

Hyperspectral Information Detection and Global-view Net for Enhanced Classification of Mold Stages in Cigarette Tobacco

Jinxia Liu,^{1,2} Qiaoyu Zhang,² Zhe Jin,¹ Feng Li,¹ Yan Shi,² and Hong Men^{2*}

¹Technical Research and Development Center, Jilin Tobacco Industrial Co., Ltd., Changchun 130000, China

²School of Automation Engineering, Northeast Electric Power University, Jilin 132012, China

(Received August 1, 2024; accepted December 26, 2024)

Keywords: cigarette tobacco, hyperspectral technology, attention mechanism, mold detection

The quality and flavor of tobacco are of paramount importance, and mold can significantly degrade these attributes. To address this issue, we utilize hyperspectral technology to capture the spectral information of cigarette tobacco across three distinct mold stages: nonmoldy, near-moldy, and moldy. We introduce a novel global-view (GV) attention mechanism, which leverages both 3×3 local spatial convolution and global attention to integrate global features effectively. Subsequently, the GV-Net model is designed and implemented for the classification of tobacco mold levels. The model achieves an accuracy of 94.39%, precision of 95.02%, recall of 95.42%, and F1 score of 0.9516. These metrics surpass those of multiple existing attention mechanisms, demonstrating our model's superior generalization capabilities. The findings of this study not only facilitate the rapid detection and management of mold in tobacco by manufacturers but also help maintain product quality and consumer trust, underscoring our model's significant practical value and application potential.

1. Introduction

The tobacco industry represents a substantial sector of China's national economy. However, the issue of mold has become a significant economic burden during the production and distribution of tobacco products. It is estimated that tobacco losses due to mold are equivalent to approximately 1–3% of total sales, with an estimated 11.3% of finished cigarettes sold in the market being affected by mold.⁽¹⁾ The hygroscopic nature of tobacco and its nutrient content provide an optimal environment for mold growth, particularly in circumstances where the warehouse environment is not effectively regulated.⁽²⁾ It is therefore imperative that a rapid and accurate quality inspection of tobacco be implemented at every stage of the production and distribution process, in order to guarantee product quality and ensure compliance with the requisite quality standards.

Moldy tobacco not only affects the flavor and quality of tobacco, but more importantly, the mycotoxins produced by mold, such as aflatoxins, may pose a serious threat to human health. Inhalation of smoke containing mycotoxins can increase the risk of pneumonia and other

*Corresponding author: e-mail: menhong@neepu.edu.cn
<https://doi.org/10.18494/SAM5266>

respiratory diseases, and long-term exposure may lead to more severe health issues. Currently, the detection of mold in tobacco leaves and tobacco cuttings primarily relies on visual inspection and sensory evaluation methods.⁽³⁾ However, these methods are not only inefficient but also highly subjective, leading to potential misjudgments. They are insufficient to meet the quality control needs of modern tobacco production, and sensory evaluation can also compromise the health of the assessors. To address these issues, there is a need for objective and reliable alternative methods for assessing tobacco mold. With the advancement of spectral technologies, hyperspectral imaging has gained increasing application in quality assessment owing to its noninvasive and high-efficiency characteristics.

Hyperspectral imaging is a three-dimensional detection technology that is widely employed in the fields of crop nutrient detection,^(4–6) pest and disease detection,^(7–9) and agricultural product quality inspection.^(10–12) Hyperspectral imaging is achieved by combining image information with spectral information. Yang *et al.* used hyperspectral imaging combined with a depth-stacked sparse self-coding algorithm to identify the degree of mold of maize infected by a fungus during the storage period.⁽¹³⁾ In their study, Jia *et al.* employed hyperspectral images of maize seeds exhibiting varying degrees of mold as a data source. They then proceeded to select a combination of standard normal variables (SNVs) and proposed uninformative variables for the processing of these images, ultimately establishing an ant colony optimization back propagation neural network (BP) classification model.⁽¹⁴⁾ Xue *et al.* constructed 12 prediction models based on seven preprocessing techniques and three algorithms for fast and nondestructive detection using hyperspectral images of maize seeds with different water contents.⁽¹⁵⁾ Yang *et al.* introduced a new nondestructive technique for determining the flavor value of rice by combining hyperspectral images with the particle swarm optimization support vector machine (PSO-SVM) algorithm.⁽¹⁶⁾ Dhakal *et al.* correlated hyperspectral images with the mycotoxin deoxynivalenol (DON) content to achieve the classification of wheat grain samples with different levels of damage.⁽¹⁷⁾ Jiang *et al.* used a hyperspectral imaging system with visible-near-infrared spectroscopy to identify three different natural mold degrees in oil tea fruits,⁽¹⁸⁾ demonstrating that hyperspectral technology provides a fast and nondestructive technique for identifying mold degrees.

However, existing hyperspectral analysis methods largely rely on traditional machine learning algorithms. Although these methods have achieved preliminary success in detecting mold in certain crops, they struggle with complex spectral data, which is susceptible to noise and data imbalance, leading to decreased recognition accuracy. Additionally, these methods often fail to effectively utilize the global information contained in hyperspectral data, resulting in poor performance during the detection of complex backgrounds and early stages of mold. Therefore, the global-view (GV)-Net model, based on a GV perspective, can effectively overcome the limitations of existing methods, enhancing the accuracy and robustness of mold detection.

In this study, we utilize hyperspectral imaging technology to acquire spectral information of cigarette tobacco at various stages of mold infestation and combine the GV attention mechanism with neural networks to construct GV-Net, a lightweight network for the nondestructive identification of tobacco quality across different mold levels. The results of this research not

only contribute to enhancing the overall quality of tobacco products, but also provide a way to effectively reduce the risk of moldy tobacco entering the market, thereby protecting consumers from the potential health hazards of mold-infested tobacco.

