### Check for updates

### OPEN ACCESS

EDITED BY Xing Yang, Anhui Science and Technology University, China

REVIEWED BY Chao Kang, Guizhou University, China Budi Yanto, Universitas Pasir Pangaraian, Indonesia

\*CORRESPONDENCE Weimin Guo Imaguoweimin1984@sina.com Jianjun Liu Iuijj325@163.com

RECEIVED 07 April 2025 ACCEPTED 06 June 2025 PUBLISHED 02 July 2025

#### CITATION

Xu Q, Zhang Y, Wang A, Chen G, Cai X, Zhou S, Li J, Jin B, Yan D, Huang J, Chen Z, Zhang H, Wang J, Guo W and Liu J (2025) Intelligent recognition of tobacco leaves states during curing with deep neural network. *Front. Plant Sci.* 16:1604382. doi: 10.3389/fpls.2025.1604382

#### COPYRIGHT

© 2025 Xu, Zhang, Wang, Chen, Cai, Zhou, Li, Jin, Yan, Huang, Chen, Zhang, Wang, Guo and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Intelligent recognition of tobacco leaves states during curing with deep neural network

Qiang Xu<sup>1</sup>, Yanling Zhang<sup>1</sup>, Aiguo Wang<sup>1</sup>, Guangqing Chen<sup>2</sup>, Xianjie Cai<sup>3</sup>, Shuoye Zhou<sup>2</sup>, Junying Li<sup>4</sup>, Baofeng Jin<sup>5</sup>, Ding Yan<sup>3</sup>, Jiajie Huang<sup>5</sup>, Zuxiao Chen<sup>6</sup>, Heng Zhang<sup>4</sup>, Jianwei Wang<sup>1</sup>, Weimin Guo<sup>1\*</sup> and Jianjun Liu<sup>2\*</sup>

<sup>1</sup>Tobacco Agricultral Labaratory, Zhengzhou Tobacco Research Institute of China National Tobacco Corporation (CNTC), Zhengzhou, China, <sup>2</sup>Tobacco Leaf Administration Office, Henan Provincial Tobacco Company of CNTC, Zhengzhou, China, <sup>3</sup>Tobacco Raw Materials Procurement Center, Shanghai Tobacco Group Co. Ltd., Shanghai, China, <sup>4</sup>Pingdingshan Branch, Henan Provincial Tobacco Company, Pingdingshan, China, <sup>5</sup>Technical Center, China Tobacco Guangdong Industrial Co. Ltd, Guangzhou, China, <sup>6</sup>Technical Center, Jilin Tobacco Industry Co. Ltd., Changchun, China

**Introduction:** The state monitoring of tobacco leaves during the curing process is crucial for process control and automation of tobacco agricultural production. While most of the existing research on tobacco leaves state recognition focused on the temporal state of the leaves, the morphological state was often neglected. Moreover, the previous research typically used a limited number of non-industrial images for training, creating a significant disparity with the images encountered in actual applications.

**Methods:** To investigate the potential of deep learning algorithms in identifying the morphological states of tobacco leaves in real industrial scenarios, a comprehensive and large-scale dataset was developed in this study. This dataset focused on the states of tobacco leaves in actual bulk curing barn in multiple production areas in China, specifically recognizing the degrees of yellowing, browning, and drying. Then, an efficient deep learning method was proposed based on this dataset to enhance the predictive performance.

**Results:** The prediction accuracy achieved for the yellowing degree, browning degree, and drying degree were 83.0%, 90.5%, and 75.6% respectively. The overall average accuracy, satisfied the requirements of practical application scenarios with a value of 83%.

**Discussion:** Our proposed framework effectively enables morphological state recognition in industrial curing, supporting parameter optimization and enhanced tobacco quality.

#### KEYWORDS

tobacco leaves, large-scale dataset, bulk curing barn, image recognition, deep learning

## **1** Introduction

Curing is an important process of tobacco production, converting fresh leaves into commercially cigarette raw materials. Curing quality of tobacco leaves directly determines farmers' income and the cigarette quality. Matching the optimal curing technology in real time according to the state of tobacco leaves is the key to determine the curing quality of tobacco leaves (Siddiqui, 2001; Zhao et al., 2024). At present, the identification of tobacco leaf states during the curing process mainly relies on people's subjective experience. Inaccurate cognition has caused problems such as uneven curing quality of tobacco leaves, large curing losses, and weak industrial usability (Ma et al., 2021).

