J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 7759--7776

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on micro-morphological feature recognition method of traditional Chinese medicinal herbs based on deep convolutional neural network

Hanying Wang[1,2], Zhi Chen[1,2], Jiabo Huo[1], Xingguo Han[1,2,✉]

[1] Guangxi Key Laboratory of Special Engineering Equipment and Control, Guilin University of Aerospace Technology, Guilin, Guangxi, 541004, China
[2] School of Mechanical Engineering, Guilin University of Aerospace Technology, Guilin, Guangxi, 541004, China

ABSTRACT
Based on the excellent achievements of deep learning technology in image recognition in several fields, then the convolutional neural network is expected to play its superior performance in the recognition of micro features of Chinese herbal medicines. The object of analysis in this paper is the microscopic feature images of Chinese herbal medicines, and the residual network will be improved in order to improve the recognition accuracy of the network model on the microscopic images of herbs. On the basis of the traditional CNN network model, CBAM based on mixed domain attention mechanism is added, and residual connection is introduced to increase the transfer of gradient and information flow, preserve image feature data and reduce feature loss. Improved from the traditional residual structure to moving inverted bottleneck convolution (MBConv), the SE module and SAM module are added to the MBConv stage respectively to optimize the feature extraction performance and improve the accuracy of the classification of microscopic features of traditional Chinese medicine. The effect of the addition of the attention mechanism on the network model is analyzed, and the network model is examined in conjunction with the constructed dataset of powdered microscopic images of commonly used Chinese medicinal herbs.The average accuracy of the Attention-TCM-Net network model on the test set reaches 96.47%, which is an improvement of 0.85 percentage points compared with that of the ResNet34 network, and meanwhile, the convergence of the model is significantly better than other models.

Keywords: CBAM, MBConv, feature recognition, powder micrographic images, microscopic features, traditional Chinese medicine herbs

✉ Corresponding author.
E-mail address: Hxingg2010@163.com (X. Han).
Received 15 January 2024; Revised 25 March 2024; Accepted 30 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-425
© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

There are many kinds of traditional Chinese medicines with different sources, including wild animals and plants, as well as home-grown, home-raised and processed traditional Chinese medicines. There are the original animal and plant products of Chinese herbal medicines, as well as various kinds of artifacts, the origin of the original animal and plant is different, the family, genus and species are also different, and the artifacts are even more varied, with different specifications [1-4]. This causes the Chinese herbal medicine sometimes real, fake products mixed, the advantages and disadvantages are difficult to distinguish, affecting the production, management and use of Chinese medicine. The main purpose of Chinese medicine identification is to identify the authenticity of Chinese medicines, to ensure the accuracy, safety and effectiveness of people's medication, to find and expand new sources of medicines, to develop the production of Chinese medicines, and to lay the foundation for the development of Chinese medicine [5-8].

Shape identification of traditional Chinese medicine is an important content in the field of traditional Chinese medicine, which is mainly judged and identified by observing and identifying the appearance of Chinese herbal medicines in terms of morphology, color, texture and other characteristics [9-11]. Common methods of shape identification of Chinese herbal medicines include observing the appearance and morphology, observing the color, touching the texture, and identifying the smell, and shape identification can also be performed based on the taste, crushing, infiltration, and other characteristics of Chinese herbal medicines [12-15]. It should be noted that the morphological characteristics of Chinese herbal medicines may be affected by a variety of factors, such as origin, treatment, etc., so it is necessary to combine the characteristics of multiple aspects for comprehensive judgment and analysis when performing shape identification [16-19].

In this paper, according to the principle of dataset construction, different Chinese herbal medicine powder microscopic images are collected by collecting and photographing to establish a common Chinese herbal medicine powder microscopic image dataset. And the microscopic images of Chinese herbal medicines in the dataset are preprocessed to improve the recognition performance of the algorithm model on the features of the herbs. Because the traditional CNN model is deficient in recognition speed and accuracy, this paper designs a micro feature recognition network of Chinese herbal medicines based on the attention mechanism by adding a hybrid attention module in the residual structure, and optimizing the feature extraction in the convolutional layer by the CBAM module. Analyze the effect of the addition of the attention mechanism in this network structure on the model performance, combine with the classical convolutional neural network model to compare the model recognition results, and carry out the actual measurement of the micro features of different categories of traditional Chinese medicinal herbs respectively.

2. Chinese herbal medicine powder microdata set construction and image pre-processing

2.1. Principles of data set construction

The construction of the dataset needs to follow several principles:

1) Generalization ability and representativeness
2) Labeling accuracy
3) Balance
4) Clarity and quality

2.2. Micrographic image preparation of Chinese herbal medicine powder

Based on the above principles, a dataset of powdered microscopic images of Chinese herbal medicines was constructed, including herb collection, grinding, screening, filming, microscopic observation, determining labels, and data storage. Each step and its operation specification will be described in detail below:

1) Herb collection. Herb collection is the first step in the construction of the dataset, which requires the selection of representative samples of Chinese herbs.

When collecting herbs, it is necessary to pay attention to the following: choose a region with good ecological environment and no pollution, and select the best time for collection according to the growth cycle and maturity of herbs. The selection of herbs should have obvious characteristics and representativeness.

The experimental team collected 150 types of Chinese herbs from Tibet, Hebei and other places and pharmacies with appropriate growth levels. For each kind of herbs, three to six samples were selected that conformed to the pharmacopoeia or relevant standards, totaling more than 700 samples. The information of origin and time was also recorded.

2) Grinding. When grinding, attention should be paid to the selection of appropriate grinding tools, such as mortar, pestle or electric grinder, to ensure that the grinding process does not lose the microscopic characteristics of the herbs.

3) Screening. Screening is to remove impurities and unnecessary particles generated during the grinding process, and retain the powder that meets the observation requirements. When screening, attention should be paid to the selection of appropriate sieve mesh to ensure that the sieved powder is of uniform particle size.

4) Preparation. Preparation is to screen the powdered herbs into slides for microscopic observation.

The experiment was carried out by taking an appropriate amount (about 1.0 mg) of ground herbal powder with clean forceps and placing it in the center of the slide. Using a dropper, a drop of water or appropriate stain (not more than 70 μL) was added to the slide to highlight the cell structure. The coverslip was then carefully placed on the slide at an oblique angle and lowered slowly to avoid air bubbles and to disperse the powder as much as possible. Finally, absorbent paper was used to absorb the excess liquid.

5) Microscopic observation. Microscope observation is a key step to obtain the microscopic characteristics of the powder of herbs. An Olympus BX50 optical microscope was chosen for this experiment.

6) Determination of labeling. Determining the label is the process of corresponding the observed microscopic features to the type of herbal medicine and the type of microscopic features. The label includes the type of herbal medicine and the type of microscopic features.