2. Materials and Methods

2.1 Sample preparation

To obtain mold samples of cigarette tobacco with three different degrees of mold, namely, nonmoldy, near-moldy, and moldy, the following experimental conditions were determined from the pre-experiment: 60 g of commercially available finished cigarette tobacco was placed in a constant-temperature and constant-humidity chamber at 25 ± 1 °C and $85 \pm 1\%$ RH for 22 days. The sample collection process was as follows.

- 1) The initial sampling was undertaken prior to the placement of the samples within the constant-temperature and constant-humidity chamber (day 0), with the subsequent sampling conducted on day 1.
- 2) Sampling was conducted every two days from days 2 to 18, comprising the 3rd to 11th samplings.
- 3) From the 19th day to the 22nd day, sampling was conducted once a day, comprising the 12th to 15th samples.

This method enabled the effective capture of the changes in cigarette tobacco at different mold stages, resulting in the acquisition of 450 sets of cigarette tobacco samples with various levels of mold.

2.2 Detection systems and methods

In this experiment, a GaiaSorter Hyperspectral System (Beijing Zollihan Optoelectronic Instrument Co., Ltd.) is used to obtain the spectral information of cigarette tobacco. The testing principle is illustrated in Fig. 1. The spectral system comprises a spectral camera, a uniform light source, a powered mobile platform, control software (SpecView), and a computer. The cigarette tobacco sample is situated at the center of the motorized mobile platform. The spectral camera captures images by recording the reflected light from the cigarette tobacco samples, which are illuminated by a tungsten bromine lamp. In this study, the spectral range of the hyperspectral system was from 380 to 1038 nm, encompassing 520 bands. The spectral camera possesses 1344×1024 pixels, thereby providing a high spectral resolution of 1.26 nm.

Following the preliminary experimentation to ascertain the requisite parameters, the hyperspectral system detection experiment was conducted. The specific steps are as follows.

- 1) Initialization of System Parameters: The camera exposure time is 10 ms, the speed of the moving platform is 0.28 cm/s, and the distance between the cigarette tobacco sample and the detection platform is 10 cm.
- 2) Subsequently, black and white correction is employed to mitigate the impacts of external illumination, sensor sensitivity, and camera parameters on the original image. The calculation formula for black and white correction is

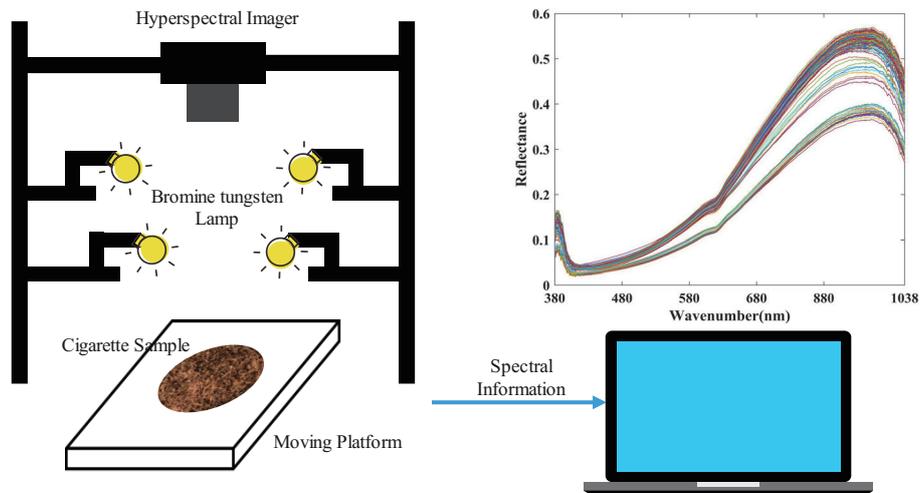


Fig. 1. (Color online) Schematic of hyperspectral system test.

$$I = \frac{I_0 - I_D}{I_W - I_D}, \quad (1)$$

where I is the black and white corrected spectral image, I_0 is the original cigarette spectral image, I_D is the standard all-black reference image, and I_W is the standard all-white reference image.

- 3) The cigarette tobacco sample is placed in the center of the moving platform and passed through the shooting area of the spectral camera at a uniform speed in order to obtain the spectral image. Once the spectral image has been obtained, the information acquisition process is completed upon the completion of black and white correction processing.
- 4) For the sake of generality, the aforementioned steps should be repeated. A total of 450 samples were obtained from cigarette tobacco with three different degrees of mold.

The use of hyperspectral images is constrained by the sheer volume of data they contain and the processing difficulties this entails. Additionally, the images comprise two distinct regions, namely, the sample and the background, which further complicates the data processing. Accordingly, we specify a region of interest in the spectral image to ensure the spatial information of cigarette tobacco samples while reducing the complexity of the processing. Following parameter adjustment and testing, a square of 240×240 pixels was selected as the region of interest. The region's average pixel value was then calculated as the final hyperspectral information. As each sample possesses 520 distinct bands, the spectral information of each sample is 1×520 , while the spectral information of 450 groups of samples is 450×520 .