The rapid development of technologies such as the Internet of Things and artificial intelligence has proposed new methods for solving such problems. Researchers conducted research on the state of tobacco leaves during the curing process by using the collected temperature, humidity and tobacco leaf images. The researches on tobacco leaf state identification mainly can be divided into two categories: temporal state and morphological state. The temporal state of tobacco leaves refers to the division of the tobacco leaf curing process into different stages based on the curing time of the tobacco leaves, such as yellowing stage, color fixing stage, and stem drying stage (Li et al., 2022; Lu et al., 2023). Although this method has achieved high accuracy (More than 90%), it is difficult to adjust the temperature and humidity of the curing room in real time based on the recognition results. Therefore, some researchers focused on the morphological state recognition (Wang et al., 2017; Wang and Qin, 2022; Zhao et al., 2024). The morphological state of tobacco leaves refers to the specific state of yellowing degree, drying degree and browning degree of tobacco leaves identified based on tobacco leaf images, thereby replacing the human eye observation and subjective analysis during the curing process, and providing more accurate, faster and scientific results for identifying the state of tobacco leaves. Meanwhile, it can also provide an important reference for the real-time adjustment of the curing technology (Condorí et al., 2020; Pei et al., 2024). To further improve the recognition accuracy, the texture information of tobacco leaf images has also begun to be gradually utilized except the widely used color information (Wang et al., 2017).

However, there are still some problems limited the accuracy and application of these recognition models. Previous research on tobacco leaf state recognition often relied on small-scale (i.e., hundreds to just over a thousand samples) (Zhao et al., 2024) or non-industrial datasets, which were collected using small ovens, experimental chambers, etc (Wu and Yang, 2021; Condorí et al., 2020). For example, some studies have acquired images through smartphone photography (Howard et al., 2017; Zhang et al., 2023), but the quality and characteristics of these images differ significantly from those captured in actual curing barns, limiting their applicability to real-world bulk curing scenarios. Some researchers tried to collect tobacco images in the actual curing barns, but the complex environment during curing process resulted in image distortion, out of focus, obvious color difference and only partial tobacco image acquisition, which are still the core problems limiting the acquisition of tobacco condition information (Condorí et al., 2020; Pei et al., 2024; Wu et al., 2014; Wu and Yang, 2019; Zhang et al., 2013) and further effected the wide application of the recognition models.

To overcome these limitations, the objective of this study is to (i) construct a comprehensive and large-scale image dataset captured directly from actual bulk curing barns; (ii),propose a deep learning approach to recognize the morphological states of tobacco leaves throughout the curing process based on this dataset; (iii) establish a benchmark framework using state-of-the-art models, including the Swin Transformer V2, to enhance predictive performance and support intelligent decision-making during tobacco curing.

## 2 Materials and methods

## 2.1 Large-scale curing tobacco leaves dataset

The large-scale curing tobacco leaves dataset involved gathering a substantial amount of real-world data from bulk curing barn and having them meticulously labeled by experts in tobacco curing. To facilitate the recognition of tobacco leaves states during the curing process, 17,420 images of tobacco leaves from 10 main production areas in China were collected, including Henan, Fujian, Yunnan, Guizhou, etc. All tobacco leaf images in the dataset were collected by a newly developed autonomous imaging device (Figure 1). The image device was installed in the middle shed on one side of the grill near the heating chamber in the curing barns (Xu et al., 2024). The sampling interval was set to 10 minutes, and the tobacco images of the curing process were obtained, which marked the time, location, temperature, and humidity in the curing barns and the status of the tobacco leaves.

Experts in tobacco curing in China conducted evaluations focusing on three distinct states of the tobacco leaf: the degree of yellowing, the degree of browning, and the degree of drying. Different degrees were categorized based on the extent of morphological differences observed in the various states of the tobacco leaves (Table 1). In Figure 2, some reference images along were provided with their corresponding yellowing degree, browning degree, and drying degree labels for further clarity. This visual representation aids in understanding the various states and degrees of tobacco leaves during the curing process.

### 2.2 Recognition algorithm

### 2.2.1 Method overview

A recognition algorithm was implemented to accurately and efficiently identify tobacco states during the curing process based on deep neural networks. As depicted in Figure 3, three components were comprised in the recognition algorithm: (i) a pre-trained backbone pre-trained on universal image recognition datasets (e.g., ImageNet), (ii) a Fourier filter module, and (iii) a common



FIGURE 1 Image acquisition device and installation photos of tobacco leaves during curing process.

