



## Regular article

## A residual neural network based method for the classification of tobacco cultivation regions using near-infrared spectroscopy sensors

Daiyu Jiang<sup>a</sup>, Guanqiu Qi<sup>b,\*</sup>, Gang Hu<sup>b</sup>, Neal Mazur<sup>b</sup>, Zhiqin Zhu<sup>a,\*</sup>, Di Wang<sup>c</sup><sup>a</sup> College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China<sup>b</sup> Computer Information Systems Department, State University of New York at Buffalo State, Buffalo, NY 14222, USA<sup>c</sup> School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400066, China

## ARTICLE INFO

## Keywords:

Tobacco leaves  
Near-infrared spectroscopy  
Residual network  
Focal loss  
Cultivation regions

## ABSTRACT

Near-infrared (NIR) spectroscopy techniques have been widely used to classify tobacco cultivation regions. NIR spectroscopy of tobacco leaves involves a large number of correlated features, so it is difficult to find the connection between spectral data and tobacco cultivation regions. This paper proposes a novel classification model of tobacco cultivation regions that integrates residual network (ResNet) and NIR spectroscopy techniques. As the number of neural network layers increases, the network may have issues such as network degradation, gradient disappearance, and the reduction of sample recognition rate. The proposed model applies a residual module to a neural network which effectively solves or alleviates the vanishing gradient issues caused by the increase of network depth. This paper also adds balance and suppression factors to the loss function to solve the issues caused by uneven sizes of tobacco samples collected from different cultivation regions in the training process. In the proposed method, tobacco samples are marked as internal and external samples respectively during the training process. Internal samples are collected from the corresponding cultivation regions in the north, northeast, and northwest of Guizhou Province, China. External samples are collected from other cultivation regions. The weight distributions of internal and external samples can be adjusted by experimental results to improve the identification accuracy of the proposed solution. The size of training samples determines the generalization ability of the network and affects the experimental results. A parametric rectified linear unit (PReLU) function is integrated into the network, in which the parameters of a linear unit are adaptively learned to further improve the identification accuracy of the proposed solution. Compared with current mainstream methods, the experimental results confirm that the proposed model is superior in accurately identifying different cultivation regions of tobacco leaves.

## 1. Introduction

The fragrance of tobacco leaves is treated as a significant factor of cigarette quality control [1,2]. Tobacco quality is evaluated by the following three criteria [3].

- Visible and detectable criteria: size, uniformity, finish, foreign matter, damage, color, texture (grainy, soft), body (thickness, density), maturity, odor and flavor;
- Physical criteria: filling power, shatter resistance, equilibrium moisture content, strip yield, combustibility and stalk position;
- Chemical criteria: nicotine, sugar, petroleum ether extract, mineral components, alkalinity of water-soluble ash, total N, protein N,  $\alpha$ -amino N, starch, non-volatile acids, and total volatile bases.

Generally, tobacco quality is determined by the interaction of many factors. The cultivation environment as a key factor contains insolation duration, temperature, precipitation, humidity, latitude, longitude, altitude, and soil condition [4,5]. So, the identification of cultivation regions widely used in the tobacco industry is an efficient way to improve the classification of tobacco leaves [6]. The early-stage manual classification cannot satisfy the increasing requirements of the large-scale tobacco industry. Therefore, it is important to establish a series of standards and a reliable method to accurately and effectively classify the cultivation regions of tobacco leaves. With the development of spectral sensor techniques, NIR spectroscopy techniques can effectively obtain abundant structure information to reflect the internal information of tobacco leaves without causing any damage to the leaves [7–9]. Therefore, NIR spectroscopy techniques have been widely used in the

\* Corresponding authors.

E-mail addresses: [daiyu93@gmail.com](mailto:daiyu93@gmail.com) (D. Jiang), [qig@buffalostate.edu](mailto:qig@buffalostate.edu) (G. Qi), [hug@buffalostate.edu](mailto:hug@buffalostate.edu) (G. Hu), [mazurnm@buffalostate.edu](mailto:mazurnm@buffalostate.edu) (N. Mazur), [zhuzq@cqupt.edu.cn](mailto:zhuzq@cqupt.edu.cn) (Z. Zhu), [diwang871106@gmail.com](mailto:diwang871106@gmail.com) (D. Wang).

tobacco industry to implement both qualitative and quantitative analysis of tobacco leaves [10–13].

