Research Article

Evaluation of Flavor Type of Tobacco Blending Module: A Prediction Model Based on Near-Infrared Spectrum

Lin Wang ⓘ,[1] Yuhan Guan ⓘ,[2] and Yaohua Zhang[1]

[1]*Technology Center, China Tobacco Hubei Industrial Co., Ltd., Wuhan 430040, China*
[2]*School of Management, Huazhong University of Science and Technology, Wuhan 430074, China*

Correspondence should be addressed to Yuhan Guan; gyhhh@hust.edu.cn

Near-infrared spectrum technology is extensively employed in assessing the quality of tobacco blending modules, which serve as the fundamental units of cigarette production. This technology provides valuable technical support for the scientific evaluation of these modules. In this study, we selected near-infrared spectral data from 238 tobacco blending module samples collected between 2017 and 2019. Combining the power of XGBoost and deep learning, we constructed a flavor prediction model based on feature variables. The XGBoost model was utilized to extract essential information from the high-dimensional near-infrared spectra, while a convolutional neural network with an attention mechanism was employed to predict the flavor type of the modules. The experimental results demonstrate that our model exhibits excellent learning and prediction capabilities, achieving an impressive 95.54% accuracy in flavor category recognition. Therefore, the proposed method of predicting flavor types based on near-infrared spectral features plays a valuable role in facilitating rapid positioning, scientific evaluation, and cigarette formulation design for tobacco blending modules, thereby assisting decision-making processes in the tobacco industry.

1. Introduction

Cigarette products need to go through tobacco blending module formula design, which is crucial in creating cigarettes with distinct flavors. It means that different modules are mixed together in a certain proportion, and spices are added for seasoning, so as to form cigarettes of various specifications [1]. The blending module of the tobacco roll group is the basic unit of cigarette cut, and the sensory quality evaluation of the blending module can better guide its use in cigarette. Different blending modules can be obtained by mixing flue-cured tobacco leaves of different varieties, origin, position, year, and grades. These modules, characterized by distinct flavors and grades, are mixed according to the requirements of cigarette products. Among sensory quality evaluation indices, the flavor of tobacco blending module is a crucial factor affecting the aroma style of cigarette and plays an important role in cigarette formula design and product maintenance [2]. The quality of formulations depends on the quality and synergy of the modules utilized. By appropriately combining

modules of various flavor types and grades that demonstrate synergistic effects, superior quality outcomes can be attained [3]. The flavor of module can be divided into three categories: clear flavor, intermediate flavor, and luzhou flavor [4]. The luzhou flavor module features a noticeable aroma with high concentration and a strong lingering sensation. It leaves a strong sweet aftertaste, but with some noticeable off-flavors. The clear flavor module is refreshing, but the aftertaste may be slightly less comfortable. The intermediate flavor module has a lower smoke concentration and intensity, making it ideal as a filling agent in the formulation. It effectively dilutes the smoke concentration and intensity, contributing to the overall balance of the blend [5]. By conducting flavor evaluation, the design of formula modules can be further optimized. This optimization not only enables each flavor module to complement one another and emphasize the aroma style characteristics of core modules, resulting in an enhanced overall sensory quality of cigarette products, but also effectively reduces the cost of cigarette production. Ultimately, this optimization aids in improving resource utilization rates and

reducing the cost of raw materials. Currently, the main methods for industry to evaluate the sensory quality characteristics of formula modules are still manual evaluation and subjective judgment [6]. In this approach, formulators assess the stylistic characteristics of modules by drawing upon their subjective experience and perception, complemented by the results of chemical measurements. However, this evaluation method possesses inherent limitations. On the one hand, it exhibits strong personal subjectivity, lacks standardized criteria, and fails to guarantee consistent evaluation results. On the other hand, the process itself is intricate, often requiring multiple individuals to evaluate a single module. Each evaluation entails procedures such as processing, grouping, laboratory sample preparation, and inspection, which can be costly. As a result, there is an urgent need for a more objective and accurate method to determine the quality of formula modules, particularly for agricultural products like tobacco, where quality is variable and the primary factors influencing product quality remain undetermined. Such a method would optimize the entire module evaluation process and enable comprehensive and precise quality determination.

In recent years, propelled by the rapid advancements in computer technology, experts in the tobacco industry have been actively exploring innovative and objective methodologies for determining flavor types. These methodologies encompass both qualitative and quantitative analyses. Researchers have utilized statistical analysis techniques to aid in categorizing modular flavors based on the measurement of specific chemicals or groups of chemicals, such as free amino acid content [7], aroma activity value [8], aromatic compound type [9], and sugar and nicotine contents [10]. Through laboratory testing, they examined the chemical content and subsequently established a relationship between the chemical composition and flavor type using various models, including hierarchical analysis and correlation analysis [11], PLS [12], SVM [13], and PCA [14]. The aforementioned process has achieved a certain level of scientific evaluation. However, it is hindered by the labor-intensive nature of chemical detection, potential sample damage, high costs, and the need for specialized expertise, resulting in limited application. In this regard, the utilization of near-infrared spectroscopy for detection proves to be a favorable approach.

Near-infrared spectroscopy (NIRS) is a widely employed technique due to its convenience, stability, and cost-effectiveness [15]. It enables rapid and nondestructive detection of sample compositions and properties. By establishing a model correlating the intensity of spectral characteristic bands with the samples under investigation, researchers can determine the product quality. NIRS is particularly effective in identifying the composition and structure of organic matter, including agricultural products and petrochemicals, as it aligns with the vibration frequency and absorption area of hydrogen groups, as well as the molecular structure containing components such as tobacco sugar, nicotine, and protein [16]. In the field of tobacco research, a multitude of scholars have engaged in the modeling of NIRS analysis to discern and uncover latent characteristics embedded within samples. Zhang et al. [17] used 1D and 2D CNNs to extract NIRS's features and establish a tobacco origin identification model,

which has accuracy rate up to 90%. Chen et al. [18] employed all preprocessed features as input directly into CNN for modeling to determine the maturity of tobacco leaves, enabling automatic feature extraction within the CNN layer. Wei et al. [19] utilized the deep transfer learning approach to extract features and model the infrared data by fine-tuning a pretrained CNN model. This model was employed to predict key component parameters, such as nicotine and sugar content, during the flue-curing process of tobacco leaves and exhibited robustness and achieved accurate real-time monitoring of tobacco leaf composition changes. Borges-Miranda et al. [20] performed regression on the intensity of 33 variables and 1050 NIR reflectance of cigars to overcome the subjectivity of raw material selection of high-grade cigars; they also calibrated and verified the model by partial least squares and support vector regression algorithm. Jiang et al. [21] introduced a regression approach based on a one-dimensional full convolutional network for the quantitative analysis of nicotine components in tobacco leaves. In this approach, a convolutional layer was employed to substitute the maximum pooling layer, thus mitigating information loss. They also proposed a classification model for tobacco cultivation regions that combined ResNet and NIRS [22]. This innovative approach effectively mitigates the vanishing gradient issues that arise from network depth expansion. Zhu et al. [23] presents a method, called TCCANN, for quantitatively analyzing the chemical components of tobacco leaves using NIRS. The TCCANN combines ResNet and LSTM neural networks to address the gradient-disappearance issue and enable simultaneous analysis of multiple chemical compositions. In general, these studies can be conducted from qualitative or quantitative and undergo the following steps [24]. Initially, NIRS data are gathered, followed by meticulous preprocessing to eliminate noise and interference. Subsequently, feature band extraction is performed to identify key spectral features that capture important information about the sample's properties. The extracted features are then used to develop a predictive model through modeling techniques, such as machine learning algorithms or chemometric methods. To ensure the reliability and accuracy of the model, comprehensive testing is conducted using independent validation datasets, assessing its performance and generalizability across different samples and conditions. This rigorous process enables the establishment of robust findings, providing valuable insights and actionable information for further analysis and decision-making [25–28].

Although NIRS gained wide acceptance in the field of tobacco research, its application has primarily focused on studying the spectral information of tobacco leaves and cigarettes themselves. The exploration of the potential of analyzing individual components within tobacco blending modules, which are directly associated with cigarettes and serve as their fundamental units, has been limited [28, 29]. Moreover, researchers have often connected NIRS with chemical information or linked chemical information with sensory evaluation, but there is a lack of direct integration between NIRS and sensory information, which hinders a comprehensive understanding of the flavor characteristics of tobacco blending modules [30–32]. Furthermore, it is worth noting that in conventional practices, the data are

often input or processed without undergoing adequate preprocessing steps. This approach of directly using or processing the data without proper preprocessing is limited in its effectiveness, especially in the presence of various types of interference information in complex samples [33]. In addition, methods based on traditional statistical learning approaches, including regression, correlation analysis, and principal component analysis, have inherent limitations and cannot fully establish the comprehensive relationship between variables [34]. Hence, there is a need to enhance the accuracy and predictive power of models used in the sensory evaluation of tobacco blending modules.

In this study, we aim to address these challenges by leveraging the NIRS data of tobacco blending modules to develop a sensory quality prediction model. To eliminate different types of interference information, we employ a series of preprocessing combinations and compare their effectiveness. In the model building phase, we utilize an improved residual network module to construct a neural network model. To enhance the stability of the network, we incorporate layer normalization techniques. In addition, label regularization is employed during the calculation of the loss function to accelerate convergence and improve the generalization capability of the model. To evaluate the performance of our proposed model, we conduct comprehensive experiments and comparisons. The results demonstrate the exceptional predictive power of the model, with an accuracy rate reaching 91.46%. This highlights the effectiveness and reliability of the model in accurately predicting the flavor type of tobacco blending modules. The main contributions of this paper are as follows:

(1) Integration of NIRS data directly with modules' sensory information: Our approach surpasses traditional practices by directly incorporating NIRS data with blending modules' sensory information. This integration enables a more comprehensive analysis and objective evaluation of tobacco blending modules, which can provide better guidance for their utilization in cigarettes. In addition, by harnessing the combined power of NIRS data integration, advanced preprocessing, and neural network modeling, we streamline the evaluation process, facilitating more efficient decision-making in the tobacco industry.