7) Data organization and storage. After completing the microscopic observation, it is necessary to organize the observed microscopic feature images with the corresponding labels and store them properly.

An Olympus DP72 microdigital camera was chosen for this experiment and captured fully outlined microscopic features at 50 μm below the objective scale.

8) Data quality control. To ensure the quality of the data set, data quality control is needed.

In this experiment, 150 kinds of herbal powders were microfilmed, and a total of 63,252 microimages of herbal powders with bit depth of 24 and pixel size of 4140×3096 were obtained for dataset construction.Next, we will categorize the microimages of herbal powders according to the types of herbs and the types of microscopic features in order to construct the dataset.

2.3. Microscopic Image Feature Generalization and Dataset Construction

The collected microscopic images of powdered Chinese herbal medicines cover a total of 150 commonly used Chinese herbal medicines. The total number of microscopic features included in the 150 herbs is more than 60. 60 microscopic features are too complicated and too many. It is easy to cause the model of the subsequent study to be easily overfitted and the performance is attenuated.

Therefore, all the microscopic features involved in this experiment were summarized and merged professionally for subsequent studies. Some of the microscopic features that were pharmacologically similar and had similar image feature details were merged. Finally, more than 67 microscopic features were combined into 20 microscopic features: secretory tissue, secretory cells, fibers, crystals, secretory cavities, glandular scales, fenestrated cells, thick-walled cells, agglomerates, hairs, conduits, starch grains, thin-walled cells, hyphae, epithelial cells, reticulocytes, non-glandular hairs, glandular hairs, sporoderms, and corky cells.

In the study of identifying the types of herbal medicines through the microscopic images of herbal powder, the objects in the images are the microscopic features of the herbal medicines, and one type of herbal medicine also contains a variety of microscopic features. Therefore, it is necessary to determine the category of microscopic features first, and then determine the type of Chinese herbal medicine on this basis. Correspondingly, when constructing a dataset of microscopic images of powdered Chinese herbal medicines, there should be two kinds of labels for the dataset. Herb species and microscopic feature categories, corresponding to the 150 types of herbal medicines mentioned above and the combined 20 types of microscopic features. Next, the 63,252 powdered microscopic images of Chinese herbal medicines acquired for constructing the dataset were produced.

Through the process of filming and feature summarization described above, the tasks of image acquisition and label setting in the construction dataset were completed. Next, it was necessary to annotate the 63,252 microimages of powdered Chinese herbal medicines according to the two labels of herb type and microfeature category to complete the construction of the construction dataset.

2.4. Preprocessing of Chinese herbal medicine powder microscopic image data

The experimental data are collected by each research center and herbal medicine institute, but operator differences can lead to the appearance of non-normalized data. Examples include image mutilation, target offset, rotation, etc. There are also problems such as optical malfunctions and excessive impurities due to environmental factors, equipment factors and the preparation process. Therefore, the data should be cleaned before the experiment to fill in the missing problems to correct the inconsistency of the data. At the same time it also enhances the generalization of the data, helps to reduce system operating costs, and also solves the problem of limited dataset size and data variability. Overall improves the overall performance of the model.

In the experiment, bilinear interpolation is performed on some of the images with inconsistent acquisition. Then the dataset is expanded and enhanced by three methods: image geometric transformation, contrast adjustment, and sharpening processing. Finally, the images processed above are normalized to achieve the model processing capability.

1) Bilinear interpolation enlargement. Based on the fact that different magnifications contain different feature information, the process specifies that the shooting objective be set at 50 μm. However, due to differences in collectors and inconsistencies in equipment, some data are collected at 100 μm objectives. In view of the fact that some species and features currently contain less data. Therefore, in the data preprocessing, the pictures that do not meet the requirements are enlarged using bilinear interpolation, so as to make the local information more complete and to unify the scale of the training data to make the model converge more quickly.

Bilinear interpolation amplification, is in x, y two directions on the picture matrix interpolation operation as shown in Figure 1, the figure first in the x direction on the Q_{12} and Q_{22} in the interpolation to complete the R_2 points, Q_{11} and Q_{21} in the interpolation to complete the R_1 points, and then on the R_2 and R_1 for the y direction of the interpolation to get the P points, that is, the whole process to complete the interpolation operation three times.

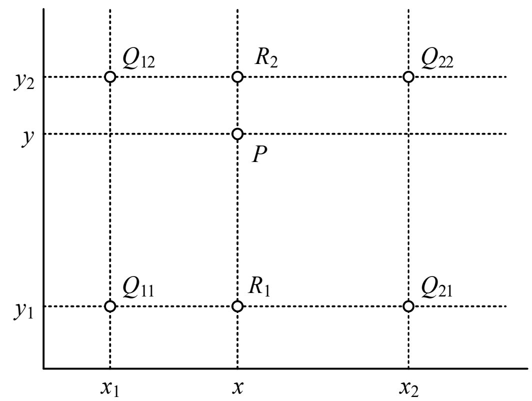


Fig. 1. Double linear interpolation

The formula is:

$$\begin{aligned} f(x,y) \approx & \frac{f(Q_{11})}{(x_2-x_1)(y_2-y_1)}(x_2-x)(y_2-y) \\ & +\frac{f(Q_{21})}{(x_2-x_1)(y_2-y_1)}(x-x_1)(y_2-y) \\ & +\frac{f(Q_{12})}{(x_2-x_1)(y_2-y_1)}(x_2-x)(y-y_1) \\ & +\frac{f(Q_{22})}{(x_2-x_1)(y_2-y_1)}(x-x_1)(y-y_1) \end{aligned} \tag{1}$$

The original image is allowed to zoom in on the target features to a certain extent through three interpolation operations, up-sampling them to a uniform scale, effectively restoring some of the information lost at low magnification.

2) Image contrast adjustment. In order to adjust the effect of different ambient light on the collected data and attenuate the image imbalance problem caused by the photosensitive elements of different sensors. In this paper, the adjustment of contrast is used as one of the ways of data expansion. The adjustment method mainly involves brightness reduction and brightness enhancement of the original image, as shown in Equation (2), which mainly controls the brightness and darkness changes of the original image through parameter β . Where I is the input image matrix and H is the output image matrix. Eq:

$$H(i,j) = I(i,j) + \beta \tag{2}$$

3) Image Sharpening. Some of the data in the dataset are blurred, resulting in serious loss of edge information. This information is of great help to the model in extracting texture features, and at this point it is necessary to experiment in prior image preprocessing to restore the lost edge details. The most classical way to make the target edge steeper is sharpening processing.

