IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 24 March 2024, accepted 6 April 2024, date of publication 16 April 2024, date of current version 23 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3388418

RESEARCH ARTICLE

SSNet: Exploiting Spatial Information for Tobacco Stem Impurity Detection With Hyperspectral Imaging

CHAO ZHOU[1], ZHENYE LI[1], DONGYI WANG[2], (Member, IEEE), SHENG XUE[1], TINGTING ZHU[1], AND CHAO NI[1]

[1]College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, Jiangsu 210037, China
[2]Department of Biological and Agricultural Engineering, University of Arkansas, Fayetteville, AR 72701, USA

Corresponding author: Tingting Zhu (tingtingzhu@njfu.edu.cn)

This work was supported by the National Natural Science Foundation of China, under Grant 62006120 and Grant 31570714, and additionally by the Jiangsu Province Graduate Student Research and Innovation Program in 2023, grant number KYCX23_1132.

ABSTRACT Tobacco stems are a critical component in cigarette production in China. However, accurately and promptly identifying impurities in tobacco stems remains a challenge with current deep learning methods, often failing to meet industry standards. This study introduces the Spectral-Spatial Network (SSNet), which is specifically optimized for rapid and precise detection of impurities in tobacco stems, leveraging spatial information in hyperspectral imaging. Designed to align with the high-speed requirements of industrial processing, SSNet effectively overcomes the complexities associated with analyzing hyperspectral data. The integration of spatial information particularly enhances recognition accuracy, especially in areas affected by fringe patterns. Our experiments reveal that SSNet significantly outperforms traditional methods in terms of speed and accuracy, marking a substantial advancement in the shift from manual to automated impurity detection in tobacco processing. Furthermore, SSNet's successful application highlights its potential for broader uses in various industrial image processing tasks.

INDEX TERMS Spectral-spatial network, hyperspectral imaging, deep learning, computer vision, tobacco sorting.

I. INTRODUCTION

Chinese tobacco industry, with its extensive brand diversity and prodigious production volume, accounts for approximately one-third of the global tobacco output [1]. However, tobacco stems comprising roughly 25-30% of the tobacco plant's mass are the prevalent underutilization. The researches show that the composition of these stems is congruence with that of the primary tobacco, so that the stems hold potential for conversion into premium cigarette fillers, thereby promising both economic and environmental dividends [2], [3].

In the realm of tobacco stem processing, the industry grapples with the intricate task of segregating and eliminating impurities. As these raw materials traverse the cigarette production trajectory, they are susceptible to degradation and contamination, thereby jeopardizing the end product's integrity. Traditional manual sorting techniques, albeit prevalent, are marred by their laborious nature, time intensiveness, and suboptimal efficiency [4]. The industry's benchmark (YC/T370-2010) meticulously classifies tobacco impurities into three distinct categories, each accompanied by rigorous content stipulations that are arduous to achieve through conventional methodologies. With the escalating demand for tobacco stems juxtaposed against the inadequacies of extant impurity removal techniques such as vibration, wind selection, and laser, there emerges a compelling imperative to enhance the precision of algorithms dedicated to impurity detection in tobacco stems [5].

Historically, a gamut of conventional strategies, encompassing vibration, wind selection, laser, and visual inspections, have been deployed to purge impurities from tobacco

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico.

stems [6]. Each modality, while possessing its unique advantages, is invariably constrained by a shared limitation: the inefficacy in addressing impurities that bear a close resemblance to tobacco stems in their properties. Specifically, laser and visual methodologies, notwithstanding their distinct operational paradigms, are beleaguered by their exorbitant costs, elevated energy demands, and labor intensiveness.

Hyperspectral imaging is an advanced imaging technique that captures information across hundreds of continuous wavelengths, ranging from visible light to infrared. Unlike traditional RGB imaging, which only provides basic color information, hyperspectral imaging records both spatial and spectral information for every pixel [7]. This capability to reveal the chemical and physical properties of materials offers a unique advantage in identifying and classifying substances, particularly when their appearance is similar. The application of hyperspectral imaging technology in detecting impurities of tobacco stems is of paramount importance. Tobacco stems and common impurities, such as paper and plastic films, may look quite similar under visible light, making traditional visual detection methods ineffective. By analyzing the spectral information of each pixel, HSI technology can identify subtle differences between tobacco stems and impurities, significantly improving the accuracy and efficiency of impurity detection. However, this technology is significantly hampered by the interference fringe problem. These interference fringes distort the spectral characteristics, making the accurate differentiation of impurities and tobacco stems a complex challenge. This distortion not only affects the accuracy but also poses a significant hurdle for most conventional and advanced methods [8], [9].

While deep learning techniques, such as CNNs, Transformer, and RNNs, have shown promise in hyperspectral image classification, they too grapple with the interference fringe challenge. Despite their advanced capabilities and successes in various domains, these methods have yet to provide a robust solution to the fringe problem in spectral images commonly encountered in factories [10], [11], [12]. This persistent issue underscores the pressing need for a novel approach that can effectively navigate the interference fringe problem while extracting accurate spectral characteristics.

Interference fringes, caused by lens chromatic aberration, distort hyperspectral signals, concealing the true characteristics of captured objects [13]. While conventional corrections like Multiplicative scatter correction (MSC) have been somewhat effective, they struggle with industrial scenarios where the fringe region can constitute up to 50% of objects like tobacco stems [14]. The high costs of hardware solutions further complicate the issue, emphasizing the need for a software-driven approach to negate this interference fringe efficiently and economically.

In summary, our article makes the following key contributions:

1) We have innovatively expended conventional classification networks for 3-channel images into multi-channel neural networks which is able to handle multi-channel hyperspectral data.
2) Based on residual models, we tailored the conventional neural network to improve overall recognition speed. This pivotal enhancement aligns perfectly with the practical exigencies of industrial production in real-world scenarios.
3) Employing the GA-DT algorithm, we effectively reduced the dimensionality of hyperspectral data, selecting optimal spectral bands combination to align with camera usage requirements. This approach streamlines the data processing pipeline, ensuring that the crucial aspects of spectral data are retained for accurate analysis.
4) We introduce the Fss score, a novel metric that holistically considers both accuracy and processing speed. This metric is pivotal in ensuring that while high accuracy is maintained, the processing speed is also taken into account, striking a balance essential for practical, time-sensitive applications.

II. MATERIALS AND METHODS

A. ACQUISITION OF MATERIALS

The materials used in this study, including tobacco stems and common impurities found in cigarettes produced by a tobacco company in Huai'an, Jiangsu Province, were provided by the company. The Fig.1 depicts the traditional manual sorting process in a tobacco factory, where workers separate impurities from tobacco stems by hand. The tobacco stems and impurities were directly collected from the production line and utilized in our experiments conducted on the actual production system to analyze real-world industrial production scenarios.

FIGURE 1. The image depicts the traditional manual sorting process in a tobacco factory, where workers separate impurities from tobacco stems by hand.

