employing multi-layer convolutional networks to extract and integrate multidimensional visual features like texture, shape, color, and edges, the system enhances its ability to classify complex and similar-shaped medicinal materials, achieving efficient and accurate detection and classification. To address issues such as subtle morphological differences, background interference, and class imbalance in the identification of medicinal materials, the system optimizes through techniques like data augmentation, class weight adjustment, feature fusion, image denoising, and standardization, significantly improving recognition accuracy, generalization capability, and robustness. Studies have shown that multi-strategy platforms are of great significance in the origin and authenticity identification of Chinese medicinal materials, while this system focuses on the rapid identification task at the visual end, aiming to improve the efficiency and reliability of on-site identification. This system provides strong support for intelligent recognition of medicinal materials, contributing to the digital and intelligent development of the traditional Chinese medicine industry.

2. Review of Relevant Studies

2.1 Research Progress in Image-Based Identification of Traditional Chinese Medicinal Materials

With the rapid development of computer vision and deep learning technologies, the demand for intelligent identification of traditional Chinese medicinal materials (TCMs) has been increasingly rising. Scholars at home and abroad have conducted extensive research focusing on "image-based TCM identification". Most existing studies adopt deep learning models such as convolutional neural networks (CNNs) to classify TCMs and have achieved certain results: Zhu et al. combined the "segmentation-combination data augmentation strategy" with a "dual attention mechanism", which improved the recognition performance of

TCM micro-images in small-sample and fine-grained tasks; Hou et al. introduced "knowledge distillation" and a "cross-attention mechanism" to construct a lightweight TCM recognition network, which not only has advantages in deployment on mobile devices but also further enhances the model's generalization ability and accuracy [5-6].

2.2 Limitations of Existing Research on TCM Identification

Current research still faces challenges in multiple aspects: First, at the data level, public TCM image datasets are insufficient in quantity, lack sample diversity, and have unbalanced category distributions, making them difficult to meet the training needs of deep learning models. Second, at the model and method level, most studies rely on a single model structure, with limited exploration of optimization directions such as "multi-scale feature fusion" and "category weighting", resulting in restricted generalization ability of the models. Third, at the practical application level, although existing models perform well in image classification, they are highly sensitive to complex backgrounds and lighting changes and prone to overfitting issues, leaving significant room for improvement in TCM recognition performance.

3. System Design and Implementation

3.1 System Architecture

This system employs deep convolutional neural networks as the core algorithm and has designed a lightweight convolutional neural network for the task of identifying traditional Chinese medicinal herbs. This reduces model parameters and computational load while ensuring accuracy and improving operational efficiency. The overall architecture includes four major modules: image acquisition and processing, model construction and training, feature extraction and classification recognition, and model optimization, forming an intelligent identification process for traditional Chinese medicinal herbs. The system structure is shown in Figure 1.

3.2 Image Acquisition and Processing Module

In this project, we collected 10,000 photos of various types of traditional Chinese medicinal materials to form a dataset. The system uses industrial-grade imaging equipment as the hardware foundation for image acquisition, featuring a camera equipped with high-performance CMOS sensors capable of achieving a high resolution of 3027×2048, which meets the need for high-precision capture of the fine textures and intricate structures of traditional Chinese medicinal materials. The device supports real-time image acquisition, as illustrated in Figure 2, and features strong low-light imaging capabilities and a high dynamic range, ensuring stable image quality under various lighting conditions.

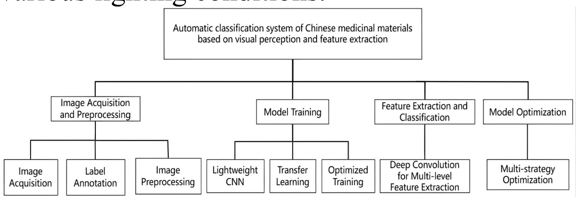

Figure 1. System Module Flow Chart

Figure 2. Image Acquisition

To meet the identification needs of different types of traditional Chinese medicinal materials, the system is equipped with corresponding label annotation mechanisms during the collection process, as shown in Figure 3.