Sharpening is mainly achieved by taking derivatives (gradients) or finite differences. For example, the following equation is the derivation of the Laplace operator for image enhancement with second order derivatives. Namely:

$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{3}$$

Differentials are replaced by differentials:

$$\frac{\partial^2 f}{\partial y^2} = [f(i+1,j) - f(i,j)] - [f(i,j) - f(i-1,j)] \tag{4}$$

$$\frac{\partial^2 f}{\partial y^2} = [f(i,j+1) - f(i,j)] - [f(i,j) - f(i,j-1)] \tag{5}$$

Equation (4) is added to equation (5) to obtain equation (6):

$$\begin{aligned} \nabla^2 f(x,y) = f(i+1,j) + f(i-1,j) + f(i,j+1) \\ + f(i,j-1) - 4f(i,j) \end{aligned} \tag{6}$$

The Laplace operator utilizes the second-order derivative information of the image, which can be used to judge the image edge information. Because of its isotropic nature, the gradient result is unchanged after rotating the coordinate axes, which makes it widely used in image processing.

3. Deep learning-based network design for recognition of Chinese herbal medicines

3.1. Description of the problem

Most of the different classes of Chinese medicines have large differences in basic visual features such as color and shape, so traditional CNN networks can achieve good results. However, such methods usually contain a large number of parameters leading to slow recognition speed, which makes it difficult to migrate to mobile devices and thus be used in the daily market. In addition, for some Chinese herbal medicines of the same broad category, due to the small gap in visual features, they are mainly recognized by local features, and strengthening the focus on these features will be able to improve the accuracy of identification. Based on the above factors, this paper designs a lightweight TCM recognition network (Attention-TCM-Net) based on the attention mechanism.

3.2. Deep Convolutional Neural Networks

Traditional convolutional neural networks are mainly composed of three parts: input layer, hidden layer and output layer. Preprocessing operations are performed on the data in the input layer and then fed into the hidden layer. The hidden layer mainly contains four common constructions: convolutional layer, excitation layer, pooling layer and fully connected layer. The convolutional and excitation layers work in tandem to perform layer-by-layer feature extraction on the input data, followed by a pooling layer that downsamples and compresses these extracted features to simplify the dimensionality of the data and retain key information. Eventually, the fully connected layer receives these processed features and makes the final classification decision based on them [20-21].

The convolutional layer is the core structure in a convolutional neural network, and the convolutional operation is essentially a filtering process that is widely used in the fields of signal processing and image processing.

In the field of image processing, this output signal is often referred to as a feature mapping, which can be viewed as an abstraction of the input image signal in a convolutional neural network, and each individual feature mapping can characterize different features extracted by the convolutional neural network from the input image. Therefore, in order to enhance the feature representation capability of the convolutional neural network, a strategy of multiple feature mappings is required in each convolutional layer. This multi-feature mapping approach can effectively enhance the network's ability to capture and process complex features.

Input signals in image processing are usually two or three dimensional, assuming that there is an image signal I and a two or higher dimensional linear system K, which are linearly convolved, the output signal is represented by Eq:

$$I' = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} I(u-i, v-j)\cdot K(i,j) \tag{7}$$

You can also write:

$$I' = I * K \tag{8}$$

The transformation function K is often referred to as the convolution kernel function.

When performing image convolution operation, a series of additional parameters of the convolution layer should be set to ensure the accuracy and diversity of the operation. The convolution kernel size, step size and padding together determine the size of the output feature map, and the size of the output data is denoted as O. The formula for calculating the size of the output data is:

$$O = (W - F + 2P)/S + 1 \tag{9}$$

Convolutional layers occupy an important position in CNN models, which directly affect the network's ability to extract features from the input image. By carefully designing the structural distribution of the convolutional layers and adjusting the parameters of each convolutional unit, the fitting ability of the network can be significantly improved. In shallower convolutional layers, the network is usually able to capture basic image features such as lines, edges and corners. However, as the depth of the network gradually increases, the convolutional layer is able to further extract more complex and abstract features to more fully understand and characterize the image information.

The activation function is the core of the excitation layer, which is a nonlinear function, and the introduction of the activation function at the output of each neuron plays the role of activating the neuron. It introduces a nonlinear relationship between the inputs and outputs of the neurons, thus enabling the network to handle more complex and diverse data patterns.

The pooling layer can also be seen as a special kind of convolutional layer. Common pooling operations include average pooling, maximum pooling, etc. Different pooling algorithms have different convolution kernels, the formula is expressed as:

$$s = \sum_{i=1,j=1}^{c} P_{ij} x_{ij} \tag{10}$$

where P_{ij} is the convolutional kernel parameter, x_{ij} is the input to the pooling layer, and s is the output. When $P_{ij} = \frac{1}{c^2}$ is the average pooling:

$$s_{i,j} = \frac{1}{c^2}\left(\sum_{i=1}^{c}\sum_{j=1}^{c} x_{ij}\right) + b \tag{11}$$

Mean pooling is the process of averaging all input values within the sensory field during the most to each input image's flat sweep and using the result of the computation as a representative value for backward computation.

The fully connected layer is often connected behind the classifier to perform the final classification of the input samples. Among many classifiers, Softmax, with its superior performance and wide range of applications, is the most basic and commonly used choice in convolutional neural networks, especially suitable for solving multi-classification problems. For input sample x, Softmax calculates the conditional probability that it belongs to category c by using the following formula:

$$p(y=c \mid x)=\frac{e^{\omega_c^T x}}{\sum_{j=1}^{c} e^{\omega_j^T x}} \tag{12}$$

where C represents the total number of classification categories, while ω_c represents the weight vector of category c, and p represents the conditional probability that sample x belongs to category c. The Softmax classifier maps the output of the fully-connected layer to the probability values in the interval [0,1] and performs the normalization, which ensures that the probability sums to one.

3.3. Mixed Domain Attention Module Based on Residual Structure

3.3.1. Attention mechanisms. SE Attention Mechanism The most widely used attention mechanism model in CNN models.SE focuses on the relationship between different channels, aiming to allow the model to automatically learn the relative importance of individual channel features. It consists of two main parts, the Squeeze layer and the Excitation layer.The Squeeze layer achieves the effect of aggregating information by performing a global average pooling operation on the feature maps to obtain a one-dimensional vector containing the feature maps, which also facilitates the operation of the Excitation layer.The Excitation layer takes the output of the Squeeze layer as a one-dimensional vector as input. The Excitation layer takes the output of the Squeeze layer as the one-dimensional vector as input. The weights of each channel in each feature map are obtained through the two fully connected layers, and the feature maps from the SE inputs are weighted and used as inputs to the next layer of the network.The flexibility of SE means that it can be applied directly to transformations other than scalar convolution.