To accurately capture hyperspectral images of the tobacco stems and impurities mixture on a conveyor belt, the imaging system depicted in Fig.2 was employed. This system, comprising a hyperspectral camera (SPECIM FX10E), a conveyor belt, and air nozzles for the air separation of impurities, represents the actual setup used in our experiments. The setup was an augmentation to the existing production line in the factory, specifically designed for this study. The camera was supposed to adopt a commonly used lens having a field of view of 37°. While it could not cover the entire conveyor belt with one camera, making it unsuitable for industrial production. As a result, a 47° wide-angle lens was utilized to capture the entire

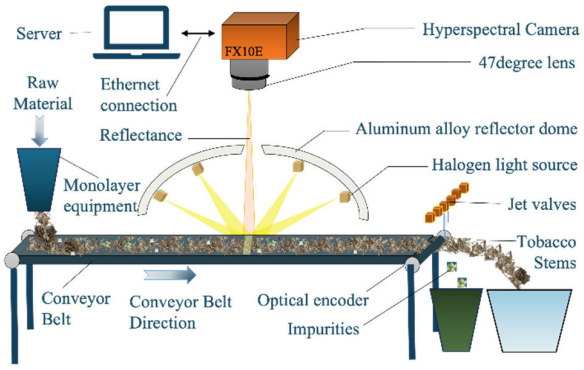


FIGURE 2. Schematic of the imaging system used in the experiments.

conveyor belt, reducing production costs. However, the fringe effect of the wider lens was more noticeable and required careful consideration. The camera had a field of view of 1.2 m, the lens used was 15 mm, and two rows of halogen lamps were used simultaneously to achieve full-range illumination. The captured spectral image size was 600 × 1000× 448 pixels (height × width × bands). The spectral range was 400 ∼ 1000 nm, and the spectral resolution was 2.2 nm. The spectral image was captured at a rate of 30 frames per second. In the case of full-band acquisition, spectral images can be acquired at 30 lines per second, and after band selection, the maximum acquisition speed can reach 200 lines per second.

B. DATA PREPROCESS AND BAND SELECTION FOR HYPERSPECTRAL IMAGES

The black and white correction is a commonly used method to eliminate the influence of the unstable light source and the dark current in the spectral camera for hyperspectral images. The black and white correction [15] is performed by misusing the hyperspectral image by the dark reference image and then dividing the result by the difference between the white and black reference image. The black reference images are captured by the hyperspectral camera when the camera shutter is closed. And the white reference images are captured when the correction board is placed on the conveyor belt. The black and white correction is expressed as follows:

$$I_{corrected} = \frac{I_{raw} - I_{dark}}{I_{white} - I_{dark}}, \tag{1}$$

where $\mathbf{I}_{raw}$ is the raw hyperspectral image, $\mathrm{I}_{\mathrm{dark}}$ is the dark reference image, $\mathrm{I}_{\mathrm{white}}$ is the white reference image, and $\mathbf{I}_{corrected}$ is the corrected hyperspectral image.

Traditional recognition algorithms predominantly rely on the spectral information of individual pixels for classification. This spectral-based classification, often referred to as ''pixel-level classification'' is susceptible to inaccuracies due to the fringe phenomenon. In such scenarios, spectral information from regions of the same category can exhibit variations, leading to compromised recognition accuracy. To address this, our study emphasizes the extraction of spatial features from adjacent areas, aiming to enhance recognition accuracy through the incorporation of domain information.

The extraction process involves constructing a sliding window of shape $n \times n$, where n is an odd number greater than 1. This window slides over the hyperspectral image, collecting neighborhood features [16]. The spatial features gathered by each sliding window are employed to train our model, represented by matrix W:

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \cdots & \cdots & \cdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix} \tag{2}$$

In real-world production scenarios, impurities are considerably less frequent than tobacco stems. Without data augmentation, the initial dataset's significant proportion of tobacco stem categories could bias the learning algorithm. This imbalance could lead to an overemphasis on tobacco stem categories at the expense of impurity categories [17]. To counteract this, we employed data augmentation, a method that synthesizes more data from a limited dataset, thereby enhancing model training.

Given our emphasis on spatial classification accuracy, and considering the time and cost constraints of manual annotation, data augmentation serves as an efficient strategy to optimize our data utilization. For the neighborhood features obtained by each sliding window, represented as matrix W, we applied transformations such as rotations to generate augmented data. These transformations not only expanded the dataset but also enriched the data for the impurity category, ensuring a balanced representation. Such data augmentation techniques bolstered the robustness of our model and diversified the dataset, addressing the challenges posed by the fringe phenomenon.

In our study, the hyperspectral image with 448 wavelengths contains a large amount of redundant information, which will increase the computational complexity of the algorithm and reduce the efficiency of the algorithm. Therefore, it is necessary to select the appropriate band combination to reduce the computational complexity of the algorithm and improve the efficiency of the algorithm to adapt to the industrial application. For this study, we collected and selected a total of 40 hyperspectral images directly from the actual production line, each containing a large variety of tobacco stems and impurities. These images not only reflected the actual conditions of the production environment but also ensured the diversity of the dataset. In order to prevent data leakage and ensure the integrity of our model validation process, the validation dataset was derived from the dataset. The division was based on a random selection from the 40 hyperspectral images collected from the actual production line, ensuring a representation of the actual data distribution.

Feature selection is vital in data analysis, aiming to choose and maintain the most significant features for achieving optimal model performance. In specific scenarios, such as tobacco stem image acquisition where real-time processing is

imperative, the traditional dimension reduction might not be the most suitable approach given the considerable number of wavelengths needed to perform transform in algorithms like Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA). To address this, band selection methods like the Genetic Algorithm (GA) [18], [19] and Greedy Algorithms [20] are often employed.

The genetic algorithm (GA) plays a crucial role in feature selection for our hyperspectral image analysis. GA's power lies in its ability to efficiently navigate through a vast search space to identify the most relevant features, in this case, spectral bands. The encoding process in GA represents potential solutions to the feature selection problem as chromosomes. In our context, a chromosome is a binary string where each bit corresponds to the presence (1) or absence (0) of a particular spectral band.

The crossover operation in GA is a genetic operator used to combine the genetic information of two parents to generate new offspring. It is an analogy to biological reproduction. The specific crossover mechanism we employ in our study is the single-point crossover. Here, a crossover point is chosen randomly within the chromosome, and the genetic information is exchanged between the two parents at this point to produce new offspring. This process allows the GA to explore new regions of the search space and is crucial for the convergence to an optimal set of spectral bands.

The genetic algorithm is a well-established method for feature selection. It is an efficient, parallelizable, and global search technique that continuously garners knowledge about the search space, adaptively guiding the search to locate the best solution. To further refine the selection process, Decision Trees (DT) can be synergized together according to [21]. Unlike GA, which is primarily an optimization tool, DT is a supervised learning method suitable for both classification and regression. By recursively partitioning data into subsets, which are visually represented as tree structures, and recognizing spectral bands with high Gini importance, DT can discern the most relevant bands for classification. In this study, the prowess of both GA and DT is harnessed follow the instruction of [22], with DT being integrated into the GA's cost function, ensuring the selection of the most apt spectral band subset for tobacco stem recognition.