(2) Utilization of an improved residual network architecture: During data preprocessing, we effectively eliminate various forms of interference information by employing a combination of diverse preprocessing methods. To build our predictive model, we leverage an enhanced residual network module, which ensures more precise predictions. Furthermore, we incorporate layer normalization techniques to stabilize the network and apply label regularization to expedite convergence, further enhancing the model's performance.

Overall, our innovative approach offers significant advancements in the evaluation of tobacco blending modules, opening up new possibilities for quality assessment and optimization in this domain. The remaining sections of this paper are organized as follows. Section 2 presents the methodology employed in this study; Section 3 delves into a comprehensive discussion of the experimental procedures; Section 4 shows the comparative experiments and analyzes the related results; Section 5 concludes this paper.

2. Methods

As an indirect measurement method, NIRS does not provide direct predictions of the content or category of a specific substance in the sample. Instead, it relies on stoichiometry to establish an association model for prediction. The utilization of NIRS data for model development involves several essential processes that are discussed in the introduction, including spectrum pretreatment, band selection, and model selection. These processes play a crucial role in enhancing the accuracy and reliability of the predictive model.

2.1. Preprocessing Methods. Due to various factors such as sample size, environmental conditions, and human operation, the NIRS data obtained often contain significant amounts of noise and irrelevant data. Moreover, the presence of stray light and baseline drift can further contribute to fluctuations and distortions in the original data. This can lead to the misinterpretation of certain trend items as genuine spectral data during the model construction process, consequently impacting the accuracy of the model [35]. To address these challenges, preprocessing of the spectrum is essential. However, it is important to acknowledge that there is no one-size-fits-all pretreatment method that can be universally applied in all scenarios. The choice of suitable pretreatment methods should be approached with careful consideration, taking into account the specific characteristics of the data, the objectives of the analysis, and the models being employed. Commonly used NIRS pretreatment methods include standardization, smoothing, trend correction, and derivation [36].

Standardization is a pretreatment method that aims to transform spectra into a standardized scale by subtracting the mean and dividing by the standard deviation. This process eliminates differences in spectral intensity and enhances comparability between samples. However, in situations where the data distribution is heavily skewed or contains extreme outliers, standardization may not effectively normalize the spectra. Furthermore, if the spectral data exhibit a high level of noise, standardization can amplify the noise during the scaling process, potentially compromising the accuracy of subsequent analysis [37]. Smoothing techniques are employed to reduce noise and eliminate high-frequency fluctuations in the spectrum. This method differentiates all spectral data and finds the slope of data points in each band to construct new data. Various smoothing algorithms such as moving average, Savitzky–Golay, or wavelet smoothing can be applied depending on the specific requirements of the data. However, excessive smoothing can result in the loss of fine details and important spectral information, particularly in regions of interest with sharp peaks or rapid changes. Selecting an

appropriate smoothing algorithm and adjusting the smoothing parameters are crucial to strike a balance between noise reduction and preservation of important spectral characteristics [38]. Trend correction plays a critical role in removing systematic variations or baseline drift from spectra. It involves fitting a mathematical function to the baseline and subtracting it from the original spectrum. However, it is important to note that trend correction methods may encounter challenges when the baseline drift is complex or exhibits nonlinear patterns. In such cases, accurately capturing the baseline variations and selecting an appropriate function for correction can be difficult. The effectiveness of trend correction also relies on the quality of the baseline estimation, which may be influenced by factors such as noise, overlapping peaks, or instrumental artifacts [39]. Differentiation is another widely used pretreatment method that aims to eliminate background interference by calculating the derivative of the spectrum. This process helps highlight specific features or spectral changes relevant for analysis. However, when overlapping peaks occur in the spectrum, differentiation attempts to differentiate between their individual contributions by calculating changes in intensity or slope at different wavelengths. This differentiation process involves calculating derivatives, which can amplify the noise present in the data. Consequently, the noise associated with the original spectrum may be magnified during the peak separation process, leading to inaccurate or distorted peak shapes. The introduction of noise during differentiation becomes particularly significant when the overlapping peaks have similar spectral profiles or are closely spaced, making it challenging to accurately differentiate between them. In such cases, the noise introduced by differentiation can hinder proper peak separation and may even create additional artifacts or false peaks in the resulting spectrum. To mitigate the impact of noise introduced by differentiation, careful consideration should be given to selecting the appropriate differentiation method, order, and window size [40]. Norris and SG are commonly used methods for differentiation. The Norris derivative method is also called direct difference method, which may cause errors for the sparse spectrum, so it is more suitable for the spectrum with more wavelength sampling points and higher resolution [41]. The SG derivative method solves the polynomial fitting matrix by least squares to obtain the derivative of the center point of the window. It overcomes the shortcoming of direct difference and is suitable for the sparse spectrum [42].

Multiple preprocessing methods can be applied simultaneously to enhance the quality of the spectral data. However, it is crucial to carefully determine the order in which these methods are implemented, considering their individual effects. In the present study, the researchers opted for a combination of two preprocessing methods, namely, multiplicative scatter correction (MSC) and second derivative (D2). The selection of the MSC and D2 combination method was based on their demonstrated efficacy in achieving the best experimental results, as discussed in the subsequent section.

MSC is employed to eliminate the baseline translation phenomenon caused by the scattering of NIR on the sample particles of uneven size. It performs well when the absorbance and the chemical properties of the sample show an obvious linear relationship [43]. By correcting this baseline translation, MSC helps improve the comparability and accuracy of the spectral data. To implement MSC, a standard spectrum is first required. In most cases, the mean of all spectra is used as the standard spectrum, as shown in the following equation:

$$\overline{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X_i}, \tag{1}$$

where $\overline{\mathbf{X}}_{1*p}$ is the standard spectrum, $(\mathbf{X_i})_{1*p}$ is every sample data, n is the number of spectral samples, and p is the number of spectral bands.

On the basis of the calculated standard spectrum, linear regression between each sample and the standard spectrum is carried out, as shown in equation (2), and the parameters $\mathbf{a_i}$ and $\mathbf{b_i}$ are obtained from the linear regression analysis.

$$\mathbf{X_i} = \mathbf{a_i}\overline{\mathbf{X}} + \mathbf{b_i}. \tag{2}$$

Finally, the values of each spectrum are corrected using the following equation:

$$\widehat{\mathbf{X}_i} = \frac{\mathbf{X_i} - \mathbf{b_i}}{\mathbf{a_i}}. \tag{3}$$

This equation represents the correction process, where $\widehat{\mathbf{X}_i}$ denotes the corrected values of the spectrum.

D2, or the second derivative, is a mathematical operation applied to the spectrum that calculates the rate of change of the spectral intensity. By taking the derivative twice, D2 highlights and amplifies changes in the spectral data. This process effectively removes unwanted background interference, such as baseline variations or fluctuations, that can obscure important spectral features. The calculation of the second derivative involves comparing the spectral intensity values of adjacent data points. The formula to calculate the second derivative, as shown in the following equation, is

$$x''_{i,k} = \frac{\mathrm{d}^2 x}{\mathrm{d}w^2} = \frac{x_{i,k+1} - 2x_{i,k} + x_{i,k-1}}{\left(w_{k+1} - w_k\right)^2}. \tag{4}$$

In this equation, $x''_{i,k}$ represent the second derivative results of the k-th band of the i-th sample, $x_{i,k}$ denotes the measurement result of the spectral intensity, and w_k represents the wavelength value of the k-th band. The second derivative operation enhances the visibility of spectral peaks, valleys, and other critical information that may be indicative of specific chemical components or characteristics. It helps to accentuate fine details and subtle variations in the spectrum, making them more distinguishable and easier to analyze.

2.2. Feature Band Screening Method.

The NIRS data obtained by the instrument contain a large amount of band information. On the one hand, the processing of a large number of features requires a robust computer performance to handle the computational load. On the other hand, a large

part of these feature bands are often redundant bands with a large amount of collinear information and useless information, which will only hinder the processing performance of the model [44]. As a result, the insufficient processing of the really important information leads to the instability of the model and poor experimental results. Therefore, before the formal establishment of the model, the bands should be screened in advance to simplify the model and improve its prediction ability. Feature screening is typically approached from two perspectives: interpretability of the labels and reduction of redundancy among independent variables. Common methods include statistical analysis, machine learning algorithms, and dimensionality reduction techniques [45]. Statistical analysis evaluates the correlation or importance between each band and the target variable, using measures such as correlation coefficients, analysis of variance, and information gain. Machine learning algorithms employ embedded methods, recursive feature elimination, and model-based feature selection. Dimensionality reduction techniques such as PCA and LDA transform the original bands into fewer new features while preserving the most informative data variations.

Determining the feature band selection method involves considering multiple factors such as the size of the dataset, feature correlations, computational resources, and model performance. Therefore, it is necessary to validate various feature extraction methods through experiments. In the next section, we provide a detailed description of the experimental process to evaluate the performance of different feature selection techniques. The results of the experiments clearly demonstrate the effectiveness of competitive adaptive reweighting sampling (CARS) and XGBoost in the feature screening process.