BAM utilizes the information carried by the space and the channels to accomplish the selection of the focal features, and BAM infers the attentional feature maps along two different paths, the channels and the space, in parallel. In this work, a simple and lightweight architecture is designed using the channel and spatial dimensions of attention, and an efficient location for placing the BAM is found, which is after the downsampling and before the pooling layer.The BAM builds a hierarchical structure similar to the human perceptual process, where low-level features, such as features of the background texture, are denoised in the early stages of the BAM, and after that, the BAM gradually puts the focus on the precise targets on them, thus achieving more efficient results.

CBAM is given an intermediate feature map, the module sequentially performs feature map processing along two independent dimensions, channel and space, to obtain the attention feature map. The obtained attention feature map is then multiplied with the input feature map thereby performing adaptive feature refinement. Since CBAM is a lightweight and general-purpose module, it can be seamlessly integrated into any CNN architecture with negligible overhead and can be trained end-to-end with the underlying CNN.CBAM can be widely used to improve the representational capabilities of CNNs [22].

3.3.2. CBAM based on hybrid domain attention mechanism. For the problem of recognition and identification of traditional Chinese medicine, due to the strong similarity in the appearance of many similar herbs, the gap between the visual features is small, for example, stove shells, green shells, and pine shells in the micrographic images of traditional Chinese medicine, which are difficult to be distinguished from each other in terms of the appearance of the color, shape, and so on. In order to further strengthen the attention to the key local features and improve the confusing problem in the classification of traditional Chinese medicine, this paper adds the CBAM based on the mixed domain attention mechanism.Meanwhile, on this basis, the residual connection is introduced to increase the transfer of the gradient and the flow of information, so that the features can be reused and the feature loss that may be caused by the weighting process is also avoided.

The CBAM module based on the residual structure firstly performs the convolution operation to get the feature map of the TCM F. Then the attention of the channel direction is carried out by the CAM module. The CAM module firstly extracts the information of the feature map in the spatial dimension, and then the extracted information is used to find the relationship between the feature channels. The expression for this is:

$$\begin{aligned} M_C\left(\hat{F}\right) &= \sigma\left(MLP\left(MaxPool\left(\hat{F}\right)\right) + MLP\left(AvgPool\left(\hat{F}\right)\right)\right) \\ &= \sigma\left(W_1\left(W_0\left(\hat{F}^C_{Max}\right)\right) + W_1\left(W_0\left(\hat{F}^C_{Avg}\right)\right)\right) \end{aligned} \tag{13}$$

where $\hat{F}$ denotes the feature map after convolutional computation. $\hat{F}^C_{Max}$, $\hat{F}^C_{Avg}$ denote the one-dimensional vectors obtained after global maximum pooling and global average pooling, W_0, W_1 denote the parameters of the shared connection layer. σ denotes the Sigmoid activation function, and $M_C(\hat{F})$ denotes the obtained channel attention weights.

Then the extracted feature map $\hat{F}$ and the attention vector $M_C(\hat{F})$ are multiplied in the channel direction, which is computed as follows:

$$F' = \hat{F} \otimes M_c(\hat{F}) \tag{14}$$

where $\otimes$ represents the channel-by-element multiplication by the channel attention weights, and F' represents the feature map weighted by the channel attention. The weight vector in the spatial direction is generated after the activation function, which is calculated as follows:

$$\begin{aligned} M_s\left(F'\right) &= \sigma\left(f^{7\times7}\left(\left[MaxPool\left(F'\right); AvgPool\left(F'\right)\right]\right)\right) \\ &= \sigma\left(f^{7\times7}\left(\left[F'^s_{Max}; F'^s_{Max}\right]\right)\right) \end{aligned} \tag{15}$$

where F'^s_{Max}, F'^s_{Max} denote the feature maps obtained after global maximum pooling and global average pooling, respectively. $f^{7\times7}$ denotes the convolutional computation using a convolutional kernel of size 7×7. $M_s\left(F'\right)$ denotes the obtained spatial attention weights.

The input feature map F' and attention vector $M_s\left(F'\right)$ are multiplied in the same spatial direction and then added to the constant mapping, which is computed as follows:

$$F' = F' \otimes M_s\left(F'\right) + F \tag{16}$$

where $\otimes$ denotes the spatial attention weights multiplied element-by-element by different channels at spatial locations. F'' denotes the final module output. By incorporating the spatial attention mechanism, it allows the network to better learn the location information of the features to enhance the representation of the features.

3.4. Improved moving inverted bottleneck convolution module

In this paper, a mobile inverted bottleneck convolution (MBConv) is proposed. The method improves on the traditional residual block by inverting it to alleviate the network degradation problem arising from deeply separable convolution, while improving the classification accuracy and memory utilization efficiency.

In this paper, we improve on the MBConv module. The channel-attention SE module and spatial-attention SAM module are added to the two stages of MBConv to optimize the feature extraction performance and strengthen the attention to the microscopic features of traditional Chinese medicine (TCM) as a way to improve the accuracy of TCM feature classification.

3.5. Network architecture

EfficientNet uses a composite factor of $\varnothing$ to scale the three dimensions of network depth, width and image resolution as shown below:

$$\begin{gathered} d=\alpha^{\phi}, w=\beta^{\phi}, r=\gamma^{\phi} \\ s.t. \alpha \cdot \beta^{2} \cdot \gamma^{2} \approx 2 \\ \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \end{gathered} \tag{17}$$

where, d , w, r denote the scaling factors of the three dimensions, which are used to adjust the size of the three. α , β , γ are constants obtained by EfficientNet using NAS technology for continuous combinatorial optimization, indicating the resource allocation parameters of the corresponding dimensions. $\varnothing$ is the control parameter of the resources, which is set by the size of the resources available to the model.

Based on the consideration of lightweight and efficiency, the convolutional stacking in the middle of the Attention-TCM-Net model in this paper refers to the design of EfficientNet-B0 with the least number of parameters. The structure of the network is designed as follows:

Step1: Initial feature extraction of TCM is carried out by the 3×3 convolutional kernel in the L1 module.

Step2: A hybrid domain attention mechanism CBAM based on residual structure is introduced in the L2 structure to enhance the attention to the features of Chinese medicine on the basis of ensuring the flow of information.

Step3: L3-L9 modules are stacked for MBConv, and the improved MBConv is used in this paper.

Step4: The L10 module utilizes a size 1×1 convolutional kernel for channel number increase. Then global average pooling is used instead of the operation of directly flattening the feature mapping in traditional convolutional neural networks, which reduces the number of over-dense parameters in the fully connected layer and improves the training speed. Finally, a fully connected layer with a node number of 20 is passed through to complete the classification of 20 classes of Chinese herbal medicine microscopic images. The single-layer fully-connected design reduces the training parameters, makes the model easier to converge, and makes the structure simpler as well.