Fig.3 shows the flowchart of the GA-DT algorithm. The GA-DT algorithm is composed of three parts: the initial population, the fitness function, and the genetic operation. The initial population is a set of chromosomes, which are randomly generated. The fitness function is used to evaluate the fitness of the chromosomes. The genetic operation is used to generate the next generation of chromosomes. The GA-DT algorithm is iterated until the termination condition is met. The termination condition is that the number of iterations reaches the maximum number of iterations or the fitness value of the optimal solution reaches the threshold. The GA-DT algorithm is used to select the optimal wavelength combination for tobacco stem recognition.

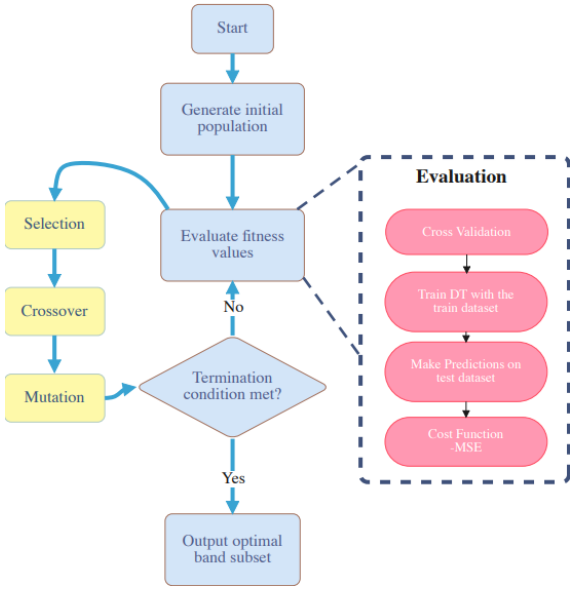


FIGURE 3. The flowchart of the GA-DT algorithm.

Typically, the GA encourages the cost function to become larger, while the DT expects the loss to become smaller. Therefore, the cost function is defined as the negative MSE. The cost function is defined as:

$$f(x) = -\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3}$$

where N is the number of samples, y_i is the actual value of the ith sample, and $\hat{y}_i$ is the predicted value of the ith sample, $f(x)$ is the cost function in this study.

C. ARCHITECTURE OF SSNET

Addressing the challenge of network degradation in traditional neural networks, we introduce SSNet, a novel architecture inspired by the residual structure proposed by Kaiming He et al. [23]. SSNet is specifically optimized for the task of hyperspectral image analysis, particularly for identifying impurities in tobacco stems. This architecture is designed to be both compact and efficient, ensuring robust performance tailored to our unique dataset comprising small, divided images.

The SSNet framework is shown as Fig.4, and it begins with an input layer that is specifically designed to encompass the spatial context of the input pixel. This is achieved by extracting the surrounding pixels and incorporating this spatial information into the network's analysis, as represented by the following equation:

$$X_{out} = \begin{bmatrix} x_{i-k,j-k} & \cdots & x_{i-k,j+k} \\ \cdots & x_{i,j} & \cdots \\ x_{i+k,j-k} & \cdots & x_{i+k,j+k} \end{bmatrix} \tag{4}$$

where $x_{i,j}$ is the central input pixel, and X_{out} is the matrix representing the surrounding pixel information.

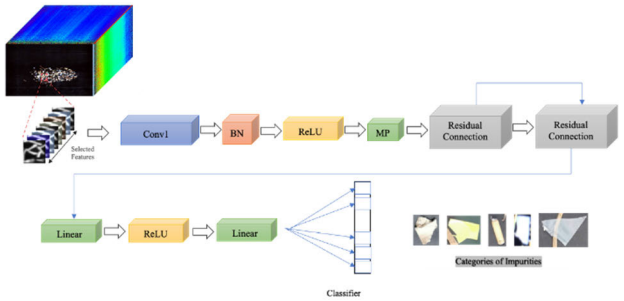


FIGURE 4. The architecture of the proposed SSNet for impurity detection in tobacco stems.

The core of SSNet consists of convolutional layers designed for efficient feature extraction. The first convolutional layer captures the essential features from the input data. This layer is followed by batch normalization and a ReLU activation function to introduce non-linearity, thereby enhancing the model's ability to learn complex patterns. The convolution operation can be mathematically represented as:

$$F\left(X_{in}\right) = \sigma(BN(W * X_{in} + b)) \tag{5}$$

where W and b are the weights and biases of the convolutional layer, σ denotes the ReLU activation function, BN represents batch normalization, and $*$ is the convolution operation.

Subsequent layers include downsampling through Max Pooling and additional convolutional layers with residual connections. These connections help alleviate the vanishing gradient problem, crucial for learning in deeper networks. The output of a residual block in SSNet can be described as:

$$X_{out} = F\left(X_{in}, W\right) + X_{in} \tag{6}$$

where F represents the residual function, and X_{in} and X_{out} are the input and output of the block, respectively.

In the final stages of the network, the feature maps are reshaped and passed through fully connected layers. These layers use dropout for regularization and ReLU activations, culminating in a linear layer that produces the classification scores.

The proposed SSNet focused on the spectral-spatial characteristics of hyperspectral data, which is crucial for precise impurity detection in tobacco stems. By optimizing the input layer to capture the spatial context of surrounding pixels, SSNet aims to enhance its sensitivity to subtle differences. Additionally, the introduction of residual connections aims to address the issue of gradient vanishing in deep networks, ensuring effective information transfer across the network. These design considerations reflect our strategy to tackle the current challenges of hyperspectral imaging, aiming to maintain high accuracy while optimizing processing speed. SSNet stands as a tailored solution for hyperspectral image classification, overcoming challenges inherent in deep neural network models while efficiently addressing the specific needs of our dataset. The details of the SSNet architecture are listed in Table 1.

TABLE 1. The structure parameters of the proposed SSNet neural network.

Layer (type:depth-idx)	Output Shape	Param #
SSNet	[512,6]	2,036,166
Sequential: 1-1	[512, 64, 7, 7]	--
- Conv2d: 2-1	[512, 64, 7, 7]	448
- BatchNorm2d: 2-2	[512, 64, 7, 7]	128
- ReLU: 2-3	[512, 64, 7, 7]	--
- MaxPool2d: 2-4	[512, 64, 7, 7]	--
Conv2d: 1-2	[512, 64, 7, 7]	36,864
BatchNorm2d: 1-3	[512, 64, 7, 7]	128
Conv2d: 1-4	[512, 64, 7, 7]	36,864
ReLU: 1-5	[512, 64, 7, 7]	--
Conv2d: 1-6	[512, 256, 3, 3]	409,600
Conv2d: 1-7	[512, 128, 5, 5]	73,728
BatchNorm2d: 1-8	[512, 128, 5, 5]	256
Conv2d: 1-9	[512, 256, 3, 3]	294,912
ReLU: 1-10	[512, 256, 3, 3]	--
Flatten: 1-11	[512, 2304]	--
Sequential: 1-12	[512, 6]	--
- Linear: 2-5	[512, 512]	1,180,160
- ReLU: 2-6	[512, 512]	--
- Linear: 2-7	[512, 6]	3,078

III. RESULTS AND DISCUSSIONS

A. EXPERIMENTAL ENVIRONMENT AND EVALUATION METRICS

The optimization algorithm for tobacco stalk impurity sorting is written in Python using the image processing library Opencv, machine learning library Scikit-Learn, and deep learning library PyTorch. The software, hardware, and compilation environment configurations used in this study are listed in Table 2.