Due to the high dimensionality and structure redundancy of the spectral data of tobacco leaves [14,7], it is difficult to build the connection between tobacco cultivation regions and the corresponding spectral data. Traditional statistical methods cannot effectively extract enough information from tobacco leaves, which may cause misidentification [14]. Machine learning methods have been widely used in the identification and classification of tobacco leaves [14,15]. As a classification method of tobacco cultivation regions, Cheng [7] proposed a feature grouping method based on a random forest to analyze tobacco leaves. Feature information is extracted from spectral data first, and then divided into different groups according to the weights of feature importance. Bin [16] proposed a modified random forest approach to classify the grades of tobacco leaves. A Monte Carlo algorithm is used to increase the diversity of classification trees in a random forest, and an uninformative variable elimination method is applied to the adaptive adjustment of variables. Ni [17] proposed an improved and simplified K-nearest neighbors (IS-KNN) classification algorithm to identify cultivation regions of tobacco leaves. The standard variance of repetition spectra collected from tobacco leaves is analyzed to obtain the maximum number of principal components first. According to the obtained principal components, the important spectra information is then selected for further analysis. Zhang [13] proposed a least-squares support vector machine (LS-SVM) method to classify the cultivation regions of tobacco leaves, in which principle component analysis (PCA) was applied to the spectra of tobacco leaves. The different numbers of the obtained principle components are processed by the proposed LS-SVM to obtain the corresponding number of principle components that can achieve the best classification performance. Wang [18] proposed a genetic algorithm support vector machine (GA-SVM) to identify the tobacco cultivation regions. In the mentioned work, the NIR spectroscopy of tobacco leaves is processed by PCA. The obtained principle components are ranked. Different combinations of principle components as inputs are trained by the proposed GA-SVM to find the optimal number of principle components for identification. Existing machine learning techniques can process low-dimensional data quickly. But they can not directly process high-dimensional raw NIR spectroscopy data. It is necessary to use pre-processing methods, such as wavelet transformation [2], standard normal variate (SNV) [8], multiplicative scatter correction (MSC) [19], and spectral derivative [2], to eliminate spectral interferences, such as light scattering and noise [20], before processing spectral data in different identification models. However, for the same NIR spectroscopy data, each pre-processing method obtains different results. Some important information in the original spectra may be lost. Compared with existing machine learning methods, deep learning methods do not require any data pre-processing, and have higher recognition accuracy of tobacco leaves.

In recent years, deep learning algorithms have shown great advantages in the analysis of NIR spectroscopy data [21,22]. Deep learning techniques can learn and extract data features from the input of classification models automatically [23,24]. Some existing research has applied NIR spectroscopy based deep learning methods to the analysis of tobacco leaves. Hana [25] employed a back propagation neural network (BPNN) to the classification of tobacco cultivation regions. However, BPNN requires many more training samples than conventional multivariate statistical methods in order to avoid overfitting. Wang [26] used a deep convolutional neural network (CNN) algorithm to classify the cultivation regions of tobacco leaves. The above mentioned deep learning methods explore the usage of neural networks in tobacco big data processing. These methods have high adaptability in data processing and do not require any data pre-processing since they can be applied to high-dimensional data directly. However, neural network based solutions do not have any ResNet structure. The vanishing gradient problem occurs during the big data processing, which can further cause the poor fitting in the big data

training and even low identification accuracy. Existing deep learning methods only achieve 93% accurately identify tobacco cultivation regions [26].

ResNet as a deep learning method has been widely used in data analysis. He [27] proposed a ResNet that included the basic structure of the residual module and the calculation method. The proposed residual module achieves identity mapping between the shallow network data and the deep network, which solves the degradation issues in a traditional deep convolutional network. ResNet has been applied to the classification of hyperspectral images in existing solutions. Gao [28] proposed a pre-activation ResNet for the classification of hyperspectral images that integrated both spectral and spatial information. In the proposed ResNet, the pre-activation and batch normalization (BN) operations are used to effectively retain the features learned by the residual block, which ensures the extraction of both spectral and spatial features and enhances the generalization performance of the neural network. Paoletti [29] proposed a ResNet model based on residual blocks of pyramidal bottleneck, which used both spectral and spatial information to achieve the fast and accurate analysis and classification of hyperspectral images. The proposed ResNet model consists of multiple convolutional layers and residual blocks, in which the connection between spectral and spatial information of hyperspectral data is built.