4. Effectiveness of identification of micro-morphological characteristics of Chinese herbal medicines

4.1. Experimental basis

4.1.1. Selection of development framework. After more than a decade of development deep learning has made great progress, during which a lot of mature development frameworks have appeared, among which the more commonly used are Caffe, TensorFlow and PyTorch.

PyTorch is a framework written in the Python programming language, which was introduced on the basis of Torch to make up for the shortcomings of the low usage of the Lua language. Its structure is simple and standardized and contains three levels of abstraction from low to high. It introduces automatic derivation on the basis of Torch, which enhances the practicality of the framework. Compared with TensorFlow, Pytorch adds the dynamic computational graph function, which possesses stronger flexibility and lighter volume. Compared with Caffe, Pytorch has rich community resources and deep learning libraries, which reduces the difficulty of designing new network layers on top of it. In addition, Pytorch is able to provide better support for GPUs, offering multi-process concurrent libraries with shared memory to accelerate model training efficiency.

After comparing the advantages and shortcomings of different deep learning frameworks, this paper concludes that PyTorch is more suitable for the development of the Attention-TCM-Net model of Chinese herbal medicine micro-image recognition, therefore, all the Attention-TCM-Net network models in the subsequent study are built and trained based on the PyTorch framework.

4.1.2. Experimental environment. As mentioned earlier, PyTorch was chosen as the development framework for the Attention-TCM-Net model in this paper, and the model was built in the Anaconda3 integrated environment. And the network model was trained, tested and validated in the same hardware and software environment. The experimental hardware environment is shown in Table 1.

Table 1. Hardware environment

Name	Parameter configuration
Processor	Intel(R) Core(TM) i7-10750H CPU@3.40GHz (12 CPUs), ~2.6GHz
Memory	32GB DDR4
Graphics card	NVIDIA GeForce RTX 2080 8GB
Hard disk	PCIe SSD

The software environment is shown in Table 2.

Table 2. Software environment

Name	Information
Operating system	Windows 11
Depth learning framework	PyTorch
CUDA	CUDA:10.0 cuDNN:7.4.1
Anaconda	Anaconda3
Python	3.6.13

4.2. Impact of Attention Mechanisms on Modeling

In order to verify the effect of the attention mechanism on the performance of the Attention-TCM-Net model, a dataset of powdered micrographic images of Chinese herbal medicines is recognized by loading the parameters of the pre-trained network model and adding a hybrid attention mechanism after the residual structure using a transfer learning method, while other parameter conditions remain unchanged. The accuracy of the training and validation sets with the hybrid attention mechanism is shown in Figure 2.

The added hybrid attention mechanism can locate the useful features of the Chinese herbal medicine powder micrographic images faster, and can adaptively adjust the attention weight information between different parts to obtain effective weight information, which enables the Attention-TCM-Net model to better understand the input data, enhances the expressive ability of the Attention-TCM-Net network model, and reduces the influence of noise on the experimental results. The effect of noise on the experimental results, weaken the invalid features of the data samples, and improve the effect of recognizing the powder micro image of Chinese herbal medicine, the accuracy rate is as high as 0.985. And the training loss is lower than the loss of the validation set, and the curve tends to be smoothed. It performs well on the Chinese herbal medicine microscopic image dataset and improves the performance of Attention-TCM-Net network model.

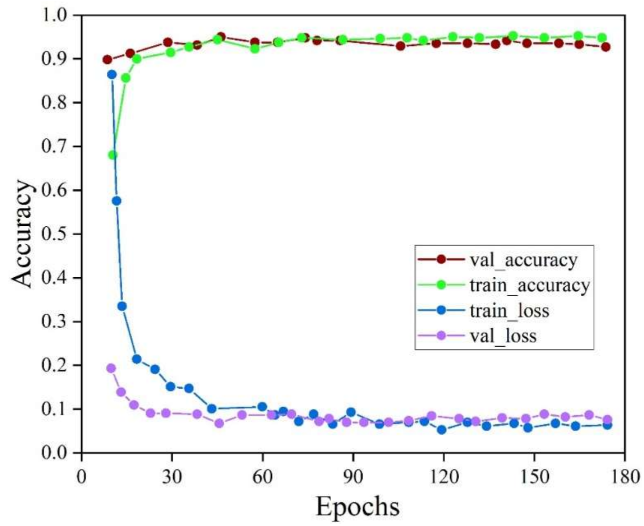


Fig. 2. The accuracy of the mixing attention mechanism

4.3. Analysis of experimental results of different models

In order to evaluate the performance of the proposed Chinese herbal medicine powder micro-recognition model, this chapter compares the Attention-TCM-Net network model with some representative convolutional neural networks including ResNet34, DenseNet121, AlexNet and VGG16 under the same parameter settings, and the experimental results are shown in Table 3.

The Attention-TCM-Net network model achieves an average accuracy of 96.47% on the test set, which is an increase of 0.85 percentage points compared to the ResNet34 network.

As can be seen from the tabular data, the Attention-TCM-Net model has the highest recognition accuracy in herbal powder micrographic image recognition compared to other network models. It is cost effective and competitive in terms of performance and storage cost.

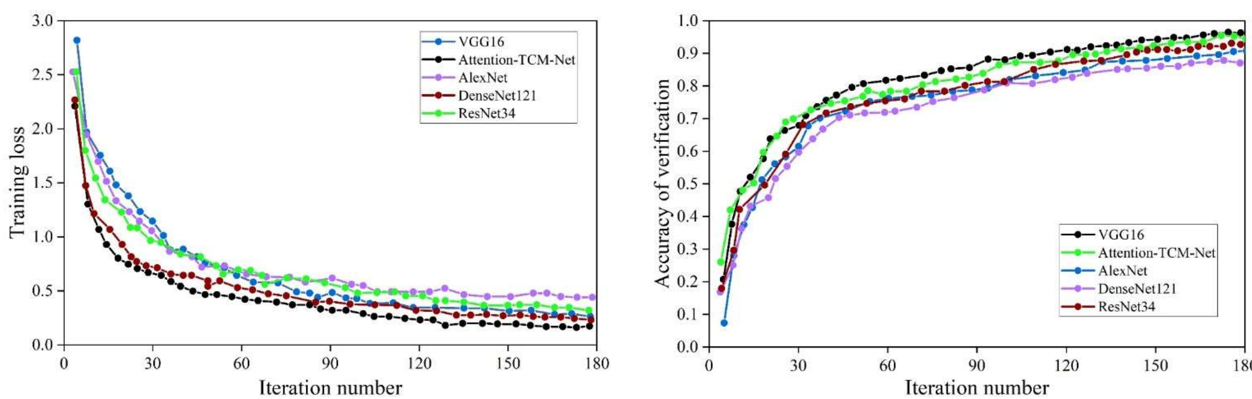
Table 3. Performance comparison of different models

Model	Average/%				Parametric memory / MB	FLOPs	FPS
	Accuracy	Accuracy ratio	Check rate	F1-Score			
ResNet34	95.62	95.32	95.66	94.36	85.36	2.85E+09	152.36
AlexNet	89.21	89.36	89.63	89.63	362.11	5.69E+08	417.43
DenseNet121	93.21	93.47	91.24	93.42	32.99	2.15E+09	53.62
VGG16	90.26	91.05	90.63	90.07	623.52	1.27E+10	129.54
Attention-TCM-Net	96.47	96.29	96.15	96.22	82.33	4.33E+0.9	103.25

The training loss and verification accuracy versus the number of iterations are shown in Fig. 3.