2.3 Data processing

2.3.1 Spectral information preprocessing

Hyperspectral imaging technology enables the acquisition of continuous spectral information of an object or scene, which is then integrated with spatial data to create a pixel point that contains both spatial and spectral information. Consequently, hyperspectral data are typically of high dimensionality and complexity, necessitating sophisticated data processing and information extraction techniques. A considerable amount of noise is present in the spectral information owing to factors such as light scattering, system noise, and dark current. Accordingly, we employ data preprocessing techniques to enhance the efficacy of the data. The raw spectral information is processed by the Savitzky–Golay (S–G) method, as detailed by Eq. (2), to enhance the signal-to-noise ratio and increase the smoothness of the data. The SNV is employed to eliminate the noise resulting from surface scattering and solid particles, as described by Eq. (3). Multiplicative scatter correction (MSC) was introduced to eliminate the spectral differences resulting from disparate scattering levels and to rectify the baseline translation and offset phenomenon observed in spectral data, as described by Eq. (4). As an illustrative example, the results of processing the average spectral information for the unmolded cigarette tobacco sample are presented in Fig. 2. The identification of spectral information from a limited number of

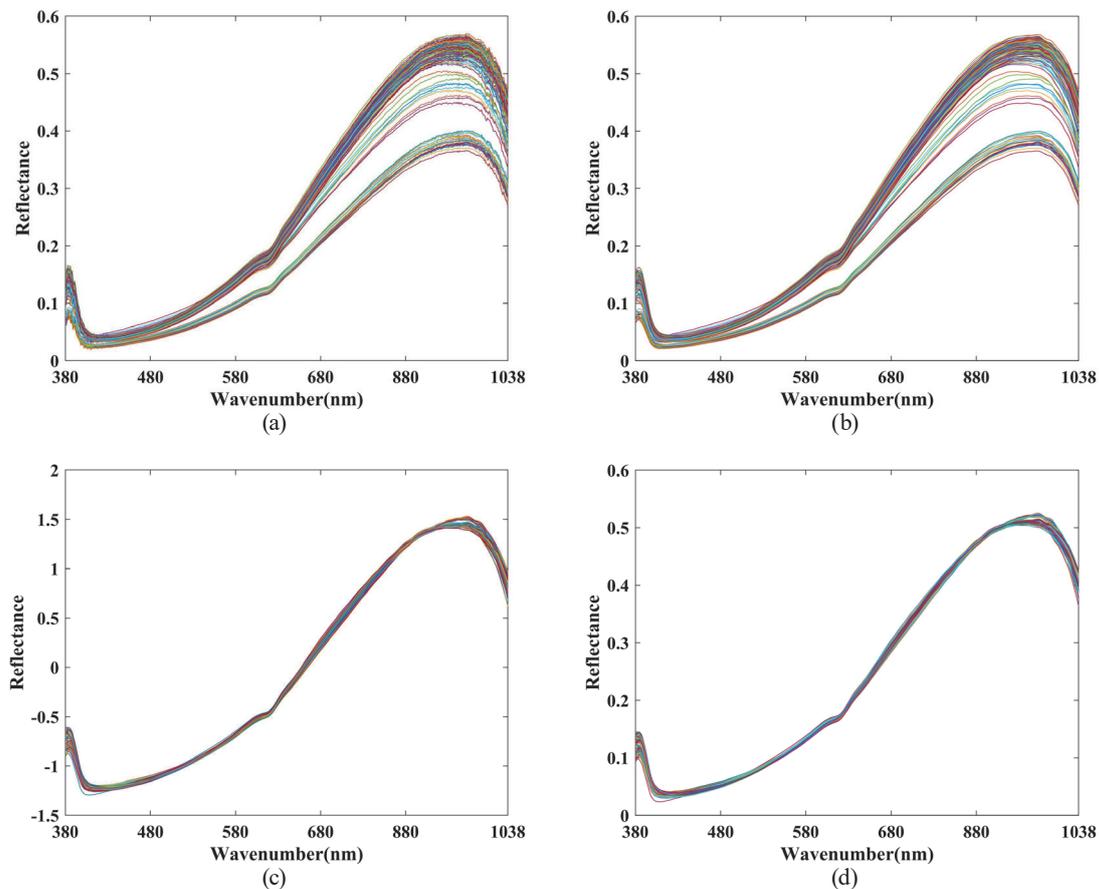


Fig. 2. (Color online) Processing results of average spectral information: (a) raw data, (b) SG processing result, (c) SNV processing result, and (d) MSC processing result.

bands may result in a reduction in the capacity for generalization. Consequently, we propose that spectral information be selected from the entire frequency range, which can enhance the classification performance of spectral data.

$$y'_i = \frac{1}{n} \sum_{j=-m}^m c_j \cdot y_{i+j} \quad (2)$$

Here, y'_i is the filtered spectral value, y_{i+j} is the original spectral value, c_j is the precomputed convolution coefficient, and n is the width of the convolution window ($2m + 1$).

$$X_{SNV} = \frac{X - \mu}{\sigma} \quad (3)$$

X is the raw data, μ is the mean value of the spectrum, and σ is the standard derivation.

$$X_{MSC} = \frac{X - \text{mean}(X)}{\text{std}(X)} \quad (4)$$

$\text{mean}(X)$ is the calculated mean value of the spectra, while $\text{std}(X)$ is the standard deviation of the spectra and is used to adjust the baseline of the spectra of each sample.

2.3.2 Local information extraction

In this study, the original dimension of the experimentally acquired spectral information is 1×520 , indicating that the spectral information of 520 bands was collected. To facilitate more efficient processing and the extraction of useful features, the spectral data are initially reshaped into a format comprising 10×52 . This entails dividing the original continuous spectral bands into 10 blocks, with each block encompassing 52 spectral bands. This reconstruction not only reduces the complexity in a single dimension but also allows for a more focused analysis of the interactions and properties of the bands within each group through grouping. The spectral data can be expressed as $X \in R^{C \times H \times W}$, where C is the number of channels, H is 10, and W is 52. The hyperspectral data is of a delicate nature and exhibits significant variations in different bands. Consequently, the spectral information is deeply correlated in both channel and spatial dimensions. The data can be viewed as a spatial remapping of the original band data in 10 and 52 dimensions, respectively. This structure allows the network to capture variations in small spatial domains as well as correlations between bands through convolution kernels. The initial extraction of spatial features of the input data is carried out using a 3×3 convolutional layer. This allows for the exploration and emphasis of local spatial and interband correlations in the spectral data, thereby facilitating the capture of key spectral and spatial features associated with mold.

$$L_i = \text{ReLU}(W * X_i + b) \quad (5)$$

Here, L_i is the extracted local feature, W is the 3×3 convolution kernel, X_i is the i -th block of the input data, b is the bias, and $*$ denotes the convolution operation.