In order to improve the recognition accuracy of tobacco leaves, this paper proposes a ResNet based classification model to classify tobacco cultivation regions by using NIR spectroscopy data of tobacco leaves called 1-dimensional ResNet (1D-ResNet). The proposed solution uses a convolutional neural network to extract key features from NIR spectroscopy data. The proposed solution can adjust the network input size to fit different NIR spectroscopy data. The training data of tobacco leaves is used as the input of the 1D-ResNet model, and the NIR spectral data features of tobacco leaves are extracted by using an 1D convolutional layer. The network output is processed by a softmax function. As the number of neural network layers increases, the gradient may disappear in the training process [30,27], which reduces the identification rate of tobacco leaves. ResNet can effectively solve or alleviate the issues of gradient disappearance caused by the increase of network depth [27,31]. In this paper, the tobacco samples are marked as internal and external samples. Internal samples are the classified tobacco leaves that are collected from the corresponding cultivation regions in the north, northeast, and northwest of Guizhou Province, China. External samples are the tobacco leaves collected from other cultivation regions outside Guizhou Province. Different samples of NIR spectroscopy have repetition and non-repetition spectra. Due to the imbalanced size of samples and the high volumes of information contained in samples, the identification difficulty of internal and external samples varies.

The balance and suppression factors are added to the loss function to solve the issue of unbalanced tobacco samples. The optimized focal loss function [32] can adjust the weight distribution of internal and external tobacco samples and alleviate the imbalance of tobacco leaves. A BN operation [33] is applied to each convolutional layer. BN can speed up the network training and improve the generalization ability of the network [33,34]. A PReLU function is used in both the convolutional layer and max-pooling layer [35]. PReLU adds a self-learning parameter called a linear unit to the negative part of the activation function. The linear unit can adaptively train during the model training process. Compared with other existing activation functions, the slope of the negative part of the activation function is used to avoid gradient vanishing and increase the sparsity of network, which can make the neural network effectively extract the key features of NIR spectroscopy data [36,37]. This paper has two main contributions:

1. A novel neural network with two residual blocks is proposed to improve the feature extraction ability of high-dimensional and redundant spectral data. Due to the high dimension of tobacco spectral data, a deep neural network is used to extract the features of spectral data. Two residual blocks are added into the proposed neural

- network to do identity mapping, which can map the data in a shallow layer to a deep layer. The identity mapping can effectively solve the issue of gradient disappearance that can increase the feature quantity of a deep neural network.
2. The balance and suppression factors are added to the loss function of the proposed neural networks, which solve the issue in training unevenly distributed data. The spectral data size of tobacco collected from different tobacco cultivation regions varies. In the training process of a neural network, the balance and suppression factors add more weight to small-size data samples. Therefore, the neural network can have the same classification ability for each data sample.

The remaining sections of this paper are organized as follows. Section 2 discusses the proposed residual neural network in detail; Section 3 shows the comparative experiments and analyzes the related results; Section 4 concludes this paper.

## 2. The proposed 1D residual neural network

As one of the most groundbreaking works in the deep learning field, ResNet has been widely discussed [31]. ResNet introduces an "identity skip connection" to the residual block. The "identity skip connection" skips multiple layers in the network and uses the output of the previous layer as a partial input of the subsequent layer, which can effectively reduce the negative effects of gradient vanishing and degeneration caused by an increase of network depth. The structure of ResNet ensures the identity mapping in the training process. The existence of a layer can be determined by the self-training of ResNet. Thus, ResNet can train thousands of layers with compelling performance. ResNet is generally used in 2-dimensional image classification and object recognition, and achieves excellent performance on large-scale well-labeled datasets [38]. The NIR spectroscopy data usually contains wide spectral bands, severe spectral peak overlap, complex internal information, and redundant structures. In this paper, the NIR spectroscopy data of tobacco leaves has a sequence of 1,609 signals [39]. Due to the data redundancy, useful information is not contained in every dimension. In order to effectively extract the spectral information, a novel 1D-ResNet is proposed to analyze NIR spectroscopy data and classify tobacco cultivation regions. The original NIR spectroscopy data is used as the input of 1D-ResNet, where the key data features with reduced dimensions can be learned by deep convolutional operations.

### 2.1. 1D-ResNet architecture design

The proposed 1D-ResNet mainly consists of five convolutional blocks, two residual blocks, and three fully connected layers. Each block may have one or multiple layers. Each convolutional block is coupled with a max-pooling layer to reduce the number of parameters, while the salient features are retained. Each residual block links to two convolutional blocks via an identity skip connection. The diagram of the overall network structure is shown in Fig. 1, where the blue, green, and gray cuboids are convolutional, max-pooling, and fully connected layers respectively. Two black lines represent the identity skip connections between two convolutional blocks through the residual blocks. This is an end-to-end classification network, where the input is a sequence of 1,609 NIR spectral signals, and the final output is a predicted result from the softmax layer, which performs multi-class classification [40].