Figure (a) gives the training loss versus iteration curves for the five models. It can be clearly seen that the training loss of each model is stabilized after 180 iterations, and there is no large fluctuation. From the loss convergence, it can be seen that the SE-ResNet36 model converges significantly better than the other models.

The validation accuracy of each model versus iteration is shown in Fig. (b), and compared with AlexNet, DenseNet121, VGG16, and the original model ResNet34, it can be clearly seen that the validation accuracy of the improved model Attention-TCM-Net rises quickly and smoothly.



(a) Training loss

(b) Accuracy of verification

Fig. 3. Training loss and validation accuracy

In order to demonstrate more intuitively the recognition effect of the five network models on the microscopic images of powdered Chinese herbal medicines of different categories, three-dimensional bar charts were drawn based on the detection rate of different models on the microscopic images of Chinese herbal medicines of each category, and the detection rate of different models on each category of images is shown in Fig. 4. Different colors were used to classify the detection rates of different models.

The detection rate of AlexNet model on the fourth class of Chinese herbal medicine micro-morphological features is only 24.32%, and the detection rate of DenseNet121 model on the micro-morphological features of this class of Chinese herbal medicine is not more than 60%, which indicates that the micro-morphological features of this class of Chinese herbal medicine are not obvious, and it is necessary to improve the recognition performance of the model. The Attention-TCM-Net model designed in this paper has a check rate of 85.25% for the recognition of micro-morphological features of this type of herbs, which is 60.93% and 31.99% higher than the AlexNet model and the DenseNet121 model, respectively. From the figure, it can be seen that the Attention-TCM-Net model is significantly better than the other models for the overall recognition of micro-morphological feature images of different Chinese medicines.

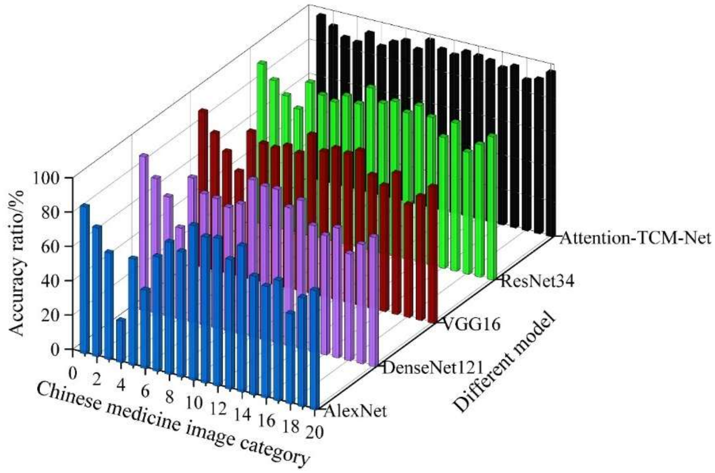


Fig. 4. The accuracy of different models on each type of image

The prediction results of micromorphological features of each type of herbal medicines can be seen at a glance by using the confusion matrix.The confusion matrix of Attention-TCM-Net model on the test set is shown in Fig. 5.

By analyzing the confusion matrix of Attention-TCM-Net network given in the figure it can be found that the image dataset of micro-morphological features of Chinese herbal medicines found that 5 samples of turmeric with label 4 were predicted as red ginseng with label 15.

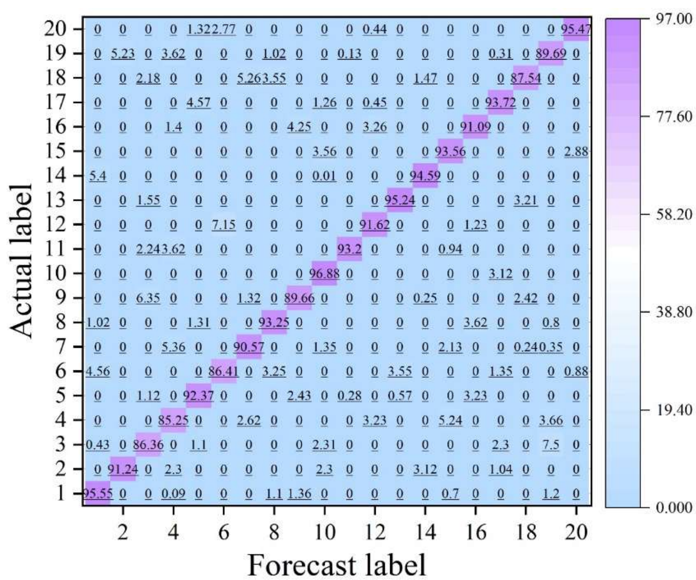


Fig. 5. The attenon tcm-net model is a confusion matrix on the test set

4.4. *Characterization of various types of Chinese herbal medicines*

The detection results of the improved model are shown in Fig. 6, which gives the detection results of the optimized model for the micro-morphological features of Chinese herbal medicines in all categories. Precision rate Precision, recall rate Recall, average recognition precision AP of each herb and average recognition precision mAP of micro-morphological features of all categories of Chinese herbal medicines are selected as the evaluation indexes of the model.

The improved image recognition model of Chinese herbal medicines, i.e. Attention-TCM-Net model, can effectively recognize and classify the micro-morphological features of Chinese herbal medicines. It can be found through comparison that the Attention-TCM-Net model has the best overall recognition effect for the categories of sand nut, lutong, nutmeg, and raspberry. It has the highest average recognition rate of 99.63% for the raspberry category, which achieves the best detection effect, which is attributed to the fact that the microscopic images of this category of Chinese herbs have obvious image features. The same high average detection accuracy was achieved for the category of Saxifrage, Lutetia, and Nutmeg, which were able to achieve better detection results. The average detection accuracies of Crow's gall, Sichuan pei, and Ullemi chinensis were lower than those of the other categories, which was attributed to the small size of these three herbs and their inconspicuous shape features, which increased the detection difficulty of the model.