CARS, as a feature selection method, leverages Monte Carlo sampling and partial least squares regression to identify the most relevant and informative wavelengths for a given problem. It effectively addresses the issue of redundant and collinear information present in the feature bands. The feature selection process involves iterative statistics, where feature elimination is performed at each iteration. Specifically, a partial least squares (PLS) regression model is constructed using the r_i feature bands with higher weights in the modeling dataset. The mean square error of the test set model is then calculated for interactive verification. After completing N rounds of sampling, the variable subset with the minimum root mean square error cross-validation (RMSECV) is selected as the optimal feature set. CARS incorporates an adaptive reweighting mechanism based on the contribution of samples to the classification task. This allows CARS to assign higher weights to more informative features while reducing the influence of irrelevant or redundant ones. Unlike traditional feature selection methods, such as principal component analysis (PCA), which rely solely on statistical measures, CARS employs a competitive mechanism that actively promotes the selection of informative features while suppressing the influence of irrelevant or redundant ones. This adaptive weighting scheme enhances the performance and robustness of the selected feature subset.

On the other hand, XGBoost, a popular machine learning algorithm, demonstrates excellent performance in feature selection as well. It is an enhanced version of the gradient boost machine algorithm, which belongs to the ensemble learning category. By combining the strengths of gradient boosting and decision tree algorithms, XGBoost effectively identifies the most crucial features for accurate prediction. It utilizes an ensemble of weak learners to iteratively learn from the data and optimize a specific objective function. XGBoost consists of multiple CART trees, with each iteration adding a new tree to capture the residuals generated in previous iterations. To optimize leaf node splitting, XGBoost incorporates second-order Taylor expansion, maximizing the objective function gain at each split. The splitting process can be performed using a greedy algorithm or by selecting candidate points through an approximate algorithm prior to segmentation. As XGBoost progresses through the trees, it assigns diminishing weights to each tree, reducing their influence in subsequent expansion steps. This adaptive weighting scheme allows XGBoost to assign higher weights to informative features, enabling the algorithm to focus on the most discriminative aspects of the data [46]. Compared to other algorithms, XGBoost stands out as a powerful machine learning algorithm. It integrates regularization techniques to mitigate overfitting and enhance model generalization, guaranteeing that the chosen features are not only relevant but also robust across diverse datasets. Furthermore, XGBoost incorporates a built-in feature importance metric that evaluates and ranks the significance of each feature based on its contribution to the model's predictive performance. This valuable information facilitates feature ranking and selection, enabling researchers to concentrate on the most influential features during their analysis. In addition, XGBoost offers the advantage of having a smaller number of parameters compared to CARS, making it more manageable and easier to fine-tune.

2.3. Model Construction Method. Common NIRS regression models include multiple linear regression, principal component regression, partial least squares regression, and support vector regression [47]. However, these regression methods often yield unsatisfactory results when applied solely due to the large amount of NIRS data and collinearity issues. As a typical nonlinear modeling method, neural network models, such as multilayer perceptron (MLP), convolutional neural network (CNN), and long short-term memory (LSTM), are also commonly used in NIRS processing [48]. Multilayer perceptron is the simplest neural network model which includes input layer, output layer, and hidden layer. It updates the model through the gradient descent process, propagates the output of the fully connected layer in the forward direction, and updates the gradient in the back direction. The one-dimensional convolutional neural network (1DCNN) is another popular approach in NIRS analysis. Leveraging the powerful feature extraction capability of CNNs, 1DCNN extracts spectral data features and fits samples by exploring various functions. It typically comprises an input layer, convolutional layers, pooling layers, and fully connected layers. Another deep learning

model commonly used in NIRS processing is the self-coding model, which integrates an encoder and a decoder to interpret features. The encoder extracts features from the input data, while the decoder reconstructs them.

To better capture the complex features present in NIRS data, this paper employs an advanced approach for feature mapping. Specifically, a multilayer convolutional neural network (CNN) integrated with an attention mechanism, known as the efficient channel attention network (ECANet), is utilized to construct the model and map the data features.

CNNs are specifically designed to capture spatial and temporal dependencies in data, making them well-suited for tasks that involve extracting meaningful features from structured or sequential data [49]. In the context of NIRS data, CNNs are particularly valuable for their ability to automatically learn and extract relevant features at different levels of abstraction. The convolutional layers apply a set of filters to the input data, enabling the network to detect and extract local patterns and spatial relationships. The pooling layers, on the other hand, downsample the feature maps, reducing the spatial dimensions while preserving the most salient features. Finally, the fully connected layers combine the extracted features and provide the final output. The calculation process is described by the following equation:

$$f_i = \sum_{k=1}^{m} x_k \cdot \omega_k^i + b_k^i, \tag{5}$$

where f_i represents the output feature map of the i-th convolution layer. The index i corresponds to the specific filter used in that layer. x_k denotes the feature map of the input x to the i-th convolution layer. The weight of the filter for this convolution layer is represented by ω_k^i and b_k^i represents the bias term associated with that filter. The equation illustrates the summation of the element-wise product of the input feature map and the corresponding filter weights, followed by the addition of the bias term, to obtain the output feature map. By incorporating CNNs as a feature extraction module in our model, we aim to leverage their capability to enhance the representation and analysis of NIRS data, leading to improved performance and insights in our study. Moreover, the integration of attention mechanism into CNN can effectively improve its feature extraction performance.

ECANet is a highly efficient network that integrates a channel attention module. It builds upon the squeeze and excitation networks (SENets) and incorporates the attention mechanism [50]. While many deep learning networks focus on improving spatial dimensions, SENet introduces attention mechanisms from the channel dimension. It automatically learns the importance of different channel features, enhancing useful features while suppressing irrelevant ones. SENet consists of the SE module, which compresses each feature channel into a real number through spatial dimension compression. It learns a parameter that represents the correlation between channels and generates the importance weight for each feature channel. This weight is then used to recalibrate the input features of the SE module, completing the feature recalculation process. However, the initial compression of spatial dimensions in the SE module is

complex and computationally expensive. In addition, the two fully connected layers after the pooling layer can weaken the weight learning and prediction ability of channel attention. To address these limitations, the ECA module improves the compression operation by adopting a local cross-channel interaction strategy without dimensionality reduction. It enables cross-channel interaction of features, significantly reducing computational complexity [51]. Consequently, the ECA module provides an extremely lightweight channel attention mechanism.

The ECANet combines the power of CNNs with attention mechanisms to effectively capture and emphasize the most relevant information in the NIRS data. By leveraging the hierarchical feature extraction capability of CNNs and the attention mechanism's ability to focus on informative features, the ECANet enhances the model's ability to extract discriminative features from the NIRS data. By integrating the attention mechanism, the ECANet dynamically adjusts the weights of different channels in the feature maps, allowing the network to selectively emphasize important channels while suppressing less relevant ones. This adaptive weighting scheme further enhances the model's capability to capture and exploit the most informative features for accurate mapping of NIRS data.

The ECA module utilizes a band matrix, denoted as W_k in equation (6), to capture interactions between feature channels. This band matrix contains parameters that control the importance of each channel in the attention mechanism. Specifically, the weight y_i is computed by considering interactions with its k neighboring channels, as shown in equation (7). In vector form, it can be expressed as equation (8). This localized attention mechanism allows the network to adaptively adjust the weights of different channels based on their relevance to the task at hand.

$$W_k = \begin{bmatrix} \omega^{1,1} & \cdots & \omega^{1,k} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \omega^{2,2} & \cdots & \omega^{2,k+1} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \omega^{C,C-k+1} & \cdots & \omega^{C,C} \end{bmatrix}, \tag{6}$$

$$\omega_i = \sigma \left(\sum_{j=1}^{k} \omega^j y_i^j \right), \quad y_i^j \in \Omega_i^k, \tag{7}$$

$$\omega = \sigma \left(C1D_k(\mathbf{y}) \right). \tag{8}$$

To realize the information interaction in the ECA module, one-dimensional convolutions are employed instead of fully connected layers. This approach efficiently captures channel dependencies without the computational burden associated with fully connected operations. The size of the convolution kernel, denoted as k, is adaptively determined based on the number of feature channels C, as shown in the following equation:

$$k = \Phi(C) = \left| \frac{\log_2(C)}{\Upsilon} + \frac{b}{\Upsilon} \right|_{odd}, \tag{9}$$

where Φ indicates nonlinear mapping, Υ and b are the parameters, and odd indicates taking the nearest odd number. This adaptive kernel size selection ensures that the ECA module effectively captures the relevant information in the feature maps. The integration of the ECANet for feature mapping in this paper aims to improve the representation and modeling of the complex features inherent in NIRS data. By dynamically adjusting the channel weights and capturing relevant dependencies, the ECANet enhances the overall performance and predictive accuracy of the model. Figure 1 illustrates the implementation process of the ECA module:

Typically, the cross entropy function is commonly chosen as the loss function in deep learning models. However, optimizing the loss function solely based on large errors with the real label can lead to overfitting, especially when working with small sample sizes. In addition, manually assigned labels may contain errors, which can significantly impact the network, particularly when the number of samples is limited. To address these concerns, label smooth regularization was introduced in the experiment to optimize the label assignment in the support set. Label smooth regularization employs weighted mixing in the calculation of the cross entropy loss function to reduce the weight of real sample labels, thereby inhibiting model overfitting and improving accuracy. It replaces the original true label distribution, represented as $q(k\,|\,x) = \delta_{(k,y)}$, with a modified label distribution as shown in the following equation:

$$q'(k\,|\,x) = (1-\epsilon)\delta_{(k,y)} + \epsilon f(k), \qquad (10)$$

where k is the real label, x is the training sample, $q(k\,|\,x)$ is the original real label distribution, q' is the replaced label distribution, $\delta_{(k,y)}$ is the class of one-hot coding vector, ϵ is the label smoothing parameter, and $f(k)$ is a prior label distribution independent of x. The formula can be regarded as the fusion of the original distribution q and the definite distribution f, whose probabilities are $(1-\epsilon)$ and ϵ, respectively. In this paper, f is a uniformly distributed function, so the label distribution after smoothing can be obtained as equation (11), where $K = 3$ is the total number of categories.

$$q'(k\,|\,x) = (1-\epsilon)\delta_{(k,y)} + \frac{\epsilon}{K}. \qquad (11)$$

In the deep learning model, the independent individual distributed data will achieve the best training effect, and the resulting model has strong generalisation and high prediction ability. However, with the increasing number of layers in the model network, the slight changes in the lower layer network may cause the input data distribution of the upper layer network to become offset. As a result, the upper layer network becomes saturated and the lower layer network gradient disappears when the back propagation occurs. The batch normalisation (BN) and layer normalisation (LN) can force data back to the standard distribution, avoiding problems such as saturation of activation functions and making the model unsensitive to initial parameters and network depth, thereby stabilising the training process. At the same time, the general deep learning model requires regularisation to stabilise the model, such as randomly ignoring neurons using the dropout optimiser, so the model can simulate a large number of network structures and improve the robustness of neuronal nodes inside the model. BN and LN also have a certain regularisation effect on the data as they make the loss function smoother, allowing to learn in larger steps and reducing training time and training costs. However, although BN and LN have the same purpose, they are used in different situations and ways. BN can normalise all samples of a single neuron for the same dimension of all data in a batch, so it is sensitive to the batch size. When the number of batches is small, the gradient will be unstable and the effect will be worse. LN is for all neurons in a layer of network, that is, all dimensions of each sample are normalised to reduce the variance of the model. As the batch size in this experiment is small, LN should be used instead of BN. After LN processing, the channel direction will be normalised, and batch size data will be output in a batch. In this method, the distribution of each layer can be stabilised, and the subsequent layer can continue to learn, stay away from the derivative saturation area, and accelerate the model convergence.

The network structure after incorporating the ECA module is illustrated in Figure 2(a), while the network structure with the built-in attention mechanism is depicted in Figure 2(b). Finally, the network is flattened, and the softmax function is used as the classification function. The model still employs the cross-entropy function as the loss function and utilizes the Adam optimizer for gradient descent during the training process.

3. Experimental

3.1. Data Acquisition.
The experimental materials used in this study comprised 238 tobacco blending modules provided by Hubei China Tobacco Industry Co., Ltd. These modules spanned the years 2017–2019. The dataset consisted of 76 samples from the clear flavor module, 104 samples from the intermediate flavor module, and 58 samples from the luzhou flavor module. The distribution of the modules across the production years is detailed in Table 1:

The instruments used for data collection are Brook MATRIX-I Fourier transform near-infrared spectrometer, Binder BD400 standard incubator, and AUARI shredder. After the sampling process, the 238 cigarette blending modules were placed in a standard incubator and dried at 40°C for a duration of 2 hours. Subsequently, the samples were ground into a 40-mesh (0.425 nm) powder using a grinder and sealed for further testing. For each module sample, approximately 50 g of the powdered material was placed in a sample cup and compacted before being subjected to NIR spectrometer data sampling. Throughout the experiment, strict environmental conditions were maintained, with a temperature of 22°C and a relative humidity of 60%. The NIR spectrometer utilized in this study collected spectral data within the band range of 3600–12500 cm^{-1} using the diffuse reflection method. The spectral resolution was set at 16 cm^{-1}. To ensure accuracy, five spectra were repeatedly collected for each sample, and the average value was taken as the representative data for analysis. The NIRS raw data obtained from this process is illustrated in Figure 3; a total of 238 modules of NIRS data were collected, each containing 1154 band values.
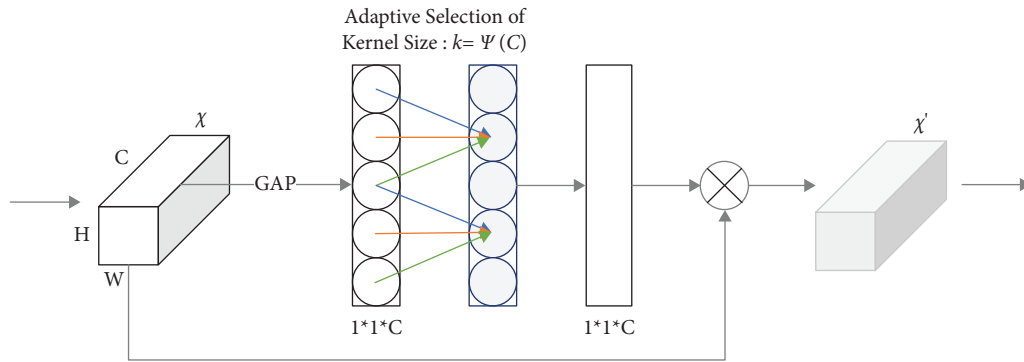
Adaptive Selection of
Kernel Size : $k = \Psi(C)$

FIGURE 1: ECA module implementation process.

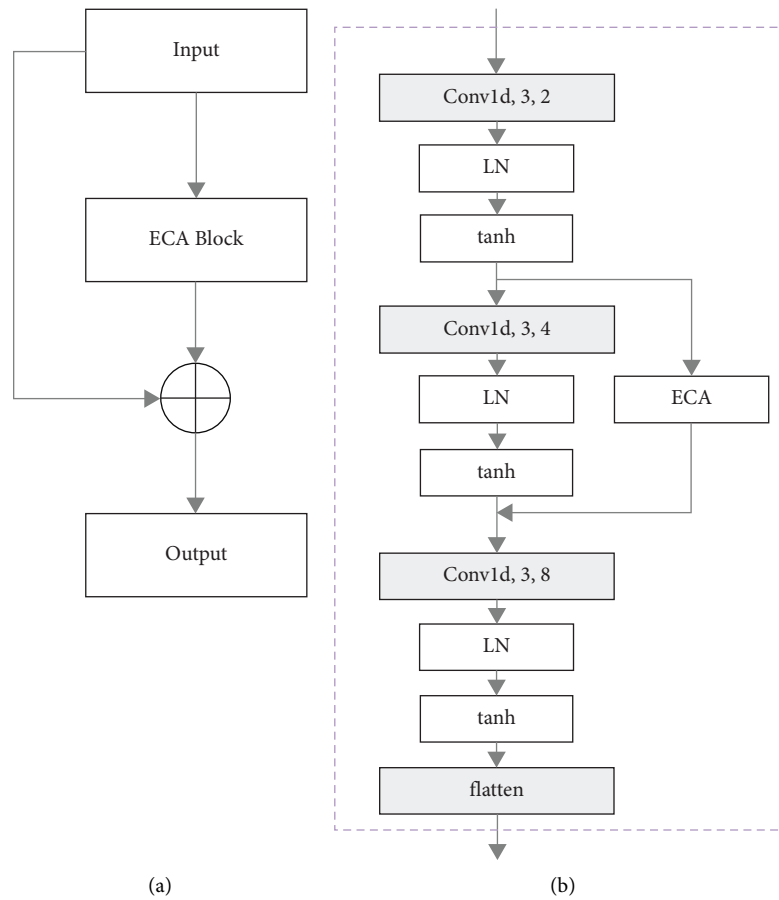(a)   (b)

FIGURE 2: Model structure. (a) ECA block structure. (b) Network structure.

3.2. Data Preprocessing. After careful evaluation, the combination of MSC + $D2$ was finally determined as the optimal pretreatment method for the NIRS data. In order to assess the impact of different pretreatment methods on the subsequent analysis, a variety of pretreatment techniques were initially considered. Figure 4 shows the data images obtained after applying MSC, first derivative ($D1$), and second derivative ($D2$) processing, respectively. By comparing the visual representations, it becomes evident that the MSC technique plays a crucial role in normalizing the spectra and enhancing their consistency. Through the application of MSC, the fluctuation range of

all spectral data is effectively reduced, leading to a more stable and uniform representation. In addition, MSC enhances the spectral overlap, ensuring a better alignment of spectral features and improving the overall quality of the data. On the other hand, the derivative methods exhibit noticeable effects on the variation range and structural characteristics of the spectral data. These methods introduce substantial changes in the shape and intensity of the spectral features. $D1$ amplifies the rate of change in the spectrum, resulting in sharper peaks and valleys, while $D2$ accentuates these changes even further, leading to more pronounced variations in the spectral shape.

TABLE 1: Experimental data statistics.