2.3.3 Global information extraction

Hyperspectral information is essentially a high-dimensional dataset reflecting the spectral variations in different bands. Each dimension corresponds to a different band signal, thereby demonstrating the global dependence of the data. To effectively extract and utilize the global features in such data, the GV attention mechanism is introduced, as illustrated in Figs. 3 and 4. The main steps of this mechanism are as follows.

Initially, the feature map is subjected to a further processing and optimization phase, which is achieved through the implementation of a 1×1 convolution operation. This is done with the objective of integrating cross-channel information. Specifically, the information from all channels is summarized at each spatial location, and the spectral information is compressed into a single value at each location. This allows the network to integrate data from different bands without altering the spatial structure. In other words, the information from each spectral block is initially integrated, which provides a foundation for subsequent feature extraction and offers a GV. This can be expressed by

$$Y = \sigma(W_c * L + b_c), \quad (6)$$

where σ denotes the activation function, usually linear activation and $*$ denotes the convolution operation.

Subsequently, the convolutional outputs are normalized in the spatial dimension (HW) using the SoftMax function, thereby generating global attention weights. These weights indicate the

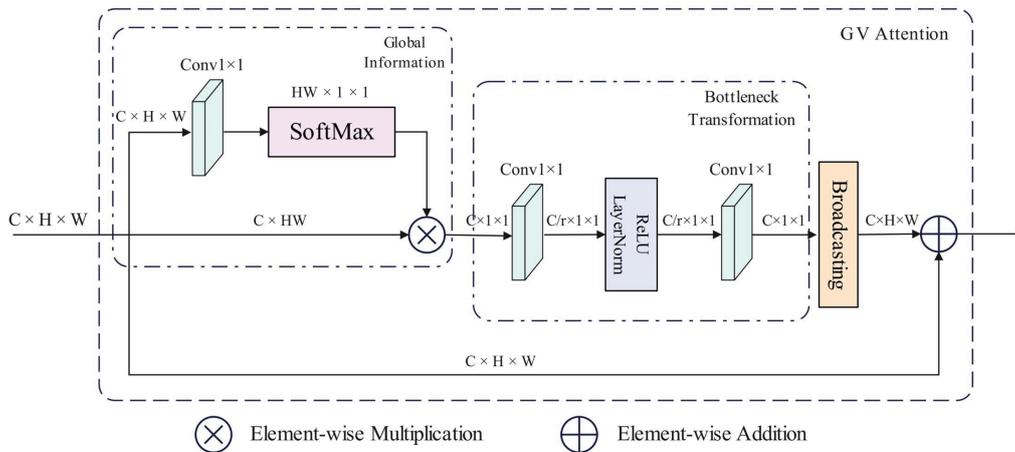


Fig. 3. (Color online) GV attention.

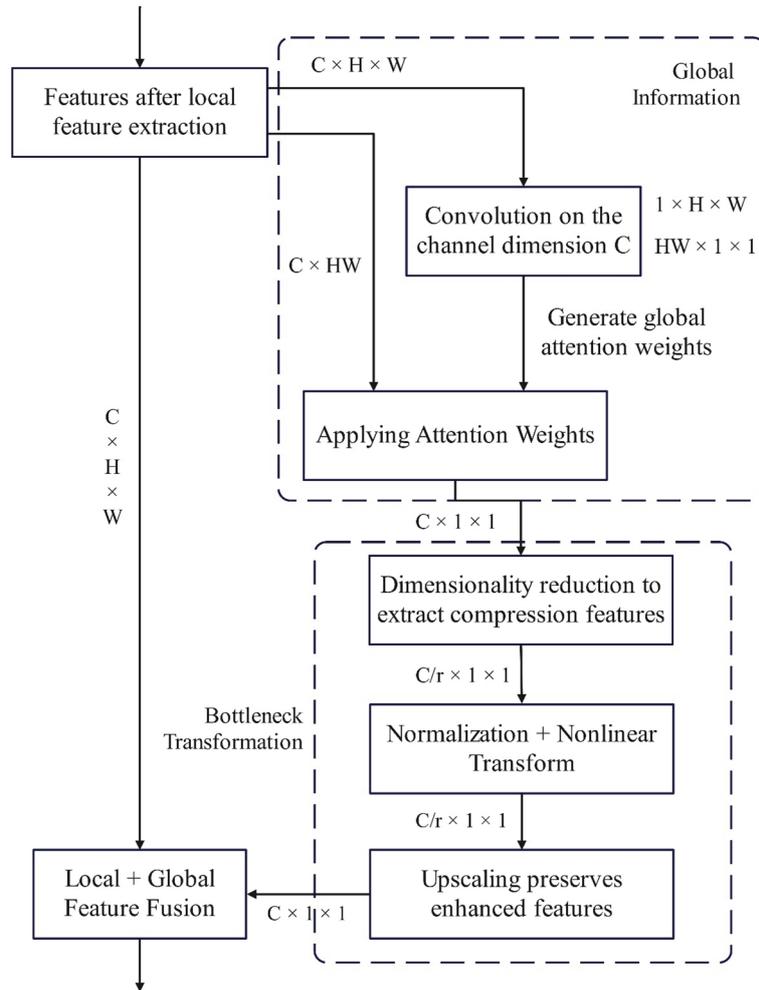


Fig. 4. GV attention mechanism process.

relative importance of each spatial location, thereby highlighting important spectral features throughout the data. This allows the model to focus more on key information in subsequent processing. The SoftMax computation can be represented in detail as

$$S_{h,w} = \frac{\exp(Y_{h,w})}{\sum_{h=1}^H \sum_{w=1}^W \exp(Y_{h,w})}, \tag{7}$$

where $S_{h,w}$ is the attentional weight at position (h, w) , ensuring that the sum of the weights at all positions is 1 and $Y_{h,w}$ is the value of the convolutional output at the same position.