#### TABLE 1 The detailed definition of the states of tobacco leaf.

States	Degrees
	0. 50%~60% yellowing
	1. 70%~80% yellowing
Yellowing degree	2. Leaves yellow, veins green and green base
	3. Leaves yellow and veins green
	4. Main veins fade cyan to white
	5. Partial main veins shrink and turn purple
	6. Main vein purpling
	0. Leaf swell and harden
	1. Leaf tip softening
	2. Leaf softening
	3. Leaf wilt completely
Drying degree	4. Leaf blade hook tip curl
	5. Leaf drying 1/2~2/3
	6. Leaf drying completely (Large roll)
	7. Main vein drying 1/2
	8. Main vein drying completely
	0. 0%
Browning degree	1.<10%
	2. 10%~20%
	3. 20%~30%
	4. 30%~50%
	5. >50%

color filter module. The pre-trained backbone extracted highly discriminative image features  $F_{img}$  by leveraging previously learned information and further fine-tunes the parameters on the proposed datasets. The Fourier filter module was designed to extract the wrinkle information of tobacco leaves  $F_{spect}$  by utilizing the Fourier spectrum map of the image and a convolution-based network. The common color filter module calculated the quantized color histogram of the image and filtered it by frequency, thereby screening out high-frequency colors and the order of the colors appearing in the image. It further employed a fully connected layer to extract high-frequency color features  $F_{color}$ . Finally, all features were concatenated and fed into the fully connected classifier to predict three states of tobacco.

In summary, a three-branch network was presented in this study incorporating two proposed modules: the Fourier filter module and the common color filter module. Through joint training, the network employs a multi-label recognition head to simultaneously classify three tobacco states.

### 2.2.2 Pre-trained backbone

To build a deep learning benchmark and verify the efficacy of deep learning networks for the tobacco leaves states recognition, four extensively employed deep neural networks pre-trained on ImageNet were utilized as backbone network, including VGG19, ResNet-152, ViT, Swin-Transformer and Swin-Transformer v2 (Simonyan and Zisserman, 2014; He et al., 2016; Dosovitskiy et al., 2020; Liu et al., 2021; 2022). VGG19 is a profound convolutional neural network comprising 16 convolution layers and 3 fully connected layers. ResNet-152 is an exceptionally deep convolutional neural network, reaching a depth of up to 152 layers, made possible by employing skip connections to bypass certain layers. ViT-Large is a model that applies the transformer architecture, which has demonstrated impressive performance in the field of computer vision recently. The Swin Transformer is a hierarchical vision model engineered for efficient image recognition. It utilizes non-overlapping windows and selfattention within each window to process images at multiple scales. Swin-Transformer V2 enhances this approach with innovations like scaled cosine attention, post- normalization, and a log-spaced continuous position bias, boosting stability, scalability, and overall performance.

These pre-trained backbones extract highly discriminative image features by leveraging the information learned before and further fine-tuning the parameters on the proposed datasets. An evaluation of the predictive accuracy of these four networks in determining the state of tobacco was conducted. To further enhance their performance, Swin-Transformer v2 was incorporated as the backbone network and its core components including the following two aspects.

### 2.2.2.1 The attention mechanism

In the Swin-Transformer v2, the attention mechanism is a crucial component (Vaswani et al., 2017). It performs multiple attention operations to extract highly discriminative features. Given N image patches within an image and their corresponding features  $F \in RN \times d'$  (the process to obtain F will be detailed in the



Reference image data of (A) Pingdingshan and (B) Zunyi of different yellowing and drying degrees of tobacco leaves during curing.

subsequent paragraph), the operation of the self-attention mechanism is as follows:

First, the query (Q), key (K), and value (V) are computed using a linear transformation with the trainable weight  $W \in R^{d' \times 3d}$ (Equations 1):

$$Q, K, V = WF \tag{1}$$

where Q, K, V  $\in \mathbb{R}^{N \times d}$ , W represents the weight matrix of a linear layer, where d' is the input channel dimension of the feature F  $\in K \ V \times d'$ , d is the output channel dimension of the linear layer. Then, the attention mechanism is applied to extract the output feature (Equations 2, 3):

Attention(Q, K, V) = SoftMax(cos(Q, K)/
$$\gamma$$
 + B)V (2)

$$SoftMax(X)_{i} = \frac{exp(X_{i})}{\sum_{j=1}^{N} exp(X_{j})}$$
(3)

where  $\cos(Q, K) \in \mathbb{R}^{N \times N}$  is the pair-wise cosine similarity,  $\gamma$  is a learnable scalar and *N* denotes the number of image patches. The B  $\in \mathbb{R}^{N \times N}$  serves as a relative positional encoding which is predicted by two trainable fully connected layers g (Equation 4):

$$B(\Delta X, \ \Delta Y) = g(\Delta X, \ \Delta Y) \tag{4}$$

### 2.2.2.2 Patch splitting and merging

The attention mechanism is implemented within the image patches, which are derived from segmenting an input image into non-overlapping patches. Each patch is regarded as an individual unit, referred to as a "token". Then a linear projection transforms each token to the token features F. The token features are subsequently fed into numerous layers of the Swin-Transformer V2. Each layer is designed to extract and refine the information embedded within the tokens. This refinement process involves a series of operations that integrate attention mechanisms. As the



tokens progress further into the depths of the network, a method known as patch merging is employed. This method reduces the number of tokens by integrating the features of neighboring patches. The result is an ensemble of tokens, thereby effectively establishing a hierarchical representation of the initial image.