Unlike the regular image classification task, the current NIR dataset of tobacco leaves is limited to a relatively small data size with imbalanced samples. This causes difficulties during the network training, such as overfitting, slow convergence, and poor prediction accuracy. Therefore, the proposed classification solution is developed based on ResNet integrating PReLU and focal loss components. Specifically, after each convolution operation, BN is applied to speed up the convergence

process. A dropout layer is added to the break-up network co-adapt layers to correct the mistakes from prior layers. The dropout layer reduces the overfitting risk in turn and makes the proposed model more robust. The default activation function is replaced by PReLU, which can adaptively learn the parameters of rectifiers to accelerate the convergence and improve the accuracy at a negligible computational cost. Finally, a balanced focal loss function is used to rectify the weight distribution of imbalanced tobacco samples. In the following sub-sections, residual block, normalization, activation function, and loss calculation will be elaborated.

### 2.2. ResNet neural network classification model

In the proposed 1D ResNet, the convolutional layer is used to extract features from low- to high-level of NIR tobacco data. The number of parameters is reduced by the max-pooling layer. In the fully connected layer, the high-level features are processed by combining every data pair based on a complete graph. The output of the last fully connected layer has four neurons that will be used in the softmax classifier to identify different cultivation areas. As a part of residual block, identity skip connection is a channel of convolutional operation, in which the output of previous block and the input of next block can be processed by point-to-point addition [27]. The second and fourth stages of the proposed 1D-ResNet use the identity skip connection. The identity skip connection allows the output of the first and third stages as the input to the third and fifth stages respectively. The network structure is shown in Fig. 2. Sky blue, beige and fuchsia cuboids are the convolutional layers (Conv), the max-pooling layer (MaxPool), and the fully connected layers (FC) respectively.

### 2.3. Residual block and batch normalization

As the number of layers in a deep neural network increases, the gradient signal will be gradually reduced as it passes through each layer during back propagation. Eventually the gradient becomes zero in shallow layers, then the network stops converging. This is a harmful learning phenomenon called gradient vanishing. Similar to the original ResNet architecture [41,31], the residual block of the proposed network shown in Fig. 3 can effectively alleviate this issue. In the shortcut layer, an identity skip connects the input  $x$  to the convolutional output  $F(x)$  directly by performing element-wise addition  $H(x) = F(x) + x$  [41,31]. Part of  $H(x)$  is derived from  $x$  directly without the convolutional process  $F(x)$ . During the back propagation, the proposed network can make the gradients received at  $H(x)$  equally flow back into  $x$  and  $F(x)$ . Even if the gradients are vanished in  $F(x)$ , its shallow block can still get the updated signals through the branch  $x$  without any reduction. This simple addition does not need any extra parameters but can greatly improve the training effect. However, extra addition would cause data shifting to bias the results in turn. Although the network is able to slowly update the related parameters through the learning process, extra intervention is needed. Thus, BN layer is introduced to the residual block to quickly achieve scaling and panning invariance [33], while preserving the characteristics of the original input. BN layer does the scaling and shifting operations on  $x$  before sending to neurons. This process can ensure the network capacity and shorten the network learning curve. Specifically, the feature data of tobacco leaves is divided into small batches  $B = \{s_1, s_2, \dots, s_m\}$  for mini-batch training. After the convolution operation, the output is obtained by Eq. (1).

$$s = wz + b \quad (1)$$

where  $z$  is the input data,  $w$  and  $b$  are the weight and bias of the convolutional layer before the BN layer respectively, and  $s$  is the input of BN layer. The average of mini batch data is shown as follows.