The SoftMax output S is applied as weights and combined with the original features L by element-by-element multiplication, resulting in the weighted feature map Z , as illustrated by Eq. (8). This step serves to enhance the salient features identified by the attention mechanism, thereby enabling the network to focus on key spectral regions. The application of attention allows for the capture of important spectral bands with greater efficiency, thereby enhancing the accuracy and reliability of the overall analysis.

$$Z = S \otimes L \quad (8)$$

Here, \otimes denotes the element-wise multiplication.

Following global feature capture by the attention mechanism, a local information fusion process is carried out with the objective of enhancing the model performance. Initially, the number of channels is reduced to $1/r$ of the original by a 1×1 convolution (where r is a reduction factor). Subsequently, the features are extracted and compressed further to reduce the computational complexity and highlight the important features. This is demonstrated by

$$W = \text{ReLU}(W_r * Z + b_r), \quad (9)$$

where W_r is the weight of the reduced dimensional convolution, b_r is the bias, and ReLU is the nonlinear activation function.

Subsequently, LayerNorm standardizes the inputs, thereby ensuring stability and rapid convergence during the training of the network. Nonlinearity is introduced through the ReLU activation function, which enables the network to learn complex patterns and relationships. This is demonstrated by

$$U = \text{ReLU}(\text{LayerNorm}(W)), \quad (10)$$

where *LayerNorm* is the layer normalization operation that normalizes each layer in W .

A second 1×1 convolution restores the reduced channels to the original number C , thereby enabling the network to re-expand the feature space subsequent to the bottleneck transformation, as illustrated in Eq. (11). This process enables the model to retain and enhance crucial local features while reducing the computational burden.

$$V = \sigma(W_e * U + b_e) \quad (11)$$

Here, W_e is the weight of the incremental convolution, b_e is the bias, and σ is usually chosen to be a linear activation.

Finally, following the broadcasting operation, the transformed feature maps are restored to their original dimensions, thus ensuring the seamless incorporation of local features into the initial hyperspectral data representation. The elemental summation operation combines the original inputs and enhanced features to incorporate the newly learned global and local features while preserving the original spectral information. This enhances the adaptability and importance of the feature representation in the channel dimension [Eq. (12)]. This residual connectivity not only enhances the model's capacity for expression but also guarantees the completeness of the information and improves the network's performance in hyperspectral data analysis.

$$\text{Output} = V + \lambda L \quad (12)$$

Here, λ is the learnable scaling factor, which is used to adjust the degree of fusion between the input features and the processed features.

2.4 Network architecture and optimization

2.4.1 Optimization of decision network structure

Given the limited sample size of the hyperspectral system dataset, which includes tobacco with varying degrees of mold contamination, it is necessary to construct a lightweight GV-Net to ensure sufficient training under these conditions. The number of convolutional kernels and GV-attention mechanism modules utilized in the construction of the GV-Net structure directly influences the depth and complexity of the network and subsequently affects the classification performance of the entire network. Accordingly, we consider the number of convolutional kernels and the number of GV modules with a view to determining the optimal configuration for the model in terms of classification performance.

In Table 1, the impact of varying the number of convolutional kernels within a single GV module on the classification results of GV-Net can be seen. As illustrated in Table 1, an increase in the number of convolutional kernels initially results in a notable improvement in the model's classification performance. The results demonstrate that when the number of convolutional kernels is 32, GV-Net attains the optimal performance with an accuracy of 94.39%, a precision of 95.02%, a recall of 95.42%, and an F1 score of 0.9516. This suggests that a moderate number of convolutional kernels can effectively enhance the overall performance of the model. However, when the number of convolutional kernels exceeds 64, the model's performance declines rather than improves. This may be attributed to the fact that an excessive number of convolutional kernels increases the model's complexity, leading to the overfitting phenomenon.

Furthermore, Table 2 shows the classification results of GV-Net when the number of convolutional kernels is 32 with different numbers of GV modules. As evidenced in Table 2, the model exhibits optimal performance when utilizing a single GV attention module. The accuracy, precision, recall, and F1 score under this configuration are optimal. As the number of GV modules increases, the model complexity rises concomitantly. However, despite this increase in complexity, the performance does not significantly improve. Rather, there is no significant difference between the models with different numbers of GV modules. This indicates that increasing the number of GV modules does not always lead to a linear improvement in the

Table 1

Effect of the number of convolutional kernels on the classification performance of GV-Net.

Kernels	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
8	87.94 ± 4.49	90.01 ± 0.03	89.05 ± 0.05	88.68 ± 0.05
16	92.94 ± 3.39	93.83 ± 0.02	94.08 ± 0.03	93.54 ± 0.03
24	92.33 ± 4.44	93.58 ± 0.03	93.46 ± 0.04	93.12 ± 0.06
32	94.39 ± 2.19	95.02 ± 0.02	95.42 ± 0.02	95.06 ± 0.03
64	92.83 ± 3.79	93.92 ± 0.02	93.64 ± 0.04	93.41 ± 0.03
128	87.11 ± 5.54	88.53 ± 0.04	88.63 ± 0.05	87.73 ± 0.06
256	83.11 ± 5.48	86.83 ± 0.04	83.40 ± 0.06	83.61 ± 0.06

Table 2

Effect of the number of GV modules on the classification performance of GV-Net.