### 2.2.3 Fourier filter module

The prediction of drying is more dependent on the morphological characteristics of the tobacco leaves compared to the prediction of yellowing and browning. As depicted in Figure 4, images of tobacco leaves with a higher degree of drying display a greater number of wrinkles, which are associated with the highfrequency information in the image's Fourier spectrum. As the drying process progresses, the formation of surface wrinkles on tobacco leaves increases, thereby amplifying the high-frequency intensity in the corresponding images.

To validate this hypothesis, the average high-frequency and lowfrequency intensities of images with different degrees of drying within the training set were computed. The results, as illustrated in Figure 5, revealed that the average high-frequency intensity of the corresponding image exhibits an upward trend as the degree of drying increases. This suggested a correlation between the image's frequency domain information and its degree of drying. Consequently, the Fourier spectrum was incorporated as information into the network and a Fourier filter was construct to aid in the prediction of the degrees of drying.

Specifically, the image was converted into a gray-scale image at first and then its two-dimensional Fourier spectrum was calculated (Equation 5):



Images of tobacco leaves with (A) low drying and (B) high drying degrees.



$$F(u,v) = \sum_{m=0}^{H-1} \sum_{n=0}^{W-1} f(m,n) e^{-j2\pi(um/H+vn/W)}$$
(5)

where F is the Fourier spectrum with the same shape as the input image. The H and W are the height and width, respectively. Then the real and imaginary components of each frequency position within this matrix are utilized as input for the neural network. Where the values of real and imaginary components are represented as separate image channels. Four-layer convolutional layers are constructed to extract spectral features Fspect from this input. This approach enables the effective capture of intricate patterns within the Fourier spectrum, thereby enhancing the robustness of the drying prediction.

### 2.2.4 Common color filter module

In conjunction with the Fourier Filter module, which is primarily designed to augment the prediction of drying, an additional module denoted as the Common Color Filter was introduced. This module was specifically engineered to enhance the prediction accuracy of the states intrinsically tied to the color of the tobacco leaves. However, the image signal frequently encompasses elements beyond the mere color of the tobacco, the presence of noise color could potentially compromise the final prediction. Therefore, it is important to eliminate as many noisy pixels as possible to mitigate color interference. To address this challenge, the characteristic that the tobacco in the bulk curing barn is densely arranged and typically occupies a consistent position were exploited. This strategy aided in the effective reduction of noise and enhanced the accuracy of the proposed model.

As depicted in Figure 6, this algorithm filtered out the most common colors in an image and extracts relevant features for subsequent use. It accomplished this through a series of steps. Firstly, it applied a center-cropping technique to the image. This process focused on the central part of the image, which contained the most important information and reduced the impact of potential noise from the image's periphery. Next, it quantized the color space. Quantization was a process that reduced the number of distinct colors used in an image, while still maintaining its overall visual construction. This step can reduce the size of the color space and the computation load when calculating the color histogram. Following this, the most common colors were selected. These colors are shown in Figure 7, it depicted typically the tobacco leaves. By focusing on these colors, the algorithm can more accurately predict the state of yellowing and browning. Finally, the common color feature Fcolor was extracted using a single fully-connected layer.

### 2.3 Preprocessing and evaluation metrics

During training, input images were resized to 384×384, the supported input size of our image backbone. Data augmentation was applied using the RandAugment method with a magnitude of 9 and a standard deviation of 0.5.

## Algorithm 1 Common Color Filter Image $I_{ori} = \{I_{ori}^{ijk}\}_{H \times W \times 3}, I_{ori}^{ijk} \in \{0, 1, 2, ..., 255\}$ HyperParameters: Quantization Parameter Q, Number of colors reserved K**Initialize:** Quantization color range $R_c = \frac{256}{Q}$ 1. Center crop image to $(\frac{H}{2} \times \frac{W}{2} \times 3)$ : $I = CenterCrop(I_{ori})$ 2. Map the RGB values of each spatial pixel to a single value: $I_{hash} = \{I^{ij0}R_c^2 + I^{ij1}R_c + I^{ij2}\}$ 2. 3. Statistical histogram of quantization color in $I_{hash}$ : N = bincount( $I_{hash}$ ) 4. Select top-k colors from the histogram: $V_{topk} = TopK(N, K)$ 5. Map the value in $V_{topk}$ back to the RGB values $I_{filter}$ . The i-th value in $I_{filter}^{i} = \{V_{topk}^{i} / / R_{c}^{2} V_{topk}^{i} / / R_{c} Mod R_{c}, V_{topk}^{i} Mod R_{c}\}, \text{ where } i \text{ represents}$ the i-th value in the top-k colors. 6. Extract feature $F_{color}$ using convolutional neural networks. FIGURE 6 Implementation procedure of filtering algorithm.