TABLE 2. Software and hardware environment configurations.

Name	Parameter
System	Linux Ubuntu 20.04
CPU	Intel Core i7-8700 CPU @ 3.2 GHz
GPU	NVIDIA GeForce GTX 2080 Ti
Memory	32 GB
Environment configuration	Python 3.10.0, scikit-learn 1.0.1, opencv-python 4.5.5.64

This study presents a classification scheme for identifying six distinct categories of objects, including background, tobacco stems, cigarette ends, paper, film, and tape. To evaluate the performance of the classification scheme, several standard metrics, including accuracy, precision, recall, and the $F1$ score, are employed. Additionally, we introduce the Fss score, a novel metric designed to factor in both the $F1$ score and the average prediction time per image, thus reflecting the dual importance of accuracy and time efficiency

in object detection. The *Fss* score is defined as:

$$F_{ss} = W_1 \times F1 + W_2 \times (1 - exp(-\frac{1}{T})) \tag{7}$$

where T represents the average prediction time per image. This innovative approach emphasizes the high value placed on rapid processing in our experiments, alongside maintaining strong precision and recall rates.

where the weighting factors W_1 and W_2 are dynamically assigned values of 0.4 and 0.6, respectively. This allocation reflects the specific requirements of industrial applications in the context of tobacco stem impurity detection, where the demand for processing speed slightly outweighs that for accuracy. In industrial settings, timely detection is crucial to ensure efficient production flow and to minimize potential downtime caused by delayed responses. Therefore, a higher weight ($W_2 = 0.6$) is attributed to the time factor, emphasizing the need for rapid processing without significantly compromising accuracy. Conversely, W_1 is set at 0.4, underscoring the importance of maintaining reliable accuracy in classification, albeit with a slightly lower priority compared to processing speed.

In our study, 'impurities' refer to non-tobacco materials such as tape, considered negative samples, contrasting with tobacco stems and background categorized as positive samples. The classification accuracy is articulated through four outcomes: true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN), where TP signifies accurate identification of positives, FP represents mistakenly identified positives, FN indicates missed positives, and TN denotes correct identification of negatives.

The selection of these weights and the categorization methodology are critical in achieving a balanced approach that aligns with the operational tempo and quality standards in the industrial processing of tobacco stems. The formulas for the aforementioned metrics are as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}, \tag{8}$$

$$Precision = \frac{TP}{TP + FP}. \tag{9}$$

$$Recall = \frac{TP}{TP + FN}. \tag{10}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}. \tag{11}$$

In the context of this study, it is imperative to acknowledge the accuracy requirement set forth for the practical application of our system. The experiments were meticulously designed with the aim of achieving an accuracy rate exceeding 96%. This target is derived not arbitrarily but from stringent industry standards and the specific stipulations of our contractual agreement with the partnering company, underscoring the high expectations for performance in real-world industrial settings.

B. EMPLOYMENT OF THE GA-DT ALGORITHM FOR BAND SELECTION

The preprocessing phase of our hyperspectral image analysis begins with a crucial black-and-white correction process, as elucidated in Eq.1. This step aims to normalize spectral values within the range [0,1], ensuring distinctiveness in the spectrum's curve profile, which is fundamental for subsequent analysis stages.

In addressing the challenge of accurately classifying impurities in tobacco stems, we have employed the GA-DT algorithm to select the most suitable spectral bands. This algorithm plays a pivotal role in determining the optimal combination of wavelengths for precise recognition, focusing particularly on pixels outside the fringe region. The underlying parameters of the GA-DT algorithm are detailed in Table 3 for selecting the optimal spectral bands. Here are brief explanations for each parameter:

TABLE 3. Hyperparameters of the GA-DT algorithm.

Hyperparameter	Values
Cross-validation	7-fold
Population size	200
Iteration number	200
Crossover rate	0.5
Mutation rate	0.2
Crossover index	0.5
Mutation index	0.05
Max depth of DT	10

Cross-validation (7-fold): This parameter specifies that the data is divided into seven parts, with six parts used for training and one part for validation in a rotating fashion. This technique helps in assessing the model's performance more reliably.

Population size (200): Indicates the number of candidate solutions (i.e., sets of selected bands) present at each generation within the genetic algorithm. A larger population size allows for a wider search of the solution space but requires more computational resources.

Iteration number (200): The number of generations over which the genetic algorithm will run. This determines how long the algorithm will search for an optimal solution.

Crossover rate (0.5): The probability with which two candidate solutions will be combined to produce offspring. A rate of 0.5 means there's a 50% chance for crossover to occur between any two selected candidates.

Mutation rate (0.2): The probability of randomly altering a gene within a candidate solution. A rate of 0.2 indicates a 20% chance of mutation for any given gene, introducing variability and preventing premature convergence.

Crossover index (0.5) & Mutation index (0.05): These parameters control the behavior of the crossover and mutation operations, respectively. The crossover index influences the blending of parent solutions during crossover, while the mutation index affects the degree of change during mutation.

TABLE 4. The results of ablation experiment for GA-DT.

Index	Methods	Bands (nm)	Accuracy
1		428.125, 438.839, 441.518, 570.089, 620.982, 863.393, 937.054	0.822
2	DA+DT	479.018, 509.821, 529.911, 602.232, 736.161, 745.536, 878.125	0.988
3	DA+SVM [24]	448.214, 629.018, 718.75, 760.268, 841.964, 872.768, 883.482	0.931
4	Greedy algorithm+DT	402.679, 576.786, 761.607, 862.054, 926.339, 945.089, 994.643	0.895
5	Greedy algorithm+SVM	448.214, 629.018, 718.75, 760.268, 841.964, 872.768, 883.482	0.853

Max depth of DT (10): Refers to the maximum depth of the decision tree used in conjunction with the GA for feature selection. A deeper tree can capture more complex patterns but may also lead to overfitting.

To rigorously evaluate the efficacy of our approach, we conducted an ablation experiment, the results of which are presented in Table 4. This experiment aimed to compare different spectral band combinations and their impact on the accuracy of impurity detection. The first entry in the table represents a baseline scenario, where spectral bands were manually selected in a random fashion, showcasing the highest accuracy achieved from 20 different manual selections. The highlighted row in the table underscores the most effective band combination as determined by the GA-DT algorithm, demonstrating a significant improvement in accuracy over the baseline.