GV module	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
1	94.39 ± 2.19	95.02 ± 0.02	95.42 ± 0.02	95.06 ± 0.03
2	92.67 ± 4.22	94.40 ± 0.02	93.52 ± 0.04	93.44 ± 0.04
3	92.94 ± 5.11	92.96 ± 0.03	92.26 ± 0.05	91.89 ± 0.05
4	92.78 ± 3.49	94.13 ± 0.02	93.72 ± 0.03	93.46 ± 0.03
5	92.72 ± 3.29	93.87 ± 0.02	93.93 ± 0.03	93.49 ± 0.03

classification performance of the model. Instead, it may introduce unnecessary computational burdens.

In conclusion, in accordance with the conceptual framework of a lightweight network structure, we selected the GV-Net network structure comprising 32 convolutional kernels and a single GV attention module to facilitate the quality assessment of cigarette tobacco with various degrees of mold. This approach not only optimizes the utilization of resources but also circumvents the potential performance deterioration associated with an overly intricate model architecture.

2.4.2 Moldy tobacco identification network

In this study, we process hyperspectral data using a deep learning method and propose a lightweight network structure graph design based on the GV attention mechanism. This is used to discriminate the quality of tobacco with different degrees of mold. The network, GV-Net, is shown in Fig. 5. The specific process is as follows.

- (1) The local information feature is extracted from the raw spectral data through the application of a 3×3 convolution.
- (2) The GV attention mechanism is employed to facilitate the extraction of global information features and to direct attention to the most salient features within the context of spectral information.
- (3) The results of the GV attention mechanism are integrated through the use of two fully connected layers, which establish a nonlinear mapping relationship between the features and the labeling space. This integration allows for the global integration of the features.
- (4) A SoftMax function is employed as a classifier following the fully connected layers, thereby facilitating the discrimination of tobacco quality with various degrees of mold.

In the network training process, the Adam optimizer was employed to train the network parameters, and the cross-entropy loss function was utilized to calculate the training loss, thereby facilitating the propagation of errors. The network training process was conducted over 500 iterations, with a batch size of 40. A total of 450 samples of cigarette tobacco with three different degrees of mold were randomly selected, with 360 samples utilized as the training set and 90 samples designated as the test set. The classification performance of GV-Net was evaluated in terms of the accuracy, precision, recall, and F1 score. To avoid randomness, the mean and standard deviation of the results of 20 calculations were utilized as the final classification results.

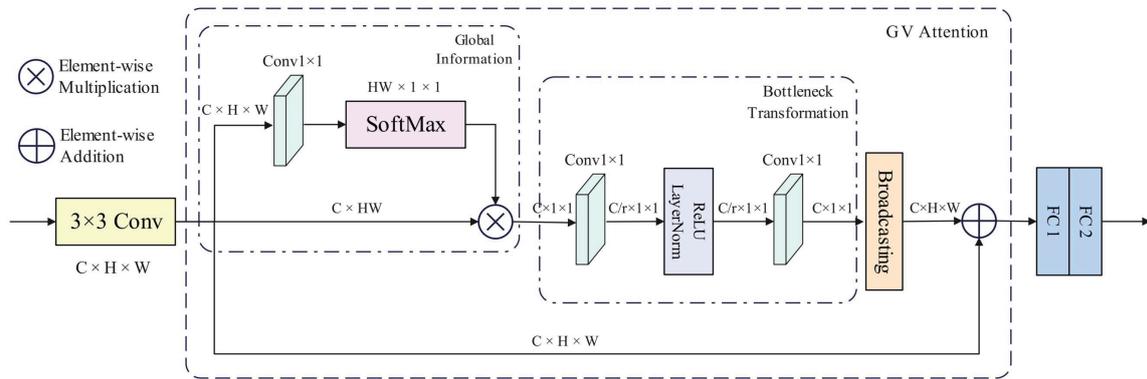


Fig. 5. (Color online) GV-Net structure

3. Results and Discussion

3.1 Ablation analysis

To conduct a comprehensive assessment of the efficacy of the proposed model in hyperspectral data analysis, two principal ablation experiments were devised in this study. These experiments were designed to facilitate a systematic investigation of the specific impact of 3×3 local spatial convolution and the GV global attention mechanism on the model performance.

Firstly, regarding the importance of local spatial convolution, it can be observed from Table 3 that the model utilizing solely 3×3 convolution demonstrates a notable proficiency in processing spatial dimensional features. This is evidenced by an enhancement in performance compared with the baseline model, with an increase in accuracy from 79.44 to 82.33%. Additionally, there is a rise in precision from 82.39 to 84.24%, and a corresponding increase in recall from 80.53 to 83.64%. Furthermore, the F1 score also exhibits an improvement, rising from 0.8035 to 0.8295. These findings suggest that local spatial convolution is a crucial technique for comprehending and analyzing spatial distribution features in hyperspectral data. Although the model in this configuration is somewhat inadequate for the extraction of spectral features, it can nevertheless enhance overall accuracy by facilitating the expression of local spatial features.

Secondly, the incorporation of the GV global attention mechanism serves to enhance the model's performance. The configuration that includes only the GV attention mechanism demonstrates a notable improvement in all metrics, with an increase of 5.12% in accuracy, 4.82% in precision, 4.23% in recall, and 0.0487 in F1 score. This suggests that the GV attention mechanism is capable of effectively integrating the spectral information across the bands and of enhancing the model's ability to comprehend the spectral data as a whole.