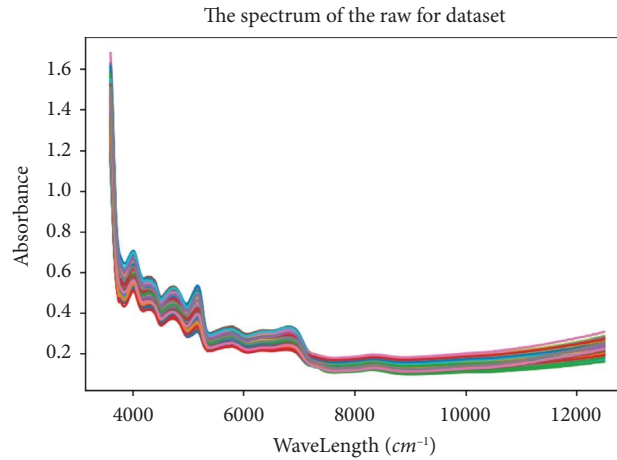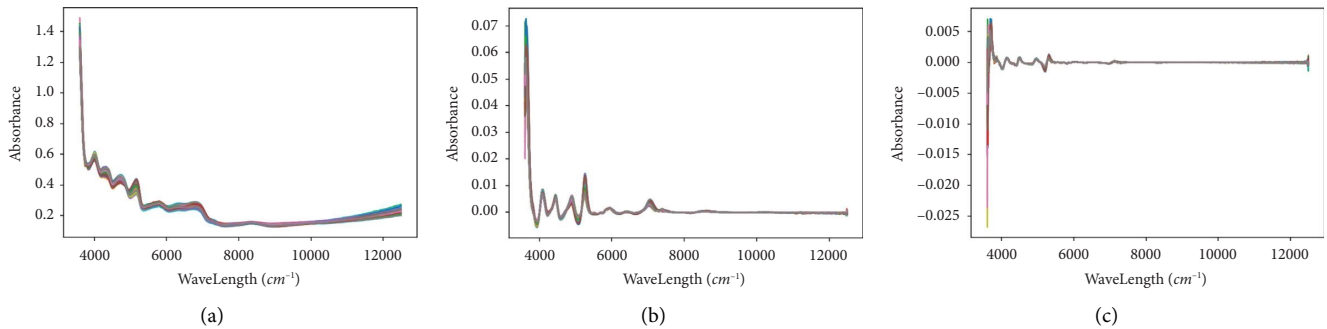| Flavor	Year		
	2017	2018	2019
Clear flavor	25	11	40
Intermediate flavor	42	10	52
Luzhou flavor	22	7	29

FIGURE 3: Raw NIRS data.

(a)   (b)   (c)

FIGURE 4: Data distribution after single preprocessing. (a) MSC. (b) $D1$. (c) $D2$.

The XGBoost model was employed in flavor evaluation based on spectral data to assess the effect of the pretreatment method. This model offers several advantages, including a lower number of parameters, a simple modeling process, and the ability to effectively validate the feature extraction effect. Data were randomly selected as the training set and test set of the experiment, and the ratio was 6 : 4, that is, 143 pieces were selected as the training data and 95 pieces as the test data. The model was constructed based on the training set, and predictions were made using the test set. This process was repeated 10 times to ensure robustness, and the final results were averaged to obtain more reliable outcomes. In order to evaluate the effectiveness of different preprocessing methods, this study also considered several alternative approaches. The test set accuracy results obtained from these different methods were recorded and are presented in Table 2.

The accuracy of several preprocessing methods was found to be lower compared to direct processing with the source data. This suggests that these methods alter the internal

TABLE 2: Accuracy of the single pretreatment model.

Methods	Accuracy (%)
Raw data	43.73
$D2$	62.50
Standard normal variate (SNV)	44.58
Standardization (SS)	41.67
Moving average (MA)	41.33
Wavelet transform (WT)	40.54
MSC	46.88
$D1$	55.23
Min-max normalization (MMN)	43.09
Mean centralization (MC)	46.34
SG smoothing (SG)	41.23
Detrend (DT)	46.98

information structure of the potential reaction module in the source data, leading to a decrease in model accuracy. The accuracy of $D2$ is the highest, indicating that the $D2$ processing mode can further clarify the internal structure of data. In

general, the effectiveness of a single preprocessing method is limited due to the presence of various interference factors in complex samples. By combining different preprocessing methods, it becomes possible to eliminate different types of interference information. This comprehensive approach allows for better extraction of relevant features and ultimately improves the accuracy of the model.

Several common pretreatment methods, as well as several single treatment methods with good performance, were selected for combination based on their specific purposes. The derivative and DT methods were employed to eliminate baseline drift, while SNV and MSC were used to reduce scattering effects. In addition, MC was applied to eliminate background noise. The aforementioned data set partitioning methods and experimental approaches were also implemented. The accuracy of the test results from two pretreatment combination models is presented in Table 3.

The combination method has a better performance than the single pretreatment method on the whole. Particularly, the combination of other methods with $D2$ demonstrates the best performance. The highest accuracy of the data results is achieved by using the combination of MC + $D2$, SNV + $D2$, and MSC + $D2$. These preprocessing methods yield similar trends and distributions in the resulting images. Figure 5 illustrates the data after MSC + $D2$ processing, which exhibits the highest accuracy.

Furthermore, three preprocessing methods were combined, with the second derivative ($D2$) applied as the final step. The accuracy test results are presented in Table 4. It can be observed that combining the three preprocessing methods increases the complexity of the preprocessing while substantially altering the original data structure. However, this combination does not lead to a significant improvement in prediction accuracy and, in some cases, even results in a slight decrease compared to the combination of two preprocessing methods. Therefore, the combination of MSC + $D2$ for data preprocessing is considered as the preferred choice.

3.3. Feature Band Screening.

While using full-spectrum modeling to increase the complexity of the model, the effect may be reduced due to the large number of interference variables. The selection of band information can effectively extract useful information and improve the model prediction effect. Therefore, before the formal establishment of the model, feature screening should be conducted. The XGBoost model was also used for flavor prediction to measure the merits of the band screening method.

In this paper, several models including principal component analysis (PCA), successive projections algorithm (SPA), competitive adaptive reweighting sampling (CARS), and XGBoost ensemble learning were compared for feature band extraction, and their parameters were adjusted, respectively. The optimal experimental results obtained are presented in Table 5.

The accuracy of PCA and SPA feature band extraction is not as high as that of all bands modeling, and the performance of PCA is extremely poor. This could be attributed to the fact that these methods solely focus on the relationship

TABLE 3: Accuracy of two pretreatment combination models.

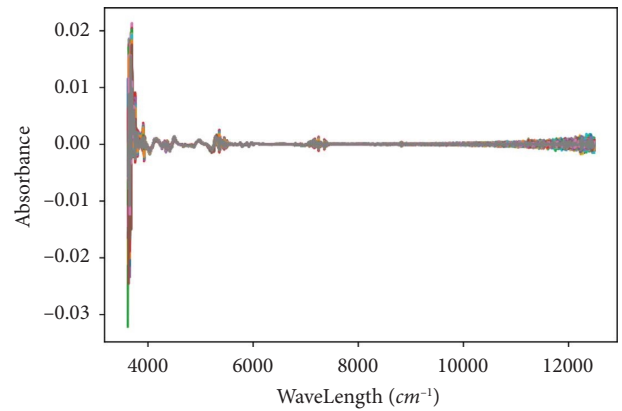| Methods	Accuracy (%)
DT + SNV	50.03
$D1$ + MSC	58.38
$D1$ + SNV	56.45
$D2$ + MSC	56.70
MC + $D1$	57.05
MSC + MC	47.44
MSC + $D1$	59.88
SNV + $D1$	57.42
MSC + $D2$	65.79
SNV + $D2$	62.31
MC + $D2$	64.14
DT + MC	46.38

FIGURE 5: Data distribution after the preprocessing of MSC + $D2$.

TABLE 4: Accuracy of three pretreatment combination models.

Methods	Accuracy (%)
MC + SNV + $D2$	62.50
SNV + MC + $D2$	61.48
MSC + MC + $D2$	63.08

TABLE 5: Optimal experimental results of different feature band extraction model.

Methods	Characteristic number	Accuracy (%)
PCA	8	34.29
SPA	9	50.00
CARS	145	64.39
XGBoost	150	66.67

between independent variable bands, neglecting the crucial relationship between the variable bands and the predicted value. As a result, the extracted bands may not adequately capture the necessary information for accurate modeling, leading to suboptimal performance. Furthermore, the limited number of bands extracted after sampling is insufficient to effectively explain the underlying model, further hampering the modeling process. In addition, PCA extracts bands in a linear manner and combines some band information, which may result in a less effective outcome, as the nonlinear relationships and intricate interactions among the bands are not fully considered in the extraction process.

However, the CARS and XGBoost classification methods demonstrate superior performance. These methods extract a sufficient number of bands while preserving the original data structure. By retaining the essential features of the original model and reducing complexity, these methods effectively filter out interfering factors. Consequently, the feature extraction methods based on CARS or XGBoost were considered. In this approach, the number of extracted bands is approximately one-tenth of the original bands, significantly reducing model complexity and enhancing processing capabilities. To summarize, this paper employs an XGBoost-based feature extraction model, which not only achieves slightly higher accuracy than CARS but also offers simpler tuning with reduced dependence on initial parameters.

The XGBoost model utilizes all bands to construct the model and assigns each band an importance score that reflects its contribution to the overall model. In this paper, the XGBoost algorithm adopts $g$btree as the weak learner and softmax as the classification function. The optimal parameter set is determined through grid search. Following the experiment, the maximum depth of the tree is set to 8, the minimum sample weight sum of the child nodes is 2, the number of established trees is 70, the learning rate is 0.06, the regularization weight is 2, and the shrinkage step is set to 0.05. In addition, cycle optimization is performed using five-fold cross-validation.

The importance scores of each band in the final model are depicted in Figure 6(a). In this study, 150 bands were chosen as the representative bands for the model. These characteristic bands accounted for 80% of the model's interpretability while comprising only 13% of the total number of original bands. This significant reduction in the number of bands greatly reduced the model's complexity. Moreover, the model also achieved high prediction accuracy. The distribution of the selected characteristic bands within the original bands is illustrated in Figure 6(b).