$$\mu_B = \frac{1}{m} \sum_{i=1}^m s_i \quad (2)$$

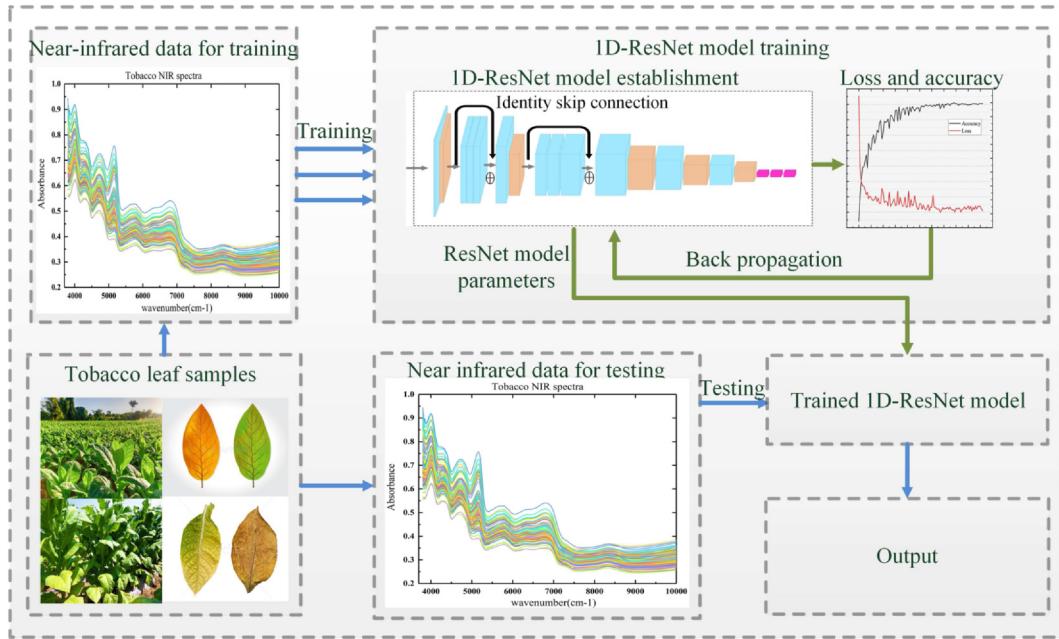


Fig. 1. The Flowchart of the Proposed 1D-ResNet.

where  $m$  represents the number of input data within a batch,  $\mu_B$  is the mean value of this batch. The standardization formula is shown as follows.

$$\hat{s}_i = \frac{s_i - \mu_B}{\sqrt{\frac{1}{m} \sum_{i=1}^m (s_i - \mu_B)^2 + \varepsilon}} \quad (3)$$

where  $\hat{s}_i$  is the normalized value of input data  $s_i$ ,  $\varepsilon$  as a small positive number is used to avoid the error caused by variance zero and is set to 0.001 in this work [33]. For each input  $s_i$ , parameter  $\gamma_i$  and  $\beta_i$  are further used to obtain the final normalized value by scaling and translation as follows.

$$y_i = \gamma_i \hat{s}_i + \beta_i \quad (4)$$

where  $y_i$  as the output of the BN layer is the value obtained after the scaling and translation of  $\hat{s}_i$ , both  $\gamma_i$  and  $\beta_i$  are learnable. In summary, the normalization of a tobacco data batch is achieved by calculating the mean and variance of each batch, and performing the scaling and shifting based on learnable hyperparameters.

#### 2.4. Activation

The activation function mainly solves the linear indivisible issue of the neural network. It adds non-linear factors into the neural network to improve fitting ability of network training [35,42]. The original activation function of rectified linear unit (ReLU) [36] zeros out all negative inputs, which makes some neurons "dead" (the "dead" neurons will not be updated during the training process). The "dead" neurons do not play any role in network and are completely useless. Too many zero

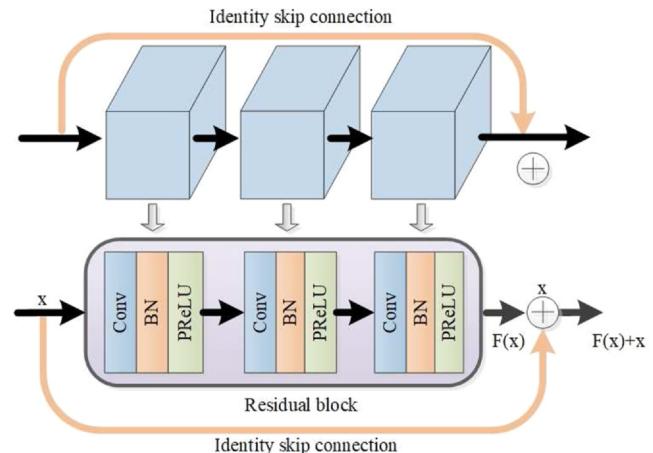


Fig. 3. The Structure of the residual block.

values of output cause network sparsity, reduce the dependence among different parameters, and inhibit overfitting. PReLU uses a different strategy to mitigate the overfitting issue. PReLU returns a certain percentage value from a negative input to enhance the network learning ability, which can reduce the number of "dead" neurons. As shown in Eq. (5),  $\alpha_i$  as the learnable parameter is added to the negative part of activation function, so the fitting ability of the model improves around the zero value. In the proposed network, PReLU activation function is used after each convolutional and fully connected layer.