The GA-DT algorithm's approach involves encoding potential spectral band combinations as chromosomes and utilizing genetic operations to iteratively select the combination that yields the highest accuracy. The success of this method is evident in the ablation experiment, showcasing the GA-DT algorithm's ability to fine-tune the band selection for enhanced recognition of impurities in tobacco stems. A comprehensive analysis of 448 full-spectrum data points via the genetic algorithm led to the identification of seven optimal spectral bands, pivotal for tobacco stem recognition. These bands, as shown in Fig.5, are at wavelengths: 479.018nm, 509.821nm, 529.911nm, 602.232nm, 736.161nm, 745.536nm, and 878.125nm. The selection of these bands is based on the absorption properties of different substances, ensuring a clear distinction between tobacco stalks and impurities. This optimization reflects a significant advancement over traditional methods, providing a more focused and efficient analysis of hyperspectral images.

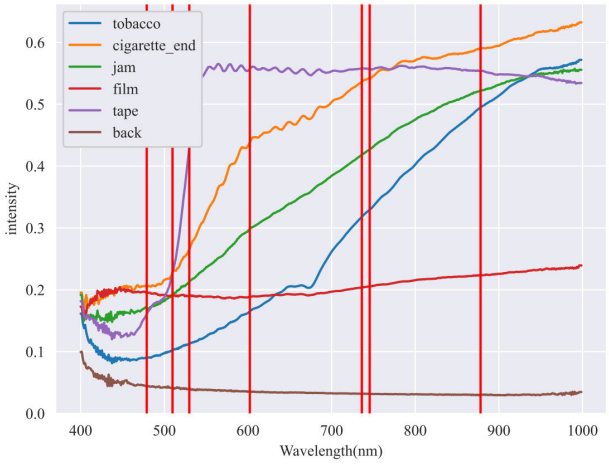


FIGURE 5. Optimal wavelength combination by the GA-DT. The vertical lines denote the bands essential for classifying the specific material, while the curves illustrate the average hyperspectral profile for each material category.

C. COMPARATIVE ANALYSIS OF MODELS PERFORMANCE WITH VARIED SPATIAL CONTEXT

In this study, the effectiveness of incorporating spatial context in hyperspectral image analysis was thoroughly investigated by comparing model performances under three configurations: $k = 1$ using conventional machine learning algorithms, and $k = 2$ and $k = 3$ employing deep learning approaches with the addition of innovative spatial domain information.

For $k = 1$, where no spatial domain information was included, various machine learning algorithms were assessed. These included Decision Tree [25], Random Forest [26], Extremely Randomized Tree [27], AdaBoost [28], and GradientBoost [29]. Among them, the Random Forest algorithm exhibited the best performance, achieving an accuracy of 83%. This superior performance made Random Forest the chosen algorithm for the $k = 1$ configuration. The comparative analysis of these algorithms is summarized in Table 5.

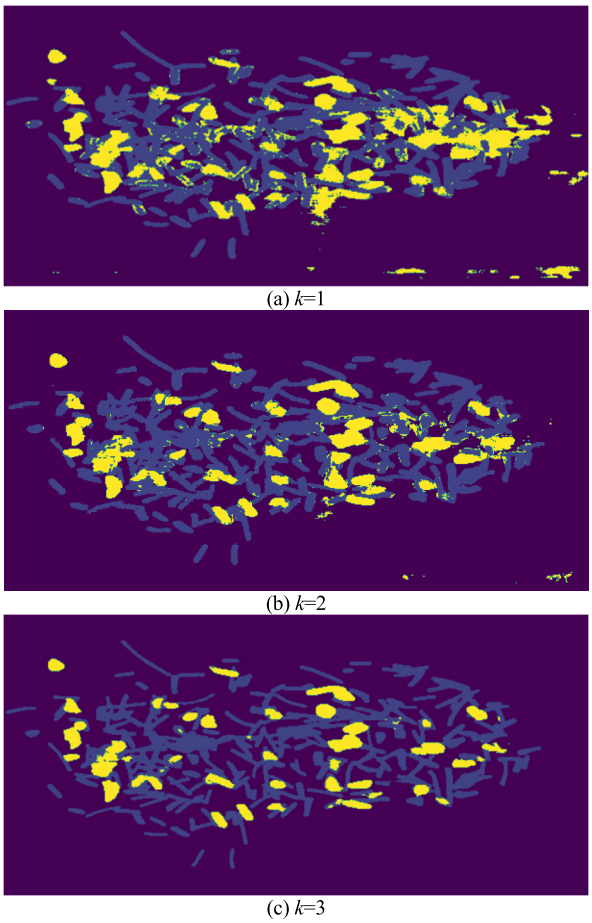
TABLE 5. Machine learning algorithms prediction metrics.

Methods	Accuracy	Precision	Recall	F1-score
DecisionTree	71%	73%	71%	71%
RandomForest	**83%**	**85%**	**83%**	**83%**
ExtraTrees	80%	84%	80%	80%
AdaBoost	74%	75%	74%	74%
GradientBoost	74%	74%	74%	74%

The study progressed to configurations $k = 2$ and $k = 3$, wherein deep learning models were enhanced with spatial domain information. This addition aimed to leverage the spatial context, which is crucial for accurately interpreting complex patterns within the data. The inclusion of this spatial information significantly improved the classification accuracy. However, configurations with k values higher than 3 were not tested due to hardware limitations. Specifically, larger values of k would have required an inordinate amount

of memory allocation during training, exceeding the capabilities of the available hardware, even with reduced batch sizes.

Fig.6 presents the comparative results for $k = 1$, 2, and $k = 3$, with the $k = 3$ configuration demonstrating the most accurate predictions. The enhanced performance of $k = 3$ underscores the importance of spatial domain information in deep learning models for hyperspectral image analysis, particularly in overcoming challenges such as fringe effects.



(a) k=1

(b) k=2

(c) k=3

FIGURE 6. Comparative results for configurations with k = 1, 2 and 3.

In conclusion, the experimental results highlight the critical role of spatial context in hyperspectral image classification. The conventional machine learning algorithms at $k = 1$, although less computationally demanding, were limited in handling complex image patterns. In contrast, the deep learning models at $k = 2$ and $k = 3$, augmented with spatial domain information, showed a marked improvement in accuracy, validating the proposed method's effectiveness.

D. COMPARATIVE EVALUATION OF DEEP LEARNING MODELS IN INDUSTRIAL CONTEXT

This study evaluates the performance of the proposed model and the other three convolutional neural networks—ResNet, MobileNetv2 [30], and ShuffleNetv2 [31]—in addressing the fringe problem in hyperspectral imaging, and the results are

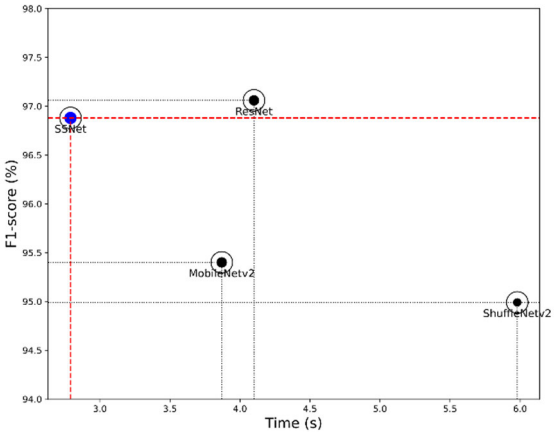


FIGURE 7. Comparative performance analysis of deep learning models.

listed in Table 5. The analysis is crucial in determining the most effective model for industrial applications, particularly in impurity removal from tobacco stems.