Ultimately, the model demonstrates optimal performance in the configuration that incorporates both 3×3 local convolution and a GV global attention mechanism. This configuration yields the highest metric performance across the board. The accuracy reached 94.39%, the precision 95.02%, and the recall 95.42%, with the F1 score being 0.9516. These findings underscore the significance of a combined approach to feature extraction in spectral

Table 3
Results of ablation experiments with different structures of GV-Net.

3×3 Conv	GV	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
		79.44 ± 4.52	82.39 ± 0.04	80.53 ± 0.05	80.35 ± 0.04
√		82.33 ± 4.39	84.24 ± 0.03	83.64 ± 0.04	82.95 ± 0.05
	√	84.56 ± 4.92	87.21 ± 0.04	84.76 ± 0.05	85.22 ± 0.05
√	√	94.39 ± 2.19	95.02 ± 0.02	95.42 ± 0.02	95.06 ± 0.03

data analysis, whereby local and global characteristics are integrated to enhance the overall performance of the model.

In conclusion, the results of these ablation experiments not only corroborate the necessity of 3×3 convolutional computation, but also illustrate the significance of the GV global attention mechanism in improving the accuracy of model prediction. These findings offer novel insights and methodologies for addressing complex hyperspectral data, which necessitates the simultaneous comprehension of local and global information.

3.2 GV-Net performance analysis

In this study, we introduced and successfully implemented GV-Net for detecting mold in cigarette tobacco. Our model integrates 3×3 local spatial convolutions with a global attention mechanism to effectively assimilate global features and demonstrated superior performance in tobacco mold detection tasks compared with traditional attention models. Using GV-Net as the fundamental network framework, we performed comparative and analytical examinations of five established attention mechanism modules, with a view to substantiating the efficacy of the GV attention mechanism. The contrasted attention mechanisms were the selective kernel networks (SK),⁽¹⁹⁾ squeeze and excitation (SE),⁽²⁰⁾ efficient channel attention (ECA),⁽²¹⁾ convolutional block attention module (CBAM),⁽²²⁾ and coordinate attention (CA).⁽²³⁾

Table 4 shows the performance outcomes of distinct attentional mechanisms for the task of discriminating the quality of cigarette tobacco with various degrees of mold contamination. The results demonstrate that GV-Net exhibits superior performance compared with the other models across all evaluation metrics. Specifically, GV-Net achieves 94.39% accuracy, 95.02% precision, 95.42% recall, and a 0.9516 F1 score. For the optimal ECA-Net classification outcomes in channel attention, GV-Net markedly outperforms the others in the integrated assessment of accuracy, precision, recall, and F1 score, thereby substantiating the superiority of GV-Net in data characterization. In comparison, CBAM, as an alternative model that integrates channel and spatial attention, demonstrates superior performance to a subset of single-channel attention models. However, its outcomes remain inferior to those of GV-Net. This result underscores the significance of the residual structure in the configuration of GV-Net. The residual structure enables the model to learn more profound feature representations, thereby facilitating more effective navigation of uncertainty and complexity in classification tasks.

Furthermore, the introduction of GV-Net not only represents a theoretical innovation but also demonstrates significant practical advantages. We provide the tobacco industry with a new, efficient tool for mold detection, which will significantly enhance the quality control in tobacco

Table 4
Performances of GV-Net and other attention networks.

Attention Networks	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
SK-Net	83.17 ± 4.96	85.47 ± 0.05	83.71 ± 0.06	83.84 ± 0.06
SE-Net	85.33 ± 5.89	87.64 ± 0.03	86.86 ± 0.06	85.99 ± 0.06
ECA-Net	86.83 ± 4.53	89.14 ± 0.03	88.37 ± 0.04	88.13 ± 0.04
CA-Net	83.11 ± 7.05	86.45 ± 0.03	84.23 ± 0.07	83.92 ± 0.07
CBAM-Net	84.83 ± 4.13	86.33 ± 0.02	86.96 ± 0.04	86.18 ± 0.04
GV-Net	94.39 ± 2.19	95.02 ± 0.02	95.42 ± 0.02	95.06 ± 0.03

production and processing. Compared with traditional manual visual inspection and chemical detection methods, GV-Net reduces human error, increases operational efficiency, and significantly improves the precision of quality control. It is expected to greatly reduce losses in tobacco products due to mold, thereby protecting consumer health and showcasing potential applications in broader areas of product quality inspection.

Overall, the GV-Net model has shown excellent accuracy and efficiency in experiments of tobacco mold detection. The improvements in its robustness and reliability also indicate the model's potential to contribute to tobacco quality control and safety supervision. As hyperspectral imaging technology continues to evolve, we anticipate that GV-Net will play a more significant role in the field of product quality inspection.

4. Conclusions

In this study, hyperspectral technology was employed to obtain the spectral information of cigarette tobacco with various degrees of mold. A GV global attention mechanism based on this information was proposed, and a lightweight network was designed to effectively identify the quality of cigarette tobacco with different degrees of mold. The principal findings are as follows.

- (1) In this study, preliminary feature extraction was conducted using 3×3 local spatial convolution in accordance with the characteristics of the spectral information of moldy tobacco. The GV attention mechanism was employed to effectively integrate the spectral information across bands, thereby facilitating the extraction of global information features and spectral-information-focused feature attention.
- (2) In a comparative analysis of multiple attention mechanisms and models, GV-Net demonstrated the best performance with an accuracy rate of 94.39%, a precision rate of 95.02%, a recall rate of 95.42%, and an F1 score of 0.9516. This effectively confirmed the high-performance recognition of various degrees of mildew in cigarette tobacco. Furthermore, the model exhibited the lowest standard deviation in classification, which indicates its excellent generalization ability.

In conclusion, the GV-Net model provides a new and effective tool for detecting mold in tobacco, offering rapid and accurate detection capabilities that are expected to significantly enhance quality control in tobacco production and processing. Moving forward, we plan to further explore the potential applications of this model in the inspection of other products, as well as how this technology can be more broadly applied in actual industrial production.