Figure 3. Data Set Collected Under the Label

Despite this, existing studies still face issues such as small dataset sizes, insufficient sample diversity, model overfitting, and sensitivity to complex backgrounds and lighting. Currently, there are some publicly available Chinese herbal medicine image datasets, but they are limited in variety and have imbalanced category distribution, making it difficult to meet the training requirements of deep learning models. Moreover, most existing research focuses on a single model structure, with insufficient exploration of multi-scale feature fusion, category weight adjustment, and model generalization [7]. Although deep learning models, such as CNNs, fully connected neural networks, and deep neural networks DNNs, excel in image classification, there is still room for improvement in the field of Chinese herbal medicine recognition.

3.3 Model Construction and Training

Neural network models for image classification have always been a crucial research direction. With the development of deep learning technology, the requirements for neural network models are becoming increasingly stringent. This paper employs a lightweight convolutional neural network (CNN) to construct an image classification model for traditional Chinese medicine (TCM) [8]. The innovation in lightweight convolutional networks reduces parameters and computational costs while maintaining model performance, balancing accuracy and computational efficiency in TCM image recognition tasks. By leveraging transfer learning and fine-tuning (fine-tuning) strategies, the model effectively utilizes knowledge from large-scale image classification models to transfer features. Specifically, the model uses pre-trained weights as initial parameters and re-trains on specialized TCM datasets to adapt and optimize for the target task.

The convolutional neural network structure adopted in this study is shown in the Figure 4. The overall architecture includes two convolutional layers, two pooling layers, and two fully connected layers. It combines the ReLU nonlinear activation function with a Softmax classifier to achieve multi-classification tasks for images of traditional Chinese medicinal materials. The structure and functions of each layer of the network are described as follows:

Firstly, the input_layer receives the preprocessed image data, and the input size is determined according to the original image size and experimental configuration.

Subsequently, the first convolutional layer (conv1) is entered, which uses 205x5 convolutional kernels with a stride of 1 and padding of "same," maintaining the output feature map size consistent with the input. After the convolution operation, the results are added to the bias term, and then passed through the ReLU activation function (relu1) to introduce nonlinearity.

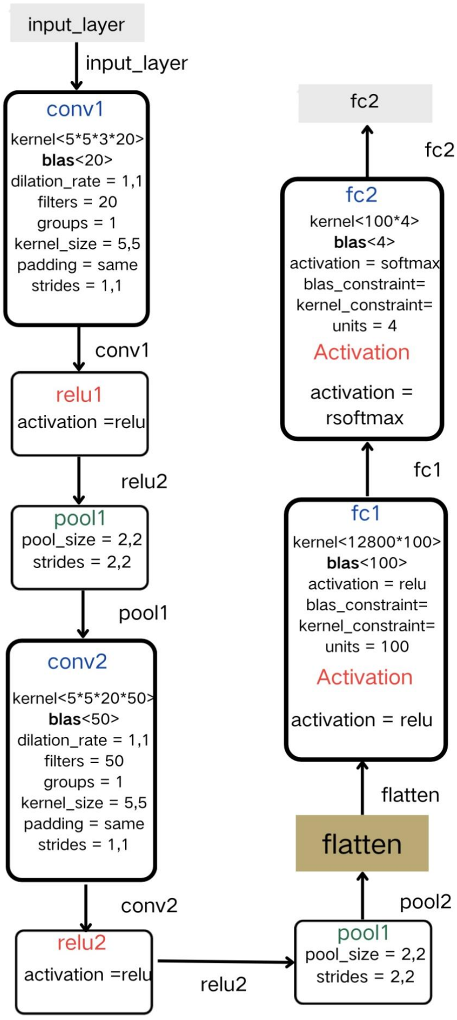

Figure 4. Model Structure

In the first pooling layer (pool1), a 2x2 maximum pooling kernel (pool_size) and stride of 2 are used to reduce the size of the feature map and extract local invariance through downsampling operation.

The second convolution layer (conv2) is followed, which uses 505x5 convolution kernels. Other parameters are the same as conv1, and the output channel number is increased from 20 to 50 to extract deeper features. The convolution output is also processed by ReLU activation function (relu2).

The second pooling layer (pool2) also uses a 2x2 pooling kernel and stride of 2 to further compress the dimension of the feature map through down sampling again, so as to prepare fixed length feature vectors for the subsequent fully connected layer.