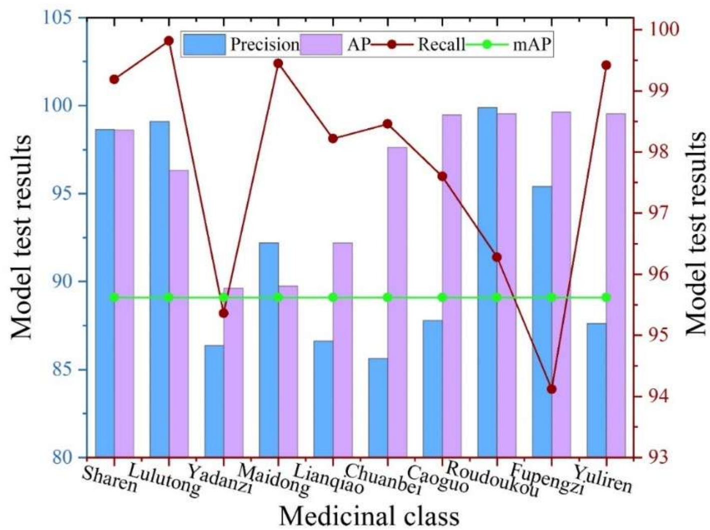


Fig. 6. Improved model test results

4.5. Herbal Characterization Measurements

In order to verify the detection ability of the improved model (Attention-TCM-Net) in the real environment, this chapter uses the micrographic image recognition model of Chinese herbal medicines to test the single-category herbs and the multi-category mixed herbs, respectively, and the test samples are purchased from a large pharmacy in a city.

4.5.1. Single class of herbal medicines. A total of ten categories of Chinese herbs were used for testing in this experiment, and five microscopic images with different shooting conditions were selected for each herb for experimental testing.

The detection results of single-category Chinese herbs are shown in Table 4, which shows that the microscopic image recognition model of Chinese herbs based on the improved convolutional neural network (Attention-TCM-Net model) is able to effectively detect and recognize most of the herbs under both normal and dim ambient light conditions. Among them, LuLuTong, ShaRen, Nutmeg, CaoGuo and MaiDong, because of the obvious appearance characteristics, are able to achieve a high level of detection, in which the average detection result of LuLuTong is 0.992. And it can be seen from the result data that the model is still able to maintain a better detection effect in the case of the image background complexity.

Table 4. The results of the single category of Chinese medicine

Medicinal class	Result 1	Result 2	Result 3	Result 4	Result 5
Lulutong	1.00	1.00	0.99	0.99	0.98
Sharen	1.00	0.99	0.99	1.00	0.99
Roudoukou	1.00	0.99	0.98	0.97	0.96

Caoguo	1.00	1.00	0.95	0.95	1.00
Fupenzi	1.00	1.00	0.93	0.93	0.95
Lianqiao	0.99	0.95	1.00	0.94	1.00
Maidong	0.99	0.97	1.00	0.99	0.96
Chuanbei	1.00	0.99	1.00	1.00	0.93
Yuliren	1.00	0.98	0.97	0.99	0.98
Yadanzi	1.00	0.93	0.99	0.95	0.99

4.5.2. Mixed testing of multiple classes of herbs. The data of mixed detection of multi-category Chinese herbs are that ten different kinds of Chinese herbs are placed in the environment with different brightness, background color and texture for testing, to test the detection ability of the image recognition model of Chinese herbs in practical application scenarios.

The results of the mixed detection of multiple categories of Chinese herbs are shown in Fig. 7, in which the average recognition accuracies of lutong, chuanbei, saren, forsythia, maitake and nutmeg are all above 95%, which can realize accurate recognition and classification. And the average detection accuracy of grass fruit and raspberry can also reach more than 90%.The overall recognition accuracy of the Attention-TCM-Net model for the micro-morphological features of multiple types of Chinese herbs is 95.03%, which proves that the model has the same good detection ability and robustness in the multi-classification task under the complex background.

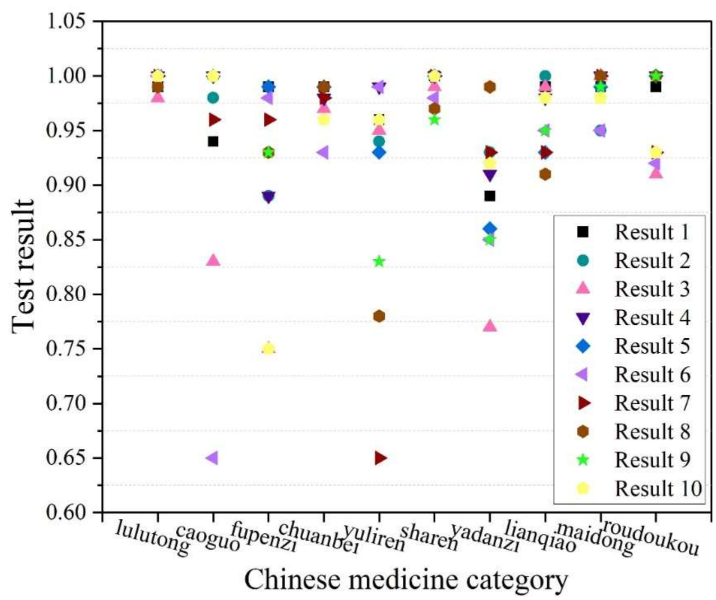


Fig. 7. The results of mixed testing of Chinese medicinal materials

5. Conclusion

In this paper, for the problem of slow recognition speed of traditional CNN network, it is proposed to add the attention mechanism on the basis of traditional CNN network for optimization, and designed the Attention-TCM-Net based on the attention mechanism for the recognition of micro-morphological features of traditional Chinese medicinal herbs. The algorithmic network is utilized to recognize the constructed dataset of powdered microscopic images of commonly used Chinese medicinal herbs.

1) Setting the algorithm model with other parameters unchanged, the influence of the hybrid attention mechanism was analyzed. Adding the hybrid attention module to the convolutional neural network can accelerate the algorithm's localization of the micro-morphological features of Chinese herbal medicines, prompting the model to better understand the image data and enhancing the

expressive ability of the Attention-TCM-Net network model. The accuracy of the Attention-TCM-Net network based on the Attention mechanism for recognizing the microscopic features of Chinese herbal medicines reaches 98.5%.

2) Comparing the performance of Attention-TCM-Net network model with ResNet34, DenseNet121, AlexNet and VGG16, the average accuracy of Attention-TCM-Net network model on the test set reaches 96.47%, and it has a comparative advantage in terms of performance and storage cost. The Attention-TCM-Net network model achieves excellent performance in different categories of micro-feature recognition of Chinese medicinal herbs, thus demonstrating that the improved micro-feature recognition network based on convolutional neural network designed and proposed in this paper has practical value.