The primary evaluation metric employed to assess the algorithms' performance was top-1 accuracy, expressed as a percentage. Specifically, the network's prediction probability for the most likely state was selected as the predicted result and compared with the ground-truth label. The accuracy is then calculated using the (Equation 6):

$$Accuracy = \frac{\#TruePositive}{\#TruePositive + \#FalsePostive}$$
(6)

where *#TruePostive* represents instances where the model accurately predicts the positive class, *#FalsePostive* denotes instances of incorrect predictions by the model, the symbol "#" indicates the number of corresponding instances or categories. The individual prediction accuracy for three different tobacco leaves states as well as their average accuracy were separately evaluated.

## 3 Results and discussion

## 3.1 Accuracy of tobacco leaves state prediction

### 3.1.1 Comparison with traditional algorithms

The performance of three traditional algorithms on the same task, including K nearest neighbor (KNN), support vector machines (SVM), and random forest (RF), was compared with our deep-learning method (Cover and Hart, 1967; Hearst et al., 1998; Breiman, 2001). Three distinct predicted states of tobacco leaves and hyperparameters for each method are shown in Tables 2, 3, respectively.

Firstly, the deep-learning method significantly outperformed the three traditional methods, highlighting the potential benefits of using deep learning for this task. Secondly, the KNN method performed well in predicting browning degree by directly computing the difference between two images as the norm, and was ineffective for yellowing degree and drying degree. This suggested that the prediction of browning degree relied more on the color information within the image.

### 3.1.2 Comparison with deep learning algorithms

As shown in Tables 4, 5, Swin-Transformer v2-Large achieved state-of-the-art performance among all other single backbones and benefited from a larger in-put size. The method proposed in this study further enhanced Swin-Transformer v2-Large's performance, demonstrating a higher accuracy with an average accuracy of 83.0%. To verify the effectiveness of the two proposed modules (FFM and CCFM), The new method was also implemented based on the Swin-Transformer-Large as the pre-trained backbone network. It can be observed that after integrating FFM and CCFM modules with the Swin-Transformer, an improvement in accuracy was achieved (81.8% vs. 81.6%). This indicated that these modules possessed a certain degree of robustness.

To further verify the impact of image resolution and pre-trained image datasets on the accuracy of tobacco condition recognition, experiments with different parameters based on Swin-Transformer and Swin-Transformer v2 were conducted. The results indicated that the accuracy of tobacco leaves states recognition can benefit from being pre-trained on a larger image dataset (i.e., ImageNet-22k), even if it was not directly related to tobacco leaves. Higher

TABLE 2 The prediction accuracy	y of different traditional methods
---------------------------------	------------------------------------

Methods		Accuracy					
		Yellowing	Browning	Drying	All		
Traditional	KNN	66.5	84.5	54.6	68.5		
	RF	71.2	84.3	64.6	73.4		
	SVM	73.9	85.7	64.4	74.7		
Deep learning	Ours	83.0	90.5	75.6	83.0		

Bold font is the best result in each experiment.

	TABLE 3	The hyperparameters	of	different	traditional	methods.
--	---------	---------------------	----	-----------	-------------	----------

Methods	Hyperparameters	Value	
	neighbors	5	
KININ	distance type	Frobenius norm	
	С	1	
SVM	kernel	Linear	
	Penalty term	L2	
	loss	squared hinge	
	Number of trees	100	
RF	Split criterion	Gini impurity	
	bootstrap	True	
	n_bins	1024	

input resolution can also improve prediction accuracy, suggesting that it is possible to further enhance the accuracy of tobacco leaf recognition by increasing the input resolution. Moreover, this method required low computational cost and provided fast output, with a less than 2GB of GPU memory for inference during testing and an average prediction time of under 0.5 seconds per image.

# 3.2 Ablation experiment of the proposed component

Ablation experiments were conducted to demonstrate the effectiveness of the three proposed components in this paper. As depicted in Table 6, when compared to the standalone backbone model, the Fourier Filter Module (FFM) contributed an absolute improvement of 1.0% in drying prediction accuracy, underscoring its efficacy in enhancing drying prediction. Similarly, the Common

TABLE 4 The prediction accuracy of different deep learning methods.

Color Filter Module (CCFM) accounted for an absolute increase of 0.8% in browning prediction, affirming its utility in tasks significantly influenced by color attributes. Moreover, as indicated in the final row, the synergistic integration of both modules lead to further enhancements in prediction precision, thereby confirming that their combined application can substantially bolster overall performance.