3.4. Model Construction. In this experiment, the near-infrared spectral data of 238 modules were randomly divided into 190 pieces of data in the training set and 48 pieces of data in the test set according to the ratio of 8 : 2. The specific data distribution is shown in Table 6.

Model details are shown in Table 7. The size parameter represents the dimensions of the convolution kernel. The number parameter refers to the number of channels in the convolutional layer, indicating the dimensionality of the output feature maps. The stride parameter determines the step size at which the convolution kernel moves across the input data. The second layer of the network consists of two components: an ordinary convolutional layer and an ECA module. The ordinary convolutional layer processes the input data from the previous layer using convolutional operations. On the other hand, the ECA module operates on the output results of the first layer, incorporating channel-wise attention mechanisms. The ECA module enhances the interdependencies between different channels, allowing the network to capture more relevant and discriminative features. The outputs from both the ordinary convolutional layer and the ECA module are then combined and passed on to the subsequent third layer for further processing.

In the training process, due to the limitation of the available dataset, the batch size for each training iteration was set to 1. This choice was made to maximize the utilization of the available data and facilitate efficient learning. We have planned a total of 120 training rounds, with each round consisting of 30 iterations. The initial learning rate of the model was set to 0.005. As the training progresses, a cosine decay schedule was employed to gradually reduce the learning rate. This decay schedule helped ensure that the model converged effectively and mitigated the risk of overfitting by fine-tuning the learning rate throughout the training process. Moreover, the label smoothness rate to 0.1. During the training process, we continuously monitor the convergence of the model. If the model demonstrates satisfactory convergence and achieves the desired performance before completing the planned training rounds, we have the flexibility to stop the training process prematurely. This allowed us to save computational resources while ensuring that the model reached its optimal performance.

The selection of activation functions plays a crucial role in enabling the model to exhibit a nonlinear structure and enhance its feature extraction capabilities. A well-chosen activation function can facilitate faster convergence and improve overall model performance. In this experiment, we explored several common activation functions, including ReLU, ELU, and tanh, to assess their effectiveness and conduct comparative experiments. The comparison of the first two activation functions is shown in Figure 7. It was observed that the ReLU activation function had limited success, which can be attributed to its characteristic of setting negative gradients to zero. This behavior can lead to the "dying ReLU" problem, where some neurons become inactive, resulting in the corresponding parameters being unable to update. Although the ELU activation function addresses this issue to some extent, its convergence efficiency was found to be inferior to that of tanh. The tanh activation function, on the other hand, demonstrated faster convergence and better handling of gradient updates near zero. Therefore, it was selected as the activation function for the middle layer of the model. In the output layer, the softmax activation function was chosen because it compresses the output values into the range of [0, 1]. This property is particularly convenient for direct classification predictions. In summary, based on the experimental results and the desirable characteristics of the activation functions, we selected tanh as the activation function for the middle layer and softmax as the activation function for the output layer. These choices aim to promote faster convergence, improve gradient updates, and facilitate more effective classification predictions.

4. Results and Discussion

The model training process is illustrated in Figure 8, where it can be observed that the model started to converge after approximately 100 training iterations and eventually reached a stable state.
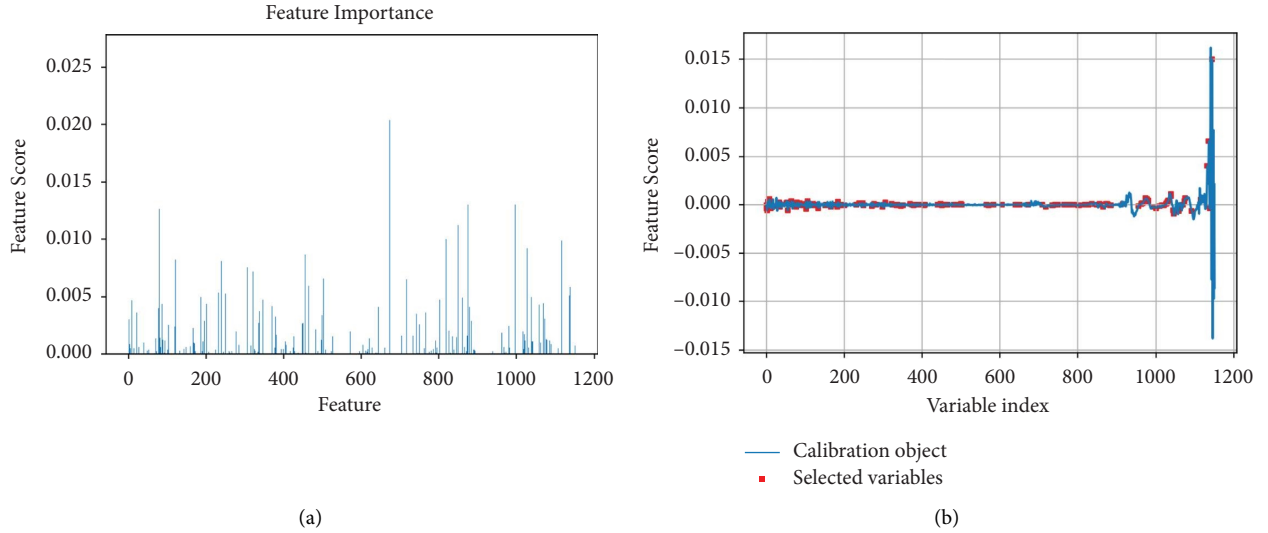
(a)

Calibration object
Selected variables

(b)

FIGURE 6: XGBoost model result. (a) Score of each band. (b) Selected characteristic distribution bands.

TABLE 6: Confusion matrix of the result.

Data set	Flavor type	Quantity
	Clear flavor	61
Training set	Intermediate flavor	83
	Luzhou flavor	46
	Clear flavor	15
Testing set	Intermediate flavor	21
	Luzhou flavor	12

TABLE 7: Detailed structure of the model.

Layer	Detailed structure	Size	Number	Stride	Output size
First layer	Conv1 + LN + tanh	1∗3	4	2	1∗74∗4
Second layer	Conv2a + LN + tanh	1∗3	8	1	1∗72∗8
	GlobalMaxPooling + reshape	3	8	—	1∗8∗1
ECA block	Conv2b + reshape	1∗3	8	1	1∗1∗8
	Mutiply	—	—	—	1∗72∗8
Third layer	Conv3 + LN + tanh	1∗3	16	1	1∗70∗16
Output layer	MaxPooling	3	16	—	1∗23∗16
	Flatten + softmax	—	—	—	1∗3

The accuracy achieved on the test set was 91.67%, while the training set accuracy reached a perfect score of 100%. The confusion matrix of the test set is presented in Table 8, demonstrating the overall good performance of the model. Particularly, the intermediate flavor module exhibited excellent predictive capability, achieving a 100% accuracy rate. However, the clear flavor module showed relatively poorer performance, which could be attributed to the imbalanced distribution of samples across different flavor categories in the dataset. To further enhance the prediction effectiveness, it may be necessary to address this imbalance by supplementing the dataset with additional samples and ensuring a balanced representation of each flavor category.

The selected characteristic variables were used as input data and the scent type was used as a classification label to further verify the superiority of the model. The BP neural network (BP), partial least squares regression (PLS), and random forest (RF) were used to learn and predict the model of the training set and test set data. The experimental results are shown in Table 9.

In some common NIRS classification models, the effect of flavor classification is poor, which may be due to their low complexity and inability to effectively extract the feature information when processing complex data, resulting in an unsatisfactory modeling effect.

Ablation experiments were conducted to verify the validity of the model established. Compared with the original model, layer normalisation was removed in model 1, and the data directly enter the activation layer after passing through the convolution layer. Model 2 took the form of
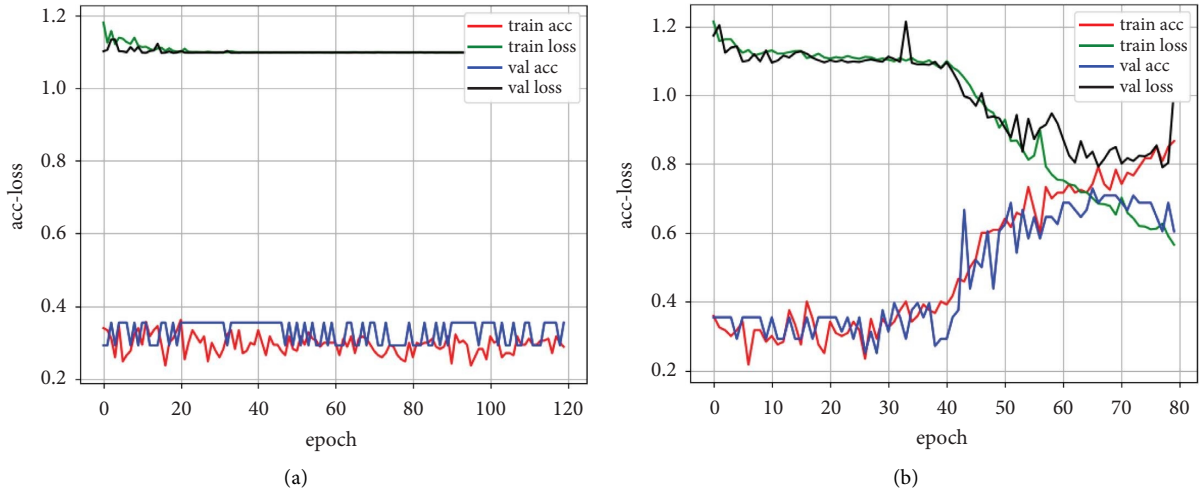
(a)   (b)

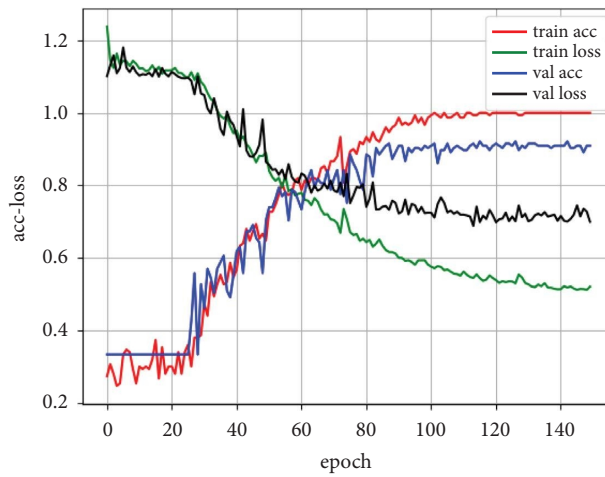FIGURE 7: Comparison of activation functions. (a) ReLU. (b) ELU.