After completing the convolution and pooling operations, the feature map is flattened into a one-dimensional vector through the flattening layer (flatten), serving as the input for the fully connected layer. The first fully connected layer (fc1) receives an input vector of length 12800 and outputs a 100-dimensional feature representation, using ReLU as the activation function.

Finally, the second full connection layer (fc2) maps the 100-dimensional features to four category nodes, and outputs the category probability distribution calculated by Softmax activation function to realize the classification prediction of four kinds of Chinese medicinal materials.

The network structure has the ability to gradually extract local and global features from shallow to deep layers. The discriminative image features of Chinese medicinal materials are extracted step by step by using convolution layer and pooling layer, and then the feature integration and final classification are completed by full connection layer.

In the training process, as shown in Figure 5. Adam optimizer is set as the optimization algorithm for model parameter update. Adam integrates momentum method and initial learning rate of 0.001, and adopts learning rate decay and early stop mechanism to prevent overfitting.

In terms of loss function, the model adopts the cross-entropy loss function to measure the difference between the predicted class distribution and the actual labels, suitable for multi-class scenarios. To further enhance training performance, the system integrates a learning rate decay strategy, gradually reducing the learning rate based on training rounds or validation set performance. This prevents the model from getting stuck in local optima or oscillating later on, thereby improving generalization capabilities.

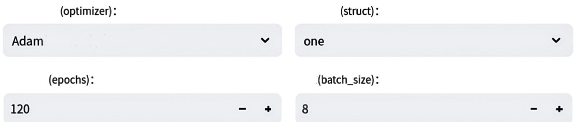

Figure 5. Model Parameter Setting

Figure 6 shows the training process of the model. Through the above strategies, the model not only effectively inherits the general visual representation ability of large-scale data sets, but also learns targeted features through re-training on Chinese herbal medicine images, so that it still has good classification accuracy and generalization performance in the context of numerous types of Chinese herbal medicine and complex texture forms.

```
Epoch 4/120 : - acc: 0.8750 - loss: 0.3301 - val_acc: 0.9179 - val_loss: 0.2173
Epoch 5/120 : - acc: 0.8887 - loss: 0.2857 - val_acc: 0.9349 - val_loss: 0.1825
Epoch 6/120 : - acc: 0.5000 - loss: 0.9027 - val_acc: 0.9489 - val_loss: 0.1375
Epoch 7/120 : - acc: 0.9224 - loss: 0.1983 - val_acc: 0.9630 - val_loss: 0.1173
Epoch 8/120 : - acc: 0.8750 - loss: 0.2150 - val_acc: 0.9610 - val_loss: 0.1115
Epoch 9/120 : - acc: 0.9451 - loss: 0.1444 - val_acc: 0.9600 - val_loss: 0.0921
Epoch 10/120 : - acc: 1.0000 - loss: 0.0321 - val_acc: 0.9560 - val_loss: 0.097
Epoch 11/120 : - acc: 0.9576 - loss: 0.1178 - val_acc: 0.9900 - val_loss: 0.039
Epoch 12/120 : - acc: 1.0000 - loss: 0.0196 - val_acc: 0.9890 - val_loss: 0.041
```

Figure 6. Model Training Process

3.4 Feature Extraction and Classification Recognition

The feature extraction module, a key system component, utilizes deep CNNs to progressively learn and extract discriminative multi-scale features from TCM herb images. Shallow layers capture local details like edges and textures, while deeper layers aggregate broader semantic information such as shape, structure, and color. This hierarchical abstraction enables robust feature identification even under complex backgrounds. To improve information integration across levels, a multi-scale feature fusion strategy combines low-level details with high-level semantics, preserving fine textures while enhancing global morphology. This boosts recognition performance under variable lighting, cluttered scenes, and high inter-class similarity. Transfer learning is applied by fine-tuning convolutional weights, adapting feature extraction to TCM-specific characteristics. Finally, the module outputs high-dimensional feature vectors rich in visual and discriminative information for classification. This end-to-end feature learning process ensures effective mapping from raw images to feature space, enabling precise classification.

The classification module, as the system's core output, uses the fully connected and Softmax layers to compute category probabilities from feature vectors. It selects the highest-confidence label as output, realizing automated recognition of traditional Chinese medicinal materials.