Funding

This research was funded by the institution of Middle-aged and Young Teachers' Basic Ability Promotion Project of Guangxi (Grant No.2021KY0802); Guangxi Nature Science Foundations (Grant No.2024GXNSFAA010338); and GUAT Special Research on the Strategic Development of Distinctive Interdisciplinary Fields (Grant No.TS2024121).

About the Authors

Hanying Wang (born in 1992), native place (Jining, Shandong), female, master's degree, lecturer. Her main research fields are machine learning and image processing.

Zhi Chen (born in 1985), native place (Guilin, Guangxi), male, master's degree, associate professor. His main research fields are intelligent algorithms and control of robots.

Jiabo Huo (born in 2009), native place (Xingtai, Hebei), male, master's degree, senior engineer. His main research field is aerodynamics.

Xingguo Han (born in 1982), native place (Laizhou, Shandong), male, master's degree, professor. His main research field is intelligent algorithms and control of robots.

References

[1] Xue, C. C., & O'Brien, K. A. (2016). Modalities of Chinese medicine. In A comprehensive guide to Chinese medicine (pp. 1-28).

[2] Zhu, Y. P. (2024). Chinese materia medica: chemistry, pharmacology and applications. CRC press.

[3] Ma, D., Wang, S., Shi, Y., Ni, S., Tang, M., & Xu, A. (2021). The development of traditional Chinese medicine. Journal of Traditional Chinese Medical Sciences, 8, S1-S9.

[4] Matos, L. C., Machado, J. P., Monteiro, F. J., & Greten, H. J. (2021, March). Understanding traditional Chinese medicine therapeutics: an overview of the basics and clinical applications. In Healthcare (Vol. 9, No. 3, p. 257). MDPI.

[5] Yang, X., Tian, X., Zhou, Y., Liu, Y., Li, X., Lu, T., ... & He, L. (2018). Evidence-based study to compare Daodi traditional Chinese medicinal material and non-Daodi traditional Chinese medicinal material. Evidence-Based Complementary and Alternative Medicine, 2018(1), 6763130.

[6] Luo, H., Zhao, Y., Hua, H., Zhang, Y., Zhang, X., Fang, Q., ... & Zhao, J. (2021). Research progress on quality assurance of genuine Chinese medicinal in Sichuan. Chinese Medicine, 16, 1-13.

[7] He, J., Feng, X., Wang, K., Liu, C., & Qiu, F. (2018). Discovery and identification of quality markers of Chinese medicine based on pharmacokinetic analysis. Phytomedicine, 44, 182-186.

[8] Wang, J., Li, Y. S., & Wang, Q. (2019). Identification of Chinese medicine constitution in public health services. Chinese Journal of Integrative Medicine, 25(7), 550-553.

[9] Chen, R. R., & Chen, Y. J. (2019, July). Intelligent Identification of Traditional Chinese Medicine Materials Based on Multi-feature Extraction and Pattern Recognition. In 2019 International Conference on Modeling, Analysis, Simulation Technologies and Applications (MASTA 2019) (pp. 390-395). Atlantis Press.

[10] Xu, A., Wang, T., Yang, T., Xiao, H. A. N., Zhang, X., Wang, Z., ... & Hu, K. (2024). Constitution identification model in traditional Chinese medicine based on multiple features. Digital Chinese Medicine, 7(2), 108-119.

[11] Ming, J., Chen, L., Cao, Y., Yu, C., Huang, B. S., & Chen, K. L. (2019). Rapid identification of nine easily confused mineral traditional Chinese medicines using Raman spectroscopy based on support vector machine. Journal of Spectroscopy, 2019(1), 6967984.

[12] Bing, W., Yan-Tao, S., Wen-Han, P., Hui, Z., & Ting-Guo, K. (2022). A new method for studying the mechanism of “Feature Identification based quality assessment” of Traditional Chinese Medicine, taking Gastrodiae Rhizoma as an example. Heliyon, 8(8).

[13] Yue, J., Zuo, Z., Huang, H., & Wang, Y. (2021). Application of identification and evaluation techniques for ethnobotanical medicinal plant of genus Panax: A review. Critical Reviews in Analytical Chemistry, 51(4), 373-398.

[14] Hu, Z., Liu, X., Chen, Z., & Zhang, P. (2024, July). Identification of Chinese medicinal materials based on Support Vector Machine. In Proceedings of the 2024 International Conference on Cloud Computing and Big Data (pp. 465-469).

[15] Tian, Z. (2022, November). Application of mathematical modeling in" identification and analysis of traditional Chinese medicine". In International Conference on Advanced Sensing and Smart Manufacturing (ASSM 2022) (Vol. 12351, pp. 405-410). SPIE.

[16] Zou, H. (2017). Two Chinese medicine species constants and the accurate identification of Chinese medicines. bioRxiv, 166140.

[17] Jiang, Y., Wei, S., Ge, H., Zhang, Y., Wang, H., Wen, X., ... & Li, P. (2025). Advances in the Identification Methods of Food-Medicine Homologous Herbal Materials. Foods, 14(4), 608.

[18] Ren, L., Guo, M. Y., & Pang, X. H. (2018). Identification and classification of medicinal plants in Epimedium. Chinese Herbal Medicines, 10(3), 249-254.

[19] Song, Y., Zhao, B., Jia, J., Wang, X., Xu, S., Li, Z., & Fang, X. (2021). A review on different kinds of artificial intelligence solutions in TCM syndrome differentiation application. Evidence-Based Complementary and Alternative Medicine, 2021(1), 6654545.

[20] Esmail Jafari,Amin Zarei,Habibollah Dadgar,Ahmad Keshavarz,Reyhaneh Manafi Farid,Habib Rostami & Majid Assadi. (2024). A convolutional neural network-based system for fully automatic segmentation of whole-body [68Ga] Ga-PSMA PET images in prostate cancer.. European journal of nuclear medicine and molecular imaging(5),1476-1487.

[21] Biren Khimji Patel,Leonardo Tariciotti,Lorenzo DiRocco,Antonio Mandile,Samir Lohana,Alejandra Rodas... & Gustavo Pradilla. (2025). Regression modeling with convolutional neural network for predicting extent of resection from preoperative MRI in giant pituitary adenomas: a pilot study.. Journal of neurosurgery11-10.

[22] Kashif Fareed, Anas Khan, Musaed Alhussein, Khursheed Aurangzeb, Aamir Shahzad & MazharIslam. (2025). CBAM Attention Gate-Based Lightweight Deep Neural Network Model for Improved Retinal Vessel Segmentation. International Journal of Imaging Systems and Technology(1),e70031-e70031.