# 3.3 Effectiveness of the hyper-parameters in the common color filter module

The effectiveness of the two hyper-parameters, quantization parameter Q and the number of colors reserved K, related to the common color filter module were evaluated. Different values of Q and K were chosen. The experimental results are shown in Table 7. The optimum value was obtained when Q = 4 and K =20. When Q = 2, the color granularity became smaller and showed higher performance in predicting browning and yellowing, and the drying slightly decreased. One possible reason was that this fine color information dominated the feature extraction. Under the same Q, sampling with different K will also lead to different results, indicating that the balance the situations of insufficient sampling and excessive noise sampling through K was required.

# 3.4 Effectiveness of production area independent prediction

The large scale dataset comprised images acquired across various production areas, sensor discrepancies, tobacco varieties and ecological environmental variations can all contribute to inherent difference of images. While a unified training approach was straightforward and widely adopted, the impact of area specificity on the recognition of tobacco states was also explored

	Desclution	Due tusined	Accuracy				
Methods	Resolution	Pre-trained	Yellowing	Browning	Drying	All	
VGG-19	224×224	ImageNet1K	79.2	87.7	70.6	79.1	
ResNet-152	224×224	ImageNet1K	78.7	88.7	70.1	79.2	
ViT-Large	224×224	ImageNet22K	80.7	90.1	71.4	80.7	
Swin-Base	224×224	ImageNet1K	80.8	89.3	73.0	81.0	
ConvNeXt V2	224×224	ImageNet22K	81.5	90.1	74.0	81.9	
Swin-Base	224×224	ImageNet22K	81.1	89.7	73.3	81.4	
Swin-Large	224×224	ImageNet22K	81.3	90.2	73.2	81.6	
Ours(Swin-Large)	224×224	ImageNet22K	81.3	90.3	73.9	81.8	
SwinV2-Large	256×256	ImageNet22K	82.2	90.2	73.9	82.1	
SwinV2-Large	384×384	ImageNet22K	82.6	89.6	75.0	82.4	
Ours(SwinV2-Large)	384×384	ImageNet22K	83.0	90.5	75.6	83.0	

Bold font is the best result in each experiment.

### TABLE 5 The hyperparameters of our methods.

Hyperparameters	Value
warmup learning rate	2e-8
base learning rate	2e-4
end learning rate	2e-7
batch size	8
decay-epoch	5
learning rate schedule	cosine
optimizer	AdamW
drop-path	0.1
gradient clip	1

in this study. Three representative production areas with substantial data volume were selected and individual training processes were conducted. In addition, the integrated training model in all areas was used to make separate predictions for these areas and compared them with the results of the separately trained model. This allowed assessments on area-specific models' effectiveness and their practicality for different tobacco production areas.

The experiment was conducted in three distinct production areas, labeled as A, B, and C. Each bulk curing barn employed an image device, with the corresponding data statistics presented in Table 8. As shown in Table 9, it was evident across three distinct areas that the integrated model outperformed models trained individually for each area in terms of prediction accuracy. This suggested that despite the inherent difference present in images from different areas, the model can still leverage a larger image dataset to enhance its performance, surpassing that of models trained individually in each area.

### 3.5 Visualization of the confusion matrix

As shown in Figure 8, a deeper analysis of the confusion matrix corresponding to the three predicted states was conducted. The results illustrated that the proposed method demonstrated remarkable predictive accuracy. For each ground-truth label across the three states, the majority of the predicted labels align with either the ground-truth labels or their neighboring labels (the cumulative probability for these exceeds 99%), which was reasonable given the inherent difficulty in distinguishing between neighboring state image features due to their significant similarity. It is difficult to capture the morphological changes at the critical stage of the tobacco leaf curing process. Even if experienced experts make judgments, they may still misjudge. The prediction of the browning degree (Figure 8b) demonstrates high accuracy across all degrees. However, further improvements are still required in the prediction of the middle degrees of yellowing (Figure 8a) and drying (Figure 8c). Generally, the accuracy of the proposed method complete recognition is 83%, the accuracy rate of adjacent stages is more than 99%, and the fault tolerance rate is within ±1 stage, which has little impact on the curing quality during the actual curing process.

### 3.6 Visualization of the gradient.

In Figure 9, GradCam (Selvaraju et al., 2017) was used to visualize the gradients for each state. The results showed that in

Packhono	EENA	CCEM		Ассі	uracy	
Backbone	FFIM	CCFM	Yellowing	Browning	Drying	All
1			82.6	89.6	75.0	82.4
1	1		83.0	89.5	76.0	82.8
1		1	82.5	90.4	75.4	82.8
1	1	1	83.0	90.5	75.6	83.0

TABLE 6 The ablation of the proposed component.