In terms of result presentation, the system not only provides text form classification tag output, but also combines visualization technology to intuitively display the recognition results in the form of highlight and image mark in the interface, as shown in Figure 7.

Figure 7. Classification Recognition Results

The module achieves efficient mapping from feature space to label space, ensuring the accuracy and real-time performance of recognition results. Additionally, it addresses common issues in traditional Chinese medicine identification, such as subtle morphological differences, background interference, and class imbalance, through targeted optimizations in algorithm design and data processing. By employing techniques like data augmentation, category weight adjustment, feature fusion, image denoising, and standardization, the system significantly enhances its classification accuracy, generalization ability, and robustness. Overall, the system design provides an efficient, reliable, and practical technical solution for intelligent identification of traditional Chinese medicines.

3.5 Model Optimization Strategy

To further enhance system performance, this study integrates various advanced optimization strategies, including multi-scale feature fusion (FPN), data augmentation, Dropout (0.5), L2 regularization, and mixed-precision training. These techniques aim to boost the model's resistance to interference, improve classification accuracy, and enhance its robustness in complex scenarios. Specifically, multi-scale feature fusion (FPN) enhances the model's ability to recognize small targets and local features by integrating information from different scales, making it better suited for handling subtle morphological differences in traditional Chinese medicine images [9]. Data augmentation techniques, such as rotation, flipping, and scaling, expand the training dataset, effectively improving the model's generalization ability and reducing the risk of overfitting. Dropout and L2 regularization work together to prevent overfitting during training, enhancing the model's generalization performance and stability. Mixed-precision training accelerates the training process and reduces memory consumption by using lower-precision floating-point operations, ensuring high computational accuracy while maintaining training efficiency.

In addition, to achieve efficient system deployment, this model supports operation on multiple platforms, including Web and mobile ends. The inference speed has been optimized to meet real-time application requirements, especially maintaining an inference time of less than 300 milliseconds on mobile devices. This ensures that the system can provide rapid responses in practical applications, meeting the immediate needs for intelligent recognition of traditional Chinese medicinal materials.

4. Experiment and Result Analysis

4.1 Data Set and Experimental Settings

This study has built an image dataset containing multiple types of traditional Chinese medicinal herbs, collecting a total of 10,000 images. The dataset covers a wide range of herbal varieties to ensure that the model can handle the diversity and complexity of the herbs 'appearances. The dataset is divided into training, validation, and test sets in a ratio of 7:2:1. During the training process, data augmentation techniques were employed, including rotation, scaling, color jittering, random cropping, blurring, and noise addition, to enhance the model's robustness and generalization ability.

4.2 Experimental Results

In the test set experiment, the system showed excellent classification effect in some test batches, and the result of all 20 samples being correctly classified appeared many times, showing the strong recognition ability of the model on specific samples, as shown in Figure 8.

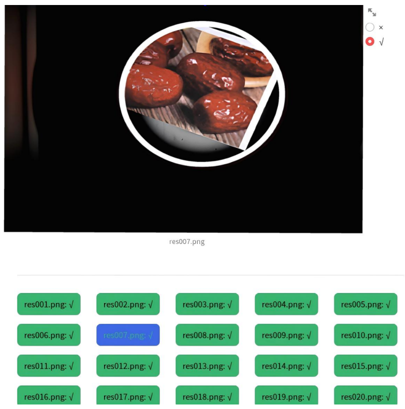

Figure 8. Model Test Results

However, in the overall statistics of multiple rounds of experiments, the system's average classification accuracy was stable at 95.5%, the average recall rate was 94.8%, and the F1 score was 95.1%. This result comprehensively reflects the overall performance of the model on large-scale test sets.

Compared to traditional Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) models, the system's classification accuracy has improved by 12% and 15%, respectively, demonstrating significant advantages. Confusion matrix analysis indicates that some herbs with similar colors still exhibit some degree of confusion during classification, suggesting that there is room for improvement in handling visually similar herbs. In the future, introducing multimodal data can further enhance the model's ability to distinguish in complex scenarios.