FIGURE 8: Model training process.

TABLE 8: Results of other models.

Real	Predict			
	Clear flavor	Intermediate flavor	Luzhou flavor	Total
Clear flavor	13	1	1	15
Intermediate flavor	0	21	0	21
Luzhou flavor	0	1	11	12
Total	13	23	12	48

TABLE 9: Results comparison.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
BP	77.08	77.34	77.08	77.21
PLS	54.17	53.90	52.30	53.09
RF	66.67	67.76	66.67	67.21
This paper	91.67	94.32	92.78	93.54

hard labels and used the original labels directly when calculating the cross entropy loss function. In model 3, the ECA block was removed and the convolution neural network without attention mechanism was adopted. Model 4 did not carry out feature band extraction and adapted full-spectrum data modeling. The experiment was repeated for 10 times, and the average accuracy is shown in Table 10. Based on the experimental results, all these factors play a crucial role in

TABLE 10: Results of ablation experiment.

Model's serial number	1	2	3	4	Original model
Accuracy (%)	73.54	82.29	80.63	62.92	91.46

TABLE 11: Comparison results of other models.

Model	Sensory evaluation	LDA	PLS	CNN	Our model
Accuracy (%)	81.41	94.44	86.11	84.57	94.54

the construction of a good model, and the prediction accuracy of all the flavor types decreased to a certain extent when these elements were removed. At the same time, the convergence rate of all models decreased, and the convergence amplitude of model 1 and model 2 decreased significantly and was unstable. The loss function value of model 1 decreased very little, and the prediction accuracy was low. The prediction accuracy of model 4 was also low, which may due to the large number of interference bands, and the model paid too much attention to irrelevant details of the data.

Furthermore, to provide a more robust evaluation of the model's effectiveness, the leave-one cross-validation method was employed in this experiment. This technique involves iteratively training the model on all but one sample and then testing its performance on the remaining sample. This process is repeated for each sample in the dataset, ensuring that every sample serves as both training and testing instances. After conducting the leave-one cross-validation, the average accuracy rate was calculated, yielding a result of 94.54%. This average accuracy rate provides a more reliable estimate of the model's predictive performance, as it takes into account the variability that may arise from different training and testing data subsets. The high average accuracy rate obtained through leave-one cross-validation further demonstrates the robustness and generalizability of the model. It indicates that the model has learned meaningful patterns and features from the training data that enable it to accurately predict the flavor of spectral data even when presented with unseen samples. It also demonstrates that data augmentation approach would help the model generalize better.

Finally, we compared the experimental results of our paper with the model accuracy results reported in the latest articles on tobacco modular flavor prediction [52, 53]. The comparison results are presented in Table 11. It is worth noting that the availability of limited data poses a challenge, but despite this limitation, our prediction model has achieved a relatively favorable outcome. The accuracy of flavor prediction in our model has shown a slight improvement compared to some of the existing models.

Our research contributes to the field of tobacco flavor prediction by presenting a promising model with improved accuracy. The next research will consider the case of obtaining more data to further improve the feature understanding ability of the model.

5. Conclusion

This study proposes a classification model based on XGBoost integrated learning and deep learning for the rapid positioning and scientific evaluation of tobacco blending module flavor styles. Characteristic variables with strong correlation with flavor types in the near-infrared spectrum data of the module were used to recognise the flavor types.

First, the combination of multiplicative scatter correction and second derivative was used to preprocess the data to eliminate noise and baseline drift. The XGBoost model was used to extract 150 relevant bands. In feature modeling, the ECA module with attention mechanism and layer normalization was introduced into the feature coding mapper. The smooth label coding was used to replace the original label one-hot coding to calculate the cross entropy loss function and optimise it.

The experiment showed that the convolutional neural network with attention mechanisms combined with the feature information extracted from the high-dimensional near-infrared spectrum by XGBoost could effectively identify the flavor style features of the tobacco blending module and realise the objectification of the flavor index evaluation. The flavor category recognition accuracy of our proposed model reached an impressive 95.54% in the leave-one cross-validation, highlighting its robustness and effectiveness. This achievement demonstrates the model's strong learning and prediction abilities, positioning it as a promising method for objective sensory quality evaluation of the tobacco rolling group's formulation module. The model's accurate predictions provide a scientific foundation for decision-making by professionals in the tobacco industry, enhancing their ability to make informed choices.

Data Availability

The data used to support the findings of the study are available at https://github.com/gyhhhhh/NIRS.

Disclosure

The authors declare that the funder has no impact on the results or outcomes of the study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] G. Wang, W. Guo, Y. Li, M. Yang, X. Cai, and J. Song, "Regional characteristics of flue-cured tobacco in different flavor areas and their relationship with sensory quality," *Chinese Tobacco Science*, vol. 43, pp. 64–70, 2022.

[2] X. Qiao, Y. Shen, Y. Ma, B. Wang, and C. Zhang, "Study on characteristic aroma notes of flue-cured tobacco leaves of different flavor styles," *Tobacco Science & Technology*, vol. 2, pp. 5–7, 2014.

[3] A. Liu, K. Yuan, Q. Li et al., "Metabolomics and proteomics revealed the synthesis difference of aroma precursors in tobacco leaves at various growth stages," *Plant Physiology and Biochemistry*, vol. 192, pp. 308–319, 2022.

[4] Z. Li, N. Wang, D. Wang, X. Zhu, and H. Zhang, "Preliminary study of aroma type styles of flue-cured tobacco in different ecological scale regions," *Chinese Tobacco Science*, vol. 30, no. 5, 2019.

[5] X. Guan, Y. Lou, Y. Wen, X. Zhang, X. Yue, and B. Xu, "Eight flue-cured tobacco producing area in central typical tobacco scent is suitable for drying temperature is obtained research," *Journal of China tobacco*, vol. 29, pp. 11-12, 2023.

[6] W. Liu, B. Liu, R. Ma, B. Xv, and B. Wang, "Formulation and Process Technology based on cigarette brand style characteristics," *Food Industry*, vol. 42, pp. 96–100, 2021.

[7] G. Jia, X. Wang, Y. Sun, C. Nie, J. He, and Y. Feng, "Analysis on the relationship between free amino acid content and tobacco quality of flue-cured tobacco in Eight major flavor ecological regions," *Crops*, vol. 10, pp. 1–6, 2023.

[8] Z. Chen, L. Cai, X. Chen, Y. Zhang, S. Ma, and G. Xi, "Analysis on the difference of flavor characteristics of intermediate flavor flue-cured tobacco from different regions based on flavor activity value," *Journal of Light Industry*, vol. 37, pp. 79–86, 2022.

[9] T. Liu, S. Guo, C. Wu et al., "Phyllosphere microbial community of cigar tobacco and its corresponding metabolites," *Frontiers in Microbiology*, vol. 13, Article ID 1025881, 2022.

[10] F. Yin, E. Karangwa, S. Song et al., "Contribution of tobacco composition compounds to characteristic aroma of Chinese faint-scent cigarettes through chromatography analysis and partial least squares regression," *Journal of Chromatography B*, vol. 1105, pp. 217–227, 2019.

[11] T. Zheng, Q. Zhang, P. Li et al., "Analysis of microbial community, volatile flavor compounds, and flavor of cigar tobacco leaves from different regions," *Frontiers in Microbiology*, vol. 13, Article ID 907270, 2022.

[12] W. Zuo, X. Cai, H. Xu, W. Guo, H. Wang, and C. Xue, "Nanling hill biome-coke sweet alcohol sweet scent producing flue-cured tobacco appearance characteristics study," *Journal of China tobacco*, vol. 28, pp. 92–100, 2022.

[13] J. Chen, X. Cai, W. Guo, H. Wang, C. Xue, and W. Zuo, "Analysis on Appearance Characteristics of flue-cured Tobacco and its relationship with conventional chemical constituents and sensory quality in some sweet-flavor producing areas in Southwest China," *Tobacco science& technology*, vol. 53, pp. 7–14, 2020.

[14] T. Yan, P. Zhou, F. Long, J. Liu, F. Wu, and M. Zhang, "Unraveling the difference in the composition/content of the aroma compounds in different tobacco leaves: for better use," *Journal of Chemistry*, vol. 2022, Article ID 3293899, 10 pages, 2022.

[15] Y. Zhou and G. Dao, "Spectral characteristics of alteration minerals and its application in gold deposit prospecting in Wulonggou, Qinghai province," *Journal of Geomechanics*, vol. 18, no. 3, 2022.