Given the industry's requirement for high accuracy due to the potential carry-over of tobacco stems during impurity removal, models need to achieve an accuracy significantly above the standard 95%. This is to compensate for the approximately 1% carry-over rate. Our results show that ResNet consistently exceeds this accuracy threshold, demonstrating its suitability for such industrial applications.

In this study, the distance between the camera and the nozzle is approximately 2 meters, and the conveyor belt speed is 0.7 m/s. Therefore, the processing speed needs to be within 2.87 seconds. While ResNet demonstrated high accuracy in addressing the fringe problem, its processing speed did not meet this requirement, prompting the need for a more efficient solution.

Recognizing the need for a network that combines the accuracy of ResNet with faster processing times, we developed SSNet, an optimized structure based on ResNet. This innovation aimed to achieve high accuracy in addressing fringe problems, similar to ResNet, but with significantly improved processing efficiency. Comparing ResNet and the newly designed SSNet in Table 6, these results highlight SSNet's superior performance, particularly in the Fss score, making it a standout choice among other networks.

Fig 8 presents a detailed comparison of four neural network models, including SSNet, by plotting F1-scores against processing times. This graph illustrates how SSNet not only achieves a F1-score competitive with that of the high-performing ResNet but also significantly reduces processing times across all models evaluated. Such a combination of high F1-score and reduced processing time is a unique attribute of SSNet, illustrating its capacity to maintain precision in results while enhancing operational efficiency. Capturing this dual advantage in a single metric is challenging, but the incorporation of the Fss score into the visualization offers a novel perspective.

TABLE 6. Performance comparison of neural networks, focusing on industrial application requirement.

Methods	Parameters	Accuracy	Precision	Recall	F1-score	Time (s)	Fss Score
MobileNetv2	3,225,222	95.47%	95.34%	95.34%	95.40%	3.87	0.5182
ShuffleNetv2	**1,260,618**	95.01%	94.56%	94.97%	94.99%	5.98	0.4723
ResNet	4,903,494	**97.08%**	**97.13%**	**97.05%**	**97.06%**	4.10	0.5181
SSNet	2,036,166	96.82%	96.82%	96.94%	96.88%	**2.79**	**0.5682**

TABLE 7. Comparative performance of SSNet across different batches of tobacco stems.

Batch	Accuracy	Precision	Recall	F1-score	Time (s)	Fss Score
Yunnan	96.82%	96.82%	96.94%	96.88%	2.79	0.5682
Guizhou	96.85%	96.80%	97.00%	96.90%	2.75	0.5705

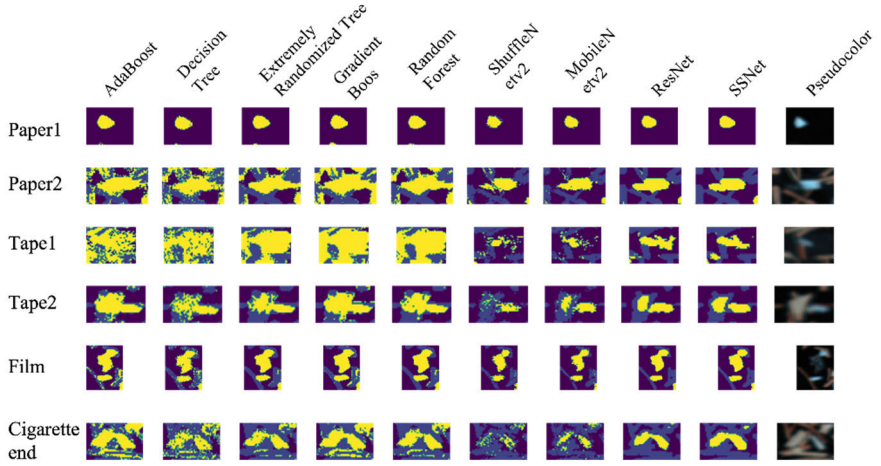


FIGURE 8. The detailed comparison of the prediction results.

In the graph, each model's performance is denoted by a solid plot point, where the size of each point reflects the Fss score, providing a visual indication of each model's balance between accuracy (as measured by the F1-score) and speed. Surrounding each solid plot point is a hollow circle, representing what the size of the point would be if the Fss score were perfect, at 1.0. This visual aid underscores the relative performance of each model in terms of this composite metric, with SSNet's plot point being notably larger within its circle, indicating a higher Fss score. This signifies SSNet's superior ability to offer an optimal blend of speed and precision.

The Fss score, depicted through the differential sizing of plot points, is designed to encapsulate the essential balance between F1-score precision and processing speed. Such balance is critical in industrial applications where decisions must be made rapidly without sacrificing accuracy. Traditional metrics often prioritize one factor over the other, but the Fss score provides a composite measure that equally values both speed and accuracy. Through Fig 8, the utility of the Fss score in offering a nuanced comparison of model performances becomes evident, with SSNet's enhanced plot point size demonstrating its effectiveness in delivering prompt and precise results. This characteristic is especially valuable in real-world scenarios, where the demands for both rapid processing and high accuracy are uncompromising.

In order to conduct a comprehensive comparison between the classification results of the proposed method and those of the state of art strategy, we selectively extracted small regions containing various types of impurities from the original image and compared their corresponding prediction outcomes. Fig. 7 illustrates these comparative analyses. Each vertical line embodies the prediction results yielded by different classification methods, while each horizontal line represents the prediction results for the same type of impurity. Overall, the proposed method excels in accurately identifying the impurities, even within areas affected by fringes.

In the analysis of Impurity Paper 1, all classification approaches delivered precise and unambiguous predictions. This degree of accuracy is partly due to the paper's distinctive white coloration, which facilitates its detection. However, the performance of these models becomes inconsistent when tasked with classifying Impurity Paper 2. Our proposed

method demonstrates exceptional accuracy in identifying Impurity Paper 2, whereas alternative methods often yield lower levels of accuracy. This disparity may be attributed to the light coloring of the surrounding tobacco stems, thereby increasing the probability of the paper being misclassified. Machine learning algorithms exhibit a higher frequency of misclassified pixels, with deep learning-based methods providing superior results in this classification endeavor.