5. Discussion and Application Analysis

The system has been validated in simulation application scenarios. Experimental results show that compared to traditional image-based medicinal plant recognition methods, offers higher accuracy and a better user experience, effectively meeting the diverse application needs of quality inspection of traditional Chinese medicines, market regulation, and education and training. In mobile deployment tests, the system's inference time for a single image is controlled within 300 milliseconds, fully satisfying the response requirements of real-time application scenarios. User feedback indicates that the system's interface design is simple and intuitive, allowing even non-professional users to quickly get started, demonstrating good practicality and potential for widespread adoption.

However, in low-light environments, when parts of the herbs are missing or obstructed by impurities, the system's recognition accuracy decreases to some extent. This phenomenon indicates that the model still has limitations when dealing with complex shooting conditions and non-ideal images [10]. To further enhance the system's adaptability and robustness, future efforts can involve incorporating multimodal information, such as infrared imaging, Raman spectroscopy, and near-infrared spectroscopy. By integrating data from different modalities, the model can improve its comprehensive perception of both internal and external characteristics of traditional Chinese medicines. Additionally, introducing more advanced neural network architectures, attention mechanisms, and graph neural networks can also enhance the model's ability to express subtle morphological differences.

6. Conclusion and Work

This paper presents an automatic classification system for traditional Chinese medicinal materials using visual perception and deep learning. The system is built on a convolutional neural network and integrates transfer learning, feature fusion, regularization, and mixed-precision training to form an end-to-end intelligent identification framework. Experiments demonstrate strong performance in accuracy, generalization, and inference speed, meeting practical needs in quality control, regulation, and education.

Future work will focus on: (1) integrating multi-modal features to improve complex sample recognition; (2) adopting deeper networks and attention mechanisms for better fine-grained feature extraction; (3) exploring reinforcement and incremental transfer learning for small-sample adaptability; and (4) enhancing model interpretability to increase transparency and user trust. These improvements aim to advance intelligent TCM recognition toward greater precision, adaptability, and broader application, supporting TCM modernization and digitization.

Acknowledgments

This work was supported by Guangzhou Huashang College Research Mentor Program (2020HSDS03).

References

[1] Lv W, Kou F, Wang G, et al. Research on road surface perception strategy for intelligent suspension based on vision and dynamics fusion. Measurement Science and Technology, 2025, 36 (9): 096118-096118.

[2] Ashurov A, Qu H. An efficient industrial defect detection based on hybrid residual attention with modified generative adversarial network and convolutional neural network model. Computers and Electrical Engineering, 2025, 127 (PA): 110580-110580.

[3] Wanqing G, Ning C, Guojiang X, et al. TCM2Vec: a detached feature extraction

deep learning approach of traditional Chinese medicine for formula efficacy prediction. Multimedia Tools and Applications, 2023, 82 (17): 26987-27004.

[4] Chen B, Yu J, Feng H, et al. Adaptive Cross-Modal Non-Invasive Detection of Crucian Carp Freshness Based on Flexible Haptics and Multi-Angle Visual Perception. Food and Bioprocess Technology, 2025: 1-16.

[5] Zhu X, Pang G, He X, et al. A segmentation-combination data augmentation strategy and dual attention mechanism for accurate Chinese herbal medicine microscopic identification. Frontiers in Plant Science, 2024, 151442968-1442968.

[6] Hou Q, Yang W, Liu G. Chinese herbal medicine recognition network based on knowledge distillation and cross-attention. Scientific Reports, 2025, 15 (1): 1687-1687.

[7] Hu K, Lin H C, Jin X, et al. Meta-YOLOv8: multi-scale few-shot object detection for Chinese medicinal decoction pieces. Journal of King Saud University Computer and Information Sciences, 2025, 37 (7): 201-201.

[8] Wang T, Chen J, Liu C, et al. Identification of H2O ice at lunar south polar region based on deep learning. Icarus, 2025, 441116682-116682.

[9] Shiney S T, Jerome A S. Optimized attention Induced multi head convolutional neural network with Densenet201 for cervical cancer diagnosis. Biomedical Signal Processing and Control, 2025, 110(PB):108166-108166.

[10]Chen Z, Lin Z, Wan H, et al. Rapid and accurate metabolite identification of traditional Chinese medicine based on UPLC-Q-TOF-MS coupled with UNIFI analysis platform and quantitative structure-retention relationship: Danshen-Honghua herbal pair as an example. Journal of Pharmaceutical and Biomedical Analysis, 2025, 257116696-116696.