TABLE 7 The ablation of the hyper-parameters in the common color filter module.

Q	К	Accuracy				
		Yellowing	Browning	Drying	All	
4	10	82.7	89.8	75.3	82.6	
4	20	83.0	90.5	75.6	83.0	
4	30	82.7	90.1	75.3	82.7	
8	20	83.1	90.2	74.9	82.8	
2	20	83.2	90.5	75.0	82.9	

Bold font is the best result in each experiment.

TABLE 8 Data statistics of three typical production areas.

Production area	А	В	С
Training data	1.2k	2.9k	1.2k
Test data	0.5k	1.2k	0.5k
All data	1.7k	4.1k	1.7k

this example, the model proposed in this study focused primarily on the brown parts of the image when predicting browning degree, on the petiole when predicting drying degree, and on larger areas of leaf content when predicting yellowing degree. This is similar to the

TABLE 9 The comparison of training of different areas separately.

reference positions and standards that people use to judge the three states of tobacco leaves during the curing process.

## 4 Conclusions

In this study, an large-scale dataset including 17,420 images of tobacco leaves from 10 main production areas in China was developed, with a specific emphasis on recognizing the degrees of yellowing, browning, and drying. This is a large-scale dataset specifically dedicated to the morphological recognition of the states of tobacco leaves within an actual bulk curing barn setting.

A #0.0	Individual model				Integrated model			
Area	Yellowing	Browning	Drying	All	Yellowing	Browning	Drying	All
All	-	-	-	-	83.0	90.5	75.6	83.0
А	84.7	87.1	76.9	82.9	84.8	87.6	78.7	83.7
В	82.2	90.9	77.7	83.6	83.7	90.6	77.5	84.0
С	86.0	89.0	77.8	84.3	84.6	90.5	79.5	84.9



The confusion matrices for the prediction of the (A) Yellowing, (B) Browning and (C) Drying states of the tobacco leaves.



Visualization of gradients with respect to ground truth labels. (A) Original image, (B) Browning, (C) Drying and (D) Yellowing states.

A deep learning benchmark was then established for this dataset using various deep learning networks. To further enhance the predictive performance of the deep backbone network, an efficient deep learning method was proposed, including Fourier filter module and common color filter module. This method integrated the spectral characteristics of tobacco leaves images and filters out color noise, which effectively enhanced the accuracy of our model with prediction accuracy for the yellowing degree, browning degree, and drying degree were 83.0%, 90.5%, and 75.6% respectively. The high overall average accuracy with a value of 83.0% and the availability and feasibility in different production areas have demonstrated the superior performance of the proposed method in this study, which provides a solid foundation for future research in this area.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

QX: Conceptualization, Methodology, Project administration, Software, Writing - original draft. YZ: Conceptualization, Funding acquisition, Resources, Writing - review & editing. AW: Data curation, Investigation, Writing - original draft. GC: Data curation, Validation, Writing - review & editing. XC: Data curation, Validation, Writing review & editing. SZ: Data curation, Validation, Writing - review & editing. JYL: Data curation, Validation, Writing - review & editing. BJ: Data curation, Validation, Writing - review & editing. DY: Data curation, Validation, Writing - review & editing. JH: Data curation, Validation, Writing - review & editing. ZC: Data curation, Validation, Writing - original draft. HZ: Data curation, Validation, Writing review & editing. JW: Data curation, Validation, Writing - review & editing. WG: Conceptualization, Data curation, Funding acquisition, Validation, Visualization, Writing - original draft, Writing - review & editing. JJL: Conceptualization, Data curation, Investigation, Validation, Writing - review & editing.

### References

Breiman, L. (2001). Random forests. Mach. Learn. 45, 5-32. doi: 10.1023/ A:1010933404324

Condorí, M., Albesa, F., Altobelli, F., Duran, G., and Sorrentino, C. (2020). Image processing for monitoring of the cured tobacco process in a bulk-curing stove. *Comput. Electron. Agric.* 168, 105113. doi: 10.1016/j.compag.2019.105113

Cover, T., and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27. doi: 10.1109/TIT.1967.1053964

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv arXiv:2010.11929*. doi: 10.48550/arXiv.2010.11929

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE. 770–778.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the foundation from the China National Tobacco Corp., No. 110202201051 (SJ-01) and No. 110202101084 (SJ-08), and the foundation from the Henan Provincial Tobacco Company No.2024410000240029. The authors declare that this study received funding from Henan Provincial Tobacco Company. The funder had the following involvement in the study: data collection.