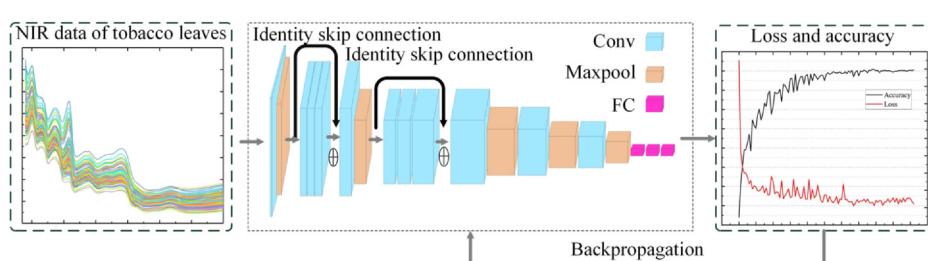
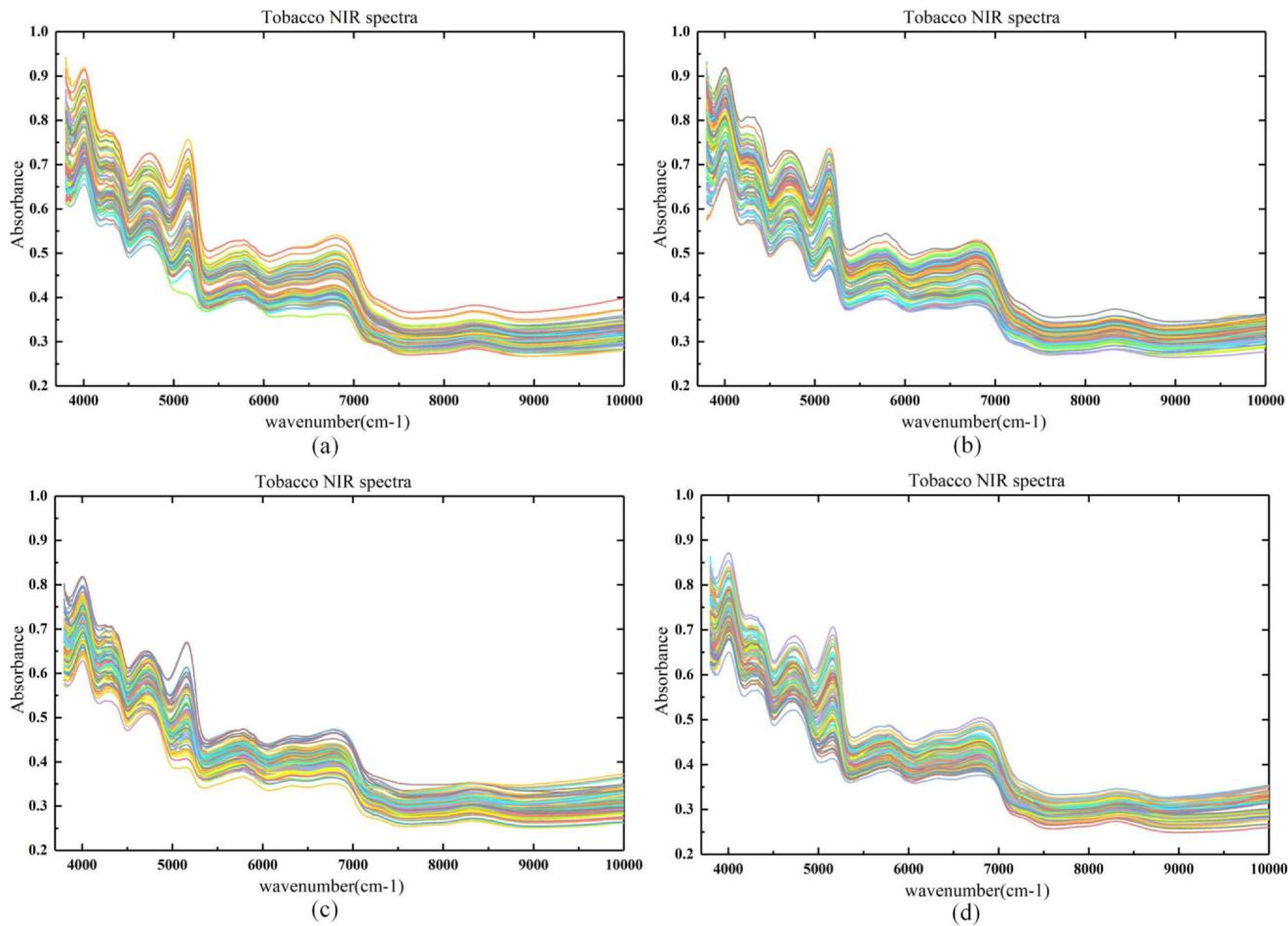


Fig. 2. ResNet neural network classification model.