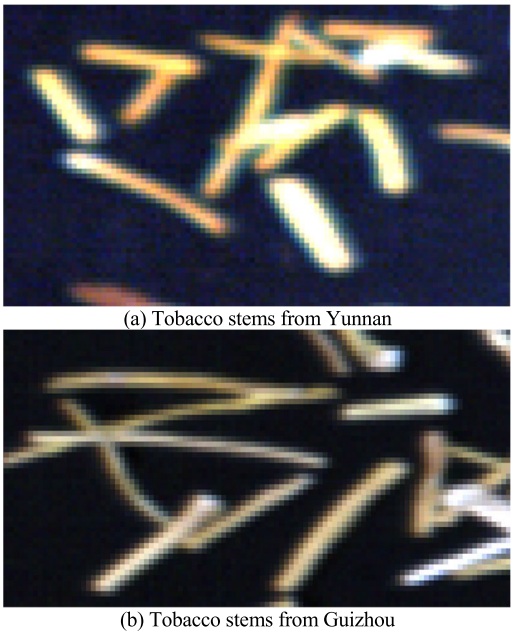
In terms of the classification outcomes of Tape 1 and Tape 2, it is evident that our proposed method delivers a more holistic set of results. This method accurately identifies Tape 1 and Tape 2, while the precision of other methods leaves room for improvement. The tape's thin profile and light coloration increase the risk of its misclassification as part of the background. It is clear that machine learning algorithms tend to produce a higher quantity of misclassified pixels, whereas deep learning-based methods demonstrate superior performance in the classification task.

With respect to the recognition of the film, all methods distinctly identify the film target, with the proposed method indicating fewer misclassified pixels. In the process of recognizing the cigarette end, even other deep learning methods fail to deliver a comprehensive recognition outcome due to the color closely resembling that of the tobacco stem. However, our proposed method presents fewer misclassified pixels as it is capable of merging spatial and spectral information, thereby enabling more precise predictions.

To further validate the robustness and general applicability of SSNet, we extended our experimental analysis to include tobacco stem samples from two distinct regions, Yunnan and Guizhou provinces. The TABLE 7 shows SSNet's performance metrics across batches collected from Yunnan and Guizhou. Despite the differences in moisture content and other regional characteristics, SSNet demonstrated remarkable stability and efficiency, with accuracy, precision, recall, and F1-scores consistently maintained at high levels. The minimal variance observed in these metrics underscores the algorithm's ability to adapt to diverse conditions without compromising on performance.

Further illustrating SSNet's adaptability, Fig 9 presents hyperspectral images of tobacco stems from Yunnan (upper) and Guizhou (lower), visually representing the algorithm's capability to accurately classify impurities across batches with inherent regional variability. This visual evidence, coupled with our quantitative analysis, reinforces SSNet's potential as a tool for the tobacco industry, capable of delivering rapid and precise sorting of tobacco stems under varying conditions.

In conclusion, the development of SSNet represents a significant stride in hyperspectral image classification. It successfully balances the need for high accuracy in complex imaging scenarios, like fringe effects, with the industrial requirement for quick processing. The comprehensive comparison underscores SSNet's suitability for practical applications, affirming its role as a more appropriate model for hyperspectral imaging challenges.



(a) Tobacco stems from Yunnan

(b) Tobacco stems from Guizhou

FIGURE 9. Hyperspectral images of tobacco stems from Yunnan and Guizhou.

IV. CONCLUSION

In this study, we proposed SSNet, a detection algorithm designed to improve the accuracy of impurity detection in tobacco stems through hyperspectral imaging. Our evaluation demonstrates SSNet's superiority over established networks like MobileNetv2 and ShuffleNetv2, offering a better trad-off between precision and speed, particularly in addressing the complex challenges posed by fringe effects. The main contribution of this work can be summarized as below:

1) By integrating the neighboring Spectral-Spatial information from the hyperspectral pixel, we improved the identification accuracy of the fringe area.
2) The conventional 3-channel image-based network is expanded into a multi-channel neural to handle multi-channel hyperspectral data.
3) The conventional neural network is tailored to be better fit the real-world scenarios.
4) The GA-DT algorithm is proposed to reduce the dimensionality of hyperspectral data, search the optimal combination for classification.
5) The Fss score is introduced to describe the balance between higher accuracy and faster recognition speed.

We have successfully addressed the limitations encountered in our study, especially the challenges of fringe effect, we propose the exploration and integration of higher-resolution cameras to complement the current hyperspectral imaging setup. This approach could allow for a more detailed capture of spectral data, facilitating the finer differentiation between impurities and tobacco stems. Additionally, we advocate for the development of algorithms that leverage prior knowledge about the spectral characteristics of tobacco stems and common impurities. These strategies aim

to enhance the robustness of impurity predictions, making the detection process more sensitive to minute discrepancies. However, the proposed method has a limitation on recognizing extremely small impurities.

In the future, enhancing the resolution of hyperspectral imaging and developing more sophisticated algorithms that incorporate prior spectral knowledge could help overcome these challenges. Moreover, expanding the diversity of the dataset to include a broader range of impurity types and sizes will be crucial for improving the robustness and applicability of SSNet.

ACKNOWLEDGMENT

(Chao Zhou and Zhenye Li contributed equally to this work.)

REFERENCES

[1] G. Yuan, M. Lu, K. Dou, J. Yu, H. Zhuang, S. Sheng, Y. Zhu, Y. Liu, X. Li, and L. Jiang, "Factors affecting the caring strength of the sheet base of reconstituted tobacco of cigarettes," *Acta Tabacaria Sinica*, vol. 29, no. 4, pp. 1–9, 2023.

[2] J. Li, Z. Gao, F. Yang, and M. Zhou, "Effect of tobacco distribution in tobacco sticks on cigarette quality," *Tobacco Technol.*, pp. 11–12+15, Aug. 2004.

[3] M. Zhao, X. An, Z. Fan, L. Liu, P. Yin, S. Wu, Y. Liu, K. Du, and H. Liu, "Research progress on pretreatment of 'Tobacco stem," *Trans. China Pulp Paper*, vol. 37, pp. 99–110, Sep. 2022.

[4] M. Chao, C. Kai, and Z. Zhiwei, "Research on tobacco foreign body detection device based on machine vision," *Trans. Inst. Meas. Control*, vol. 42, pp. 2857–2871, Nov. 2020.

[5] W. Fan, "Study on countermeasures of promoting rural vitalization strategic development by tobacco production in China," *Tobacco Regulatory Sci.*, vol. 7, no. 5, pp. 2385–2391, Sep. 2021.

[6] T. Jun, H. E. Banghua, Y. I. Bin, L. I. N. Wenqiang, M. A. Ning, T. Li, C. Wen, and Z. Bing, "Research progress of tobacco foreign material eliminating technology for cigarette processing in China," *J. Light Ind.*, vol. 37, no. 2, p. 94, Apr. 2023.

[7] A. Benelli, C. Cevoli, A. Fabbri, and L. Ragni, "Ripeness evaluation of kiwifruit by hyperspectral imaging," *Biosyst. Eng.*, vol. 223, pp. 42–52, Nov. 2022.