Acknowledgments

This work was supported by the Science and Technology Program of Jilin Tobacco Industry Co., Ltd. (KJXM-2024-12) and (KJXM-2022-11).

References

- 1 F. Yang, T. Zhang, S. Lu, Z. Li, L. Luo, C. Tian, and L. Ji: *J. Anhui Agric. Sci.* **33** (2015) 370.
- 2 N. Zhai: *Yunnan Normal Univ.* (2021).
- 3 F. Kong, J. Lin, C. Zhang, G. Jiang, J. Liu, and S. Liang: *Acta Tabacaria Sin.* **15** (2009) 78. <https://doi.org/10.3969/j.issn.1004-5708.2009.05.017>
- 4 Y. Zhang, Y. Tian, W. Sun, X. Mu, P. Gao, and G. Zhao: *Spectrosc. Spectral Anal.* **2** (2020) 535. [https://www.gpxygpf.com/EN/10.3964/j.issn.1000-0593\(2020\)02-0535-08](https://www.gpxygpf.com/EN/10.3964/j.issn.1000-0593(2020)02-0535-08)
- 5 S. Liu, H. Yu, M. Chen, Z. Piao, T. Yu., F. Li, and Y. Sui: *Spectrosc. Spectral Anal.* **5** (2020) 1575. <https://api.semanticscholar.org/CorpusID:234994825>
- 6 L. Zhang, Z. Rao, and H. Ji: *Spectrosc. Lett.* **53** (2020) 207. <https://doi.org/10.1080/00387010.2020.1726402>
- 7 X. Deng, G. Zeng, Z. Zhu, Z. Huang, J. Yang, Z. Tong, X. Yin, T. Wang, and Y. Lan: *J. South China Agric. Univ.* **6** (2020) 100. <https://doi.org/10.7671/j.issn.1001-411X.202006042>
- 8 S. Liu, X. Tan, C. Liu, C. Zhu, W. Li, S. Cui, Y. Du, D. Huang, and F. Xie: *Spectrosc. Spectral Anal.* **11** (2019) 3540. [https://www.gpxygpf.com/EN/10.3964/j.issn.1000-0593\(2019\)11-3540-07](https://www.gpxygpf.com/EN/10.3964/j.issn.1000-0593(2019)11-3540-07)
- 9 S. Zhang, J. Huang, L. Qin, and H. Li.: *Trans. Chin. Soc. Agric. Mach.* **4** (2019) 196. <https://doi.org/10.6041/j.issn.1000-1298.2019.04.022>
- 10 Z. Sun, T. Wang, J. Li, X. Zou, and X. Liu: *Spectrosc. Spectral Anal.* **7** (2020) 2208. [https://www.gpxygpf.com/EN/10.3964/j.issn.1000-0593\(2020\)07-2208-07](https://www.gpxygpf.com/EN/10.3964/j.issn.1000-0593(2020)07-2208-07)
- 11 Z. Xie, H. Xu, Q. Huang, and P. Wang: *Trans. Chin. Soc. Agric. Eng.* **13** (2019) 277. <https://doi.org/10.11975/j.issn.1002-6819.2019.13.033>
- 12 L. Liu, D. He, M. Li, X. Liu, and J. Qu: *Chin. J. Lasers* **11** (2020) 291. <https://doi.org/10.3788/CJL202047.1111002>
- 13 D. Yang, J. Yuan, Q. Chang, H. Zhao, and Y. Cao: *Infrared Phys. Technol.* **109** (2020) 103412. <https://doi.org/10.1016/j.infrared.2020.103412>
- 14 Y. Jia, Z. Li, R. Gao, X. Zhang, H. Zhang, and Z. Su: *Spectrosc. Lett.* **4** (2022) 240. <https://doi.org/10.1080/00387010.2022.2053163>
- 15 H. Xue, X. Xu, Y. Yang, D. Hu, and G. Niu: *Sensors* **24** (2024) 1855. <https://doi.org/10.3390/s24061855>
- 16 H. Yang, F. Qu, Y. Yang, X. Li, P. Wang, S. Guo, and L. Wang: *Sensors* **24** (2024) 4635. <https://doi.org/10.3390/s24144635>
- 17 K. Dhakal, U. Sivaramkrishnan, X. Zhang, K. Belay, J. Oakes, X. Wei, and S. Li: *Sensors* **23** (2023) 3523. <https://doi.org/10.3390/s23073523>
- 18 H. Jiang, X. Jiang, Y. Ru, Q. Chen, X. Li, L. Xu, H. Zhou, and M. Shi: *Phys. Technol.* **123** (2022) 104169. <https://doi.org/10.1016/j.infrared.2022.104169>
- 19 X. Li, W. Wang, X. Hu, and J. Yang: *Proc. IEEE-CVF Conf. Computer Vision and Pattern Recognition (IEEE, 2019)* 510–519. <https://doi.org/10.1109/CVPR.2019.00060>
- 20 J. Hu, L. Shen, and G. Sun: *Proc. IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2018)* 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>
- 21 Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu: *Proc. IEEE-CVF Conf. Computer Vision and Pattern Recognition (IEEE, 2020)* 11534–11542. <https://doi.org/10.1109/CVPR42600.2020.01155>
- 22 S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon: *Proc. European Conf. Computer Vision (ECCV, 2018)* 3–19. https://doi.org/10.1007/978-3-030-01234-2_1
- 23 Q. Hou, D. Zhou, and J. Feng: *Proc. IEEE-CVF Conf. Computer Vision and Pattern Recognition (IEEE, 2021)* 13708–13717. <https://doi.org/10.1109/CVPR46437.2021.01350>