**Fig. 4.** The Raw NIR Spectra of Tobacco Samples from Different Regions. (a) represents the tobacco samples from the northern region of Guizhou Province, (b) represents the samples from northeast region of Guizhou Province, (c) represents the samples from the northwest region of Guizhou Province, and (d) represents the external samples out of Guizhou Province.

$$\text{PReLU}(x_i) = \begin{cases} x_i, & \text{if } x_i > 0 \\ \alpha_i x_i, & \text{if } x_i \leq 0 \end{cases} \quad (5)$$

where  $x_i$  is the input of the activation function on the  $i$ th channel, and  $\alpha_i$  controls the slope of the negative part. PReLU in the proposed neural network can adaptively learn the parameter  $\alpha_i$  to improve the whole network [37]. When  $\alpha_i = 0$ , PReLU degenerates into a ReLU.

## 2.5. Loss function

The training performance of a neural network is determined by a loss function [37,43]. At the end of the network forward stage, the output of the softmax classifier is a probability value of class  $y_i$  within the range of (0, 1) as shown in Eq. (6).

$$\text{SoftMax}(y_i) = \frac{e^{y_i}}{\sum_{j=1}^n e^{y_j}} \quad (6)$$

where  $y_i$  is one of tobacco cultivation regions in the proposed solution [40], and  $n$  is the total number of categories ( $n$  is four in this study). As a regular method of neural network classification model, a cross entropy (CE) loss function is introduced to measure the distance between the predicted score and the ground truth. Given two probability distributions  $p$  and  $q$  from softmax representing the true region and the predicted region respectively, CE is defined in Eq. (7).

$$CE(q(x_i)) = - \sum_{i=1}^n p(x_i) \ln q(x_i) \quad (7)$$

where  $p(x_i)$  is the true probability distribution of sample  $x_i$ ,  $q(x_i)$  is the predicted probability distribution obtained by the softmax classifier.

The number of samples affects the accuracy of the training result. It is difficult to train an accurate model with small-size labeled samples. Conversely, it is easier to learn a better model with more sample data because network parameters can be updated more thoroughly. In this work, tobacco samples from different regions have imbalanced sizes. Since deep network classification is data driven, some regions with overwhelming sample data would receive most of the training efforts, while other with fewer samples would suffer an insufficient training problem. When most of the training efforts contribute to few classes, it leads to a degenerate model. To tackle the issue of imbalanced data, a focal loss (FL) function is introduced in Eq. (8).

$$FL(q(x_i)) = -\alpha_t(1 - q(x_i))^\gamma \ln q(x_i) \quad (8)$$

where  $\alpha_t$  is the balance parameter,  $\gamma$  is the penalty item to adjust the rate of weight reduction for simple samples.

Specifically, if  $\gamma = 0$ , Eq. (8) is just a standard cross-entropy loss. When the  $\gamma$  value increases, the loss of large-size data samples would be suppressed (become smaller), and the loss of small-size data samples would be enlarged accordingly. In this way, the regions (classification categories) with a large number of samples receive relatively less training efforts. On the other hand, the regions with a small number of samples get relatively more training (more parameter updates). Therefore, the focal loss function reduces the negative issues from the imbalanced data of tobacco leaves.

### 3. Experiments and analyses

This section describes the NIR dataset of tobacco leaves, the experiment platform, the network settings, detailed experiments, and result analysis.

#### 3.1. Experiment preparation and sample acquisition

Following the development of Internet techniques and cloud infrastructure, more and more transaction processing systems have been hosted in cloud platforms [44–46]. Cloud-based platforms can use numerous and powerful computation resources within reasonable costs [47–49] to implement the analysis of large-size tobacco data. The training and testing of the proposed solution are conducted on a small cloud environment that consists of six machines with NVIDIA GTX 2080 GPU, Intel(R) Core(TM) i7-8700 3.20 GHz CPU and 24 GB RAM. The neural network is built in the deep learning framework Tensorflow 2.0 on Windows Server 2019 operating system. The training and testing codes are programmed in Python using Keras library.