## Conflict of interest

Authors XC and DY were employed by the company Shanghai Tobacco Group Co. Ltd. Authors JYL and HZ were employed by the company Henan Provincial Tobacco Company. Author BJ was employed by the company China Tobacco Guangdong Industrial Co. Ltd. Author ZC was employed by the company Jilin Tobacco Industry Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Author Contributions

## **Generative AI statement**

The author(s) declare that no Generative AI was used in the creation of this manuscript

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., and Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Syst. their Appl.* 13, 18–28. doi: 10.1109/5254.708428

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv arXiv:1704.04861*. doi: 10.48550/arXiv.1704.04861

Li, Z., Meng, L., Wang, S., Gao, J., Xu, X., Zhu, X., et al. (2022). Selection of optimum discriminant model in tobacco curing stage based on image processing. *Acta Ta bacaria Sin.* 28, 65.

Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., et al. (2022). "Swin transformer v2: Scaling up capacity and resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. New Orleans, LA: IEEE/CVF. 12009–12019. doi: 10.48550/arXiv.2201.12086 Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/ CVF international conference on computer vision*. Montreal, QC, Canada: IEEE/CVF. 10012–10022.

Lu, M., Wang, C., Wu, W., Zhu, D., Zhou, Q., Wang, Z., et al. (2023). Intelligent grading of tobacco leaves using an improved bilinear convolutional neural network. *IEEE Access* 11, 68153–68170. doi: 10.1109/ACCESS.2023.3292340

Ma, J., Liu, D., Li, Y., Hu, Y., Lu, C., Gai, X., et al. (2021). Construction of benchmark data set for intelligent identific ation of tobacco pests, diseases, phytotoxicity and design of three dimensional attention model. *Acta Tabacaria Sin.* 27, 52–60. doi: 10.48550/arXiv.2106.07178

Pei, W., Zhou, P., Huang, J., Sun, G., and Liu, J. (2024). State recognition and temperature rise time prediction of tobacco curing using multi-sensor data-fusion method based on feature impact factor. *Expert Syst. Appl.* 237, 121591. doi: 10.1016/j.eswa.2023.121591

Selvaraju, R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-cam: Visual explanations from deep networks via gradient-based localization," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy: IEEE. 618–626.

Siddiqui, K. M. (2001). Analysis of a malakisi barn used for tobacco curing in east and southern africa. *Energy Conversion Manage*. 42, 483–490. doi: 10.1016/S0196-8904 (00)00066-2

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv arXiv:1409.1556*. doi: 10.48550/arXiv.1409.1556

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30, 1–15. doi: 10.48550/arXiv.1706.03762

Wang, L., Cheng, B., Li, Z., Liu, T., and Li, J. (2017). Intelligent tobacco flue-curing method based on leaf texture feature analysis. *Optik* 150, 117–130. doi: 10.1016/j.ijleo.2017.09.088

Wang, Y., and Qin, L. (2022). Research on state prediction method of tobacco curing process based on model fusion. *J. Ambient Intell. Humanized Computing* 13, 2951–2961. doi: 10.1007/s12652-021-03129-5

Wu, J., and Yang, S. X. (2019). Intelligent control of bulk tobacco curing schedule using LS-SVM- and ANFIS-based multi-sensor data fusion approaches. *Sensors* 19, 1778. doi: 10.3390/s19081778

Wu, J., and Yang, S. X. (2021). Modeling of the bulk tobacco flue-curing process using a deep learning-based method. *IEEE Access* 9, 140424–140436. doi: 10.1109/ACCESS.2021.3119544

Wu, J., Yang, S. X., and Tian, F. (2014). A novel intelligent control system for fluecuring barns based on real-time image features. *Biosyst. Eng.* 123, 77–90. doi: 10.1016/ j.biosystemseng.2014.05.008

Xu, Q., Zhang, Y., Wang, A., and Guo, W. (2024). "A novel integrated system for real-time monitoring of tobacco leaf images in the bulk curing barn," in 2024 IEEE International Conference on Advanced Intelligent Mechatronics (AIM) (Lyon, France: IEEE), 1573–1578.

Zhang, H., Jiang, X., and Chen, S. (2013). Intelligent tobacco curing control based on color recognition. *Res. J. Appl. Sciences Eng. Technol.* 5, 2509–2513. doi: 10.19026/rjaset.5.4688

Zhang, Y., Zhu, Y., Liu, X., Lu, Y., Liu, C., Zhou, X., et al. (2023). In-field tobacco leaf maturity detection with an enhanced mobilenetv1: Incorporating a feature pyramid network and attention mechanism. *Sensors* 23, 5964. doi: 10.3390/s23135964

Zhao, P., Wang, S., Xia, H., and Dai, Y. (2024). "Recognition of tobacco leaf curing stage based on deep learning," in 2024 10th International Conference on Electrical Engineering, Control and Robotics (EECR), Guangzhou, China: EEE. 305–309.