[16] L. Serva, G. Marchesini, M. Cullere, R. Ricci, and A. Dalle Zotte, "Testing two NIRs instruments to predict chicken breast meat quality and exploiting machine learning approaches to discriminate among genotypes and presence of myopathies," *Food Control*, vol. 144, Article ID 109391, 2023.

[17] L. Zhang, X. Ding, and R. Hou, "Classification modeling method for near-infrared spectroscopy of tobacco based on multimodal convolution neural networks," *Journal of Analytical Methods in Chemistry*, vol. 2020, Article ID 9652470, 13 pages, 2020.

[18] Y. Chen, J. Bin, C. Zou, and M. Ding, "Discrimination of fresh tobacco leaves with different maturity levels by near-infrared (NIR) spectroscopy and deep learning," *Journal of Analytical Methods in Chemistry*, vol. 2021, Article ID 9912589, 11 pages, 2021.

[19] K. Wei, J. Bin, F. Wang, and C. Kang, "On-line monitoring of the tobacco leaf composition during flue-curing by near-infrared spectroscopy and deep transfer learning," *Analytical Letters*, vol. 55, no. 13, pp. 2089–2107, 2022.

[20] A. Borges-Miranda, F. J. Silva-Mata, I. Talavera-Bustamante, J. Jiménez-Chacón, M. Álvarez-Prieto, and C. S. Pérez-Martínez, "The role of chemosensory relationships to improve raw materials' selection for Premium cigar manufacture," *Chemical Papers*, vol. 75, no. 8, pp. 4075–4091, 2021.

[21] D. Jiang, G. Hu, G. Qi, and N. Mazur, "A fully convolutional neural network-based regression approach for effective chemical composition analysis using near-infrared spectroscopy in cloud," *Journal of Artificial Intelligence and Technology*, vol. 1, pp. 74–82, 2021.

[22] D. Jiang, G. Qi, G. Hu, N. Mazur, Z. Zhu, and D. Wang, "A residual neural network based method for the classification of tobacco cultivation regions using near-infrared spectroscopy sensors," *Infrared Physics & Technology*, vol. 111, Article ID 103494, 2020.

[23] Z. Zhu, G. Qi, Y. Lei et al., "A Long short-term memory neural network based simultaneous quantitative analysis of multiple tobacco chemical components by near-infrared hyperspectroscopy images," *Chemosensors*, vol. 10, no. 5, p. 164, 2022.

[24] M. Rodrigues, R. Berti de Oliveira, G. Leboso Alemparte Abrantes dos Santos et al., "Rapid quantification of alkaloids, sugar and yield of tobacco (Nicotiana tabacum L.) varieties by using Vis–NIR–SWIR spectroradiometry," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 274, no. 5, Article ID 121082, 2022.

[25] J. Li, Z. Wang, L. Chen, and H. Chen, "Application of near-infrared spectroscopy in cigarette blend analysis: a comprehensive review," *Trends in Analytical Chemistry*, vol. 122, Article ID 115718, 2020.

[26] L. Zhang, X. Zhang, Y. Zhang, and Q. Hu, "Recent advances in near-infrared spectroscopy for the analysis of tobacco and tobacco products," *TrAC, Trends in Analytical Chemistry*, vol. 143, Article ID 116409, 2021.

[27] J. Wang, M. Zhang, X. Zhang, and X. Wang, "Near-infrared spectroscopy-based analysis of tobacco and its application in cigarette blend formulation," *Frontiers in Chemistry*, vol. 7, no. 135, 2019.

[28] M. Liu, X. Li, C. Wang, and Y. Shi, "Application of near-infrared spectroscopy in tobacco quality evaluation: a review," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 245, Article ID 118939, 2021.

[29] Y. Wang, C. Qiu, J. Yang, S. Li, F. Wang, and Y. Hua, "Evaluation on mixing uniformity of reroasted tobacco by near infrared spectroscopy," *Ournal of Anhui Agricultural Sciences*, vol. 50, pp. 178–180, 2022.

[30] C. Wang, Y. Liu, K. Yang, W. Xv, and Z. Tang, "Relationship between grade quality and key chemical indexes of flue-cured tobacco by near infrared rapid detection," *Journal of Measurement and Testing in China*, vol. 47, no. 2, 2021.

[31] Y. Yang, Q. Peng, X. Zhang, S. Liu, and Y. Wu, "Study on the relationship between sensory quality and chemical composition of cigarette raw materials," *Journal of Food Safety and Quality Inspection*, vol. 9, no. 20, 2018.

[32] S. Liu, X. Gan, G. Wang, Y. Li, and K. Hui, "Prediction model of sensory quality of alcoholized tobacco based on main chemical components," *Southwest China Journal of Agricultural Sciences*, vol. 33, no. 7, 2020.

[33] Y. Bi, K. Yuan, W. Xiao et al., "A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation," *Analytical Chemistry Journal*, vol. 909, pp. 30–40, 2016.

[34] T. Yuan, J. Wang, W. Ji, Y. Duan, and W. Li, "Application of near infrared spectroscopy and review of related standards," *Chinese Agricultural Science Bulletin*, vol. 29, no. 20, 2013.

[35] F. Meng, S. Luo, Y. Wu, H. Sun, and F. Liu, "Study on feature extraction and discriminant model of maize seed ear rot by near infrared spectroscopy," *Spectroscopy and Spectral Analysis*, vol. 42, no. 6, 2022.

[36] S. Li, G. Du, P. Li, L. Jiang, and X. Liu, "Nondestructive Identification of soybean milk Powder Based on near infrared spectroscopy and Optimized pretreatment method," *Food Research and Development*, vol. 41, no. 17, 2020.

[37] C. Gertz, A. Gertz, B. Matthäus, and I. Willenberg, "A systematic chemometric approach to identify the geographical origin of olive oils," *European Journal of Lipid Science and Technology*, vol. 121, no. 12, 2019.

[38] J. Wang, Y. Huo, Y. Wang et al., "Grading detection of "Red Fuji" apple in Luochuan based on machine vision and near-infrared spectroscopy," *PLoS One*, vol. 17, no. 8, 2022.

[39] Y. Xu, P. Du, R. Senger, J. Robertson, and J. L. Pirkle, "ISREA: an efficient peak-preserving baseline correction algorithm for Raman spectra," *Applied Spectroscopy*, vol. 75, no. 1, pp. 34–45, 2021.

[40] L. Rieppo, S. Saarakkala, T. Närhi, H. Helminen, J. Jurvelin, and J. Rieppo, "Application of second derivative spectroscopy for increasing molecular specificity of fourier transform infrared spectroscopic imaging of articular cartilage," *Osteoarthritis and Cartilage*, vol. 20, no. 5, pp. 451–459, 2012.

[41] H. Li, D. Wang, Y. Zhu, Z. Lin, and J. Zhang, "Analysis of formula ratio of cigarette cut by near infrared spectroscopy combined with pattern recognition method," *Physical and Chemical Examination*, vol. 202, no. 58, 2022.

[42] Z. Xia, J. Yang, J. Wang, S. Wang, and Y. Liu, "Optimizing rice near-infrared models using fractional order savitzky–golay derivation (FOSGD) combined with competitive adaptive reweighted sampling (CARS)," *Applied Spectroscopy*, vol. 74, no. 4, pp. 417–426, 2020.

[43] V. Sem, "Interpretability of selected variables and performance comparison of variable selection methods in a polyethylene and polypropylene NIR classification task," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 258, Article ID 119850, 2021.

[44] Z. Wang, Q. Wu, and M. Kamruzzaman, "Portable NIR spectroscopy and PLS based variable selection for adulteration detection in quinoa flour," *Food Control*, vol. 138, Article ID 108970, 2022.

[45] X. Cao, H. Ding, L. Yang et al., "Near-infrared spectroscopy as a tool to assist Sargassum fusiforme quality grading: harvest time discrimination and polyphenol prediction," *Postharvest Biology and Technology*, vol. 192, Article ID 112030, 2022.

[46] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference*, ACM, San Francisco, CA, USA, August 2016.

[47] S. Ribeiro, T. D. J. G. Salva, and M. Silvarolla, "Prediction of a wide range of compounds concentration in raw coffee beans using NIRS, PLS and variable selection," *Food Control*, vol. 125, Article ID 107967, 2021.

[48] J. Liu, J. Jia, H. Xue, X. Pan, and Y. Liu, "Prediction of milk protein content in milk for health hygiene based on hyperspectral and convolutional neural network," *Journal of Testing and Evaluation*, vol. 51, no. 1, Article ID 20210453, 2023.

[49] H. Xue, M. Sun, and Y. Liang, "ECANet: explicit cyclic attention-based network for video saliency prediction," *Neurocomputing*, vol. 468, pp. 233–244, 2022.

[50] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.

[51] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-net: efficient Channel Attention for deep convolutional neural networks," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Piscataway, NJ, USA, October 2020.

[52] J. Tian, W. He, S. Li, X. Hao, Y. Li, and L. Zhang, "Study on prediction method of reroasted flake tobacco origin based on cascade classification," *Journal of Light Industry*, vol. 37, pp. 94–99, 2022.

[53] J. K. Bernat, K. J. Jackson, E. J. Z. Krüsemann, S. Boesveldt, S. F. Rudy, and R. Talhout, "Sensory methods to evaluate perception of flavours in tobacco and other nicotine-containing products: a review," *Tobacco Control*, vol. 32, no. e1, pp. 95–102, 2023.