[8] Y. Ding, Z. Zhang, X. Zhao, D. Hong, W. Li, W. Cai, and Y. Zhan, "AF2GNN: Graph convolution with adaptive filters and aggregator fusion for hyperspectral image classification," *Inf. Sci.*, vol. 602, pp. 201–219, Jul. 2022.

[9] B. Lu, P. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, p. 2659, Aug. 2020.

[10] R. Moyazzoma, Md. A. A. Hossain, Md. H. Anuz, and A. Sattar, "Transfer learning approach for plant leaf disease detection using CNN with pre-trained feature extraction method Mobilnetv2," in *Proc. 2nd Int. Conf. Robot., Electr. Signal Process. Techn. (ICREST)*, Jan. 2021, pp. 526–529.

[11] R. Reedha, E. Dericquebourg, R. Canals, and A. Hafiane, "Transformer neural network for weed and crop classification of high resolution UAV images," *Remote Sens.*, vol. 14, no. 3, p. 592, Jan. 2022.

[12] Y. Ding, X. Zhao, Z. Zhang, W. Cai, N. Yang, and Y. Zhan, "Semi-supervised locality preserving dense graph neural network with ARMA filters and context-aware learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2021.

[13] L. Zhang, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Purple fringing correction method based on color information of object," *Acta Optica Sinica*, vol. 36, no. 12, 2016, Art. no. 1233001.

[14] P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, and D. N. Rutledge, "New data preprocessing trends based on ensemble of multiple preprocessing techniques," *TrAC Trends Anal. Chem.*, vol. 132, Nov. 2020, Art. no. 116045.

[15] S. Weng, P. Tang, H. Yuan, B. Guo, S. Yu, L. Huang, and C. Xu, "Hyperspectral imaging for accurate determination of Rice variety using a deep learning network with multi-feature fusion," *Spectrochimica Acta A, Mol. Biomolecular Spectrosc.*, vol. 234, Jun. 2020, Art. no. 118237.

[16] M. Imani and H. Ghassemian, "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges," *Inf. Fusion*, vol. 59, pp. 59–83, Jul. 2020.

[17] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks," *Sci. Rep.*, vol. 9, no. 1, p. 16884, Nov. 2019.

[18] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: Past, present, and future," *Multimedia Tools Appl.*, vol. 80, pp. 8091–8126, Feb. 2021.

[19] M. Sharma and P. Kaur, "A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem," *Arch. Comput. Methods Eng.*, vol. 28, no. 3, pp. 1103–1127, May 2021.

[20] Z. Zhao, S. Liu, M. Zhou, D. You, and X. Guo, "Heuristic scheduling of batch production processes based on Petri nets and iterated greedy algorithms," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 1, pp. 251–261, Jan. 2022.

[21] I. S. Damanik, A. P. Windarto, A. Wanto, Poningsih, S. R. Andani, and W. Saputra, "Decision tree optimization in C4.5 algorithm using genetic algorithm," *J. Phys., Conf. Ser.*, vol. 1255, no. 1, Aug. 2019, Art. no. 012012.

[22] R. N. Toma, A. E. Prosvirin, and J.-M. Kim, "Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers," *Sensors*, vol. 20, no. 7, p. 1884, Mar. 2020.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[24] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.

[25] A. Tariq, J. Yan, A. S. Gagnon, M. R. Khan, and F. Mumtaz, "Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest," *Geo-Spatial Inf. Sci.*, vol. 26, no. 3, pp. 302–320, 2022.

[26] Q. Hu, X.-S. Si, Q.-H. Zhang, and A.-S. Qin, "A rotating machinery fault diagnosis method based on multi-scale dimensionless indicators and random forests," *Mech. Syst. Signal Process.*, vol. 139, May 2020, Art. no. 106609.

[27] Y.-B. Zhao, Y. Song, F.-F. Li, and X.-L. Yan, "Prediction of mechanical properties of cold rolled strip based on improved extreme random tree," *J. Iron Steel Res. Int.*, vol. 30, pp. 293–304, Jul. 2022.

[28] M. Mazini, B. Shirazi, and I. Mahdavi, "Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 31, no. 4, pp. 541–553, Oct. 2019.

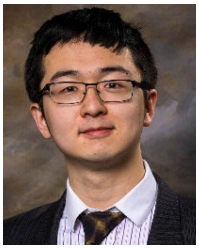
[29] M. Arif, S. Ahmad, F. Ali, G. Fang, M. Li, and D.-J. Yu, "TargetCPP: Accurate prediction of cell-penetrating peptides from optimized multi-scale features using gradient boost decision tree," *J. Computer-Aided Mol. Design*, vol. 34, no. 8, pp. 841–856, Aug. 2020.

[30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[31] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 116–131.

CHAO ZHOU was born in Yangzhou, Jiangsu, China, in 1999. He received the bachelor's degree in engineering from Nanjing Forestry University, in 2021, where he is currently pursuing the master's degree in control science and engineering. His research interests include artificial intelligence in data processing and image processing.

ZHENYE LI was born in Yangzhou, Jiangsu, China, in 1997. He received the bachelor's degree in engineering from Nanjing Forestry University, in 2019, where he is currently pursuing the master's degree. His research interests include the application of intelligent algorithms in industry, image data processing, and hyperspectral analysis.



DONGYI WANG (Member, IEEE) received the B.S. degree in electrical engineering from Fudan University, China, and the Ph.D. degree from the Department of Bioengineering, University of Maryland, College Park, MD, USA. He was a Visiting Student with the Department of Electrical Engineering, The Chinese University of Hong Kong, in 2014. He is currently an Assistant Professor of biological and agricultural engineering with the University of Arkansas, Fayetteville. His research interests include computer vision, artificial intelligence, robotics, hyperspectral imaging analysis, food, and healthcare automation.

SHENG XUE was born in Huaian, Jiangsu, China, in 2000. He received the bachelor's degree in engineering from Nanjing Forestry University, in 2022, where he is currently pursuing the master's degree. His research interests include the application of intelligent algorithms in industry and image data processing.

TINGTING ZHU received the Ph.D. degree in pattern recognition and artificial intelligence from the School of Automation, Southeast University, in 2019. She was a Visiting Student with the Department of Atmospheric and Oceanic Sciences, McGill University, Canada, from 2017 to 2018. She is currently an Associate Professor with the College of Mechanical and Electronic Engineering, Nanjing Forestry University, China. Her current research interests include machine learning, data processing and modeling, renewable energy generation forecast, and climate feedback. She achieved a fellowship jointly awarded by Fonds de Recherche du Quebec—Nature et Technologies (FRQNT) and China Scholarship Council.

CHAO NI was born in Nanjing, Jiangsu, China, in 1979. He received the B.S. degree in automation from Nanjing University of Science and Technology, Nanjing, in 2001, and the Ph.D. degree in control theory and control engineering from Southeast University, Nanjing, in 2008. From October 2017 to November 2018, he was a Visiting Scholar with the University of Maryland, College Park, MD, USA. He is currently a Professor with the Automation Department, Nanjing Forestry University, China. His research interests include artificial intelligence in industrial application, data processing, and spectroscopy analysis.