In comparative experiments, in order to make the NIR spectra of the sampled tobacco leaves represent all the characteristics of an entire tobacco plant, leaves are picked from different parts of a tobacco plant. The NIR spectra of tobacco leaves were recorded by Thermo Antaris 2 with multiple sensors (Thermo Fisher Scientific Inc. Waltham, MA, USA). Each spectrum is the average of 64 scanned interferograms at 8 cm<sup>-1</sup> resolutions. NIR range is from 3800 cm<sup>-1</sup> to 10,000 cm<sup>-1</sup>, which is shown in Fig. 4. The spectra of each sample were saved by NIR sensors and represented as the corresponding digital absorbance values at different wavelengths. OMNIC analysis software is used to read the spectra and convert them into NIR spectral data with a dimension of 1,609, which is saved in spa format. Each spectrum size is about 100 KB. The converted digital absorbance values are used as the inputs of the proposed classification model in this study.

As a landlocked and mountainous province in Southwest China, Guizhou has a humid subtropical climate with few seasonal changes [50]. Its annual average temperature is roughly 10 °C to 20 °C, with January temperatures ranging from 1 °C to 10 °C and July temperatures ranging from 17 °C to 28 °C [50]. The unique natural conditions create a favorable environment for tobacco cultivation. As the third largest producer of tobacco in China, the tobacco cultivation is spread all over Guizhou Province [50]. However, the tobacco quantity varies in different cultivation regions of Guizhou Province.

As shown in Fig. 5, a large number of samples were collected from the tobacco cultivation regions located in the northern, northeastern, and northwestern regions of Guizhou Province in four consecutive years. Due to the large size of tobacco leaves, the manual identification

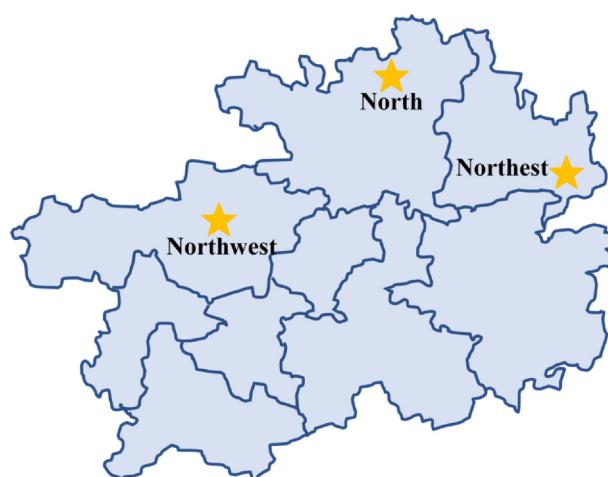


Fig. 5. Distribution of Tobacco Cultivation Regions in Guizhou Province.

Table 1

Training and testing sample sizes of tobacco from different regions.

Class	Region	Number of Samples		
		Training Set	Testing Set	Total
1	North	2325	572	2897
2	Northeast	1137	285	1422
3	Northwest	6116	1552	7668
4	External samples	1118	265	1383

of tobacco cultivation areas is highly challenging.

The comparative experiments use 13,370 tobacco samples for demonstration. About 80% and 20% of the collected data samples are randomly selected and used as training and testing sets respectively. The training set is used to train the classification model and adjust the corresponding parameters. As mentioned in Section 2, the effectiveness of the training method used was confirmed by existing research work [18,51]. The testing set is used to evaluate the recognition rate of the proposed classification model. As the internal samples, 11,987 tobacco samples are selected from the raw data of Guizhou tobacco cultivation regions. Additionally, an extra 1,383 tobacco samples were collected from other tobacco cultivation regions outside Guizhou Province as external samples. Table 1 lists the detailed numbers of tobacco samples and the related cultivation regions.

#### 3.2. ResNet network design

NIR spectroscopy data is used to identify and classify the tobacco cultivation regions. As the sequential NIR data, the input data is set to 1,609 spectral elements in a single channel, and the final output is the classification result of tobacco cultivation regions. It uses  $w \times h \times c$  to denote the data shape, where  $w$  and  $h$  represent the width and height respectively, and  $c$  represents the channel size.

The proposed 1D-ResNet has six stages including the first to fifth stages and an output stage. The second and fourth stages are residual blocks 1 and 2 respectively. The structure of a residual block is shown in Fig. 3. Each residual block contains three convolutional layers.

The detailed structures of the proposed 1D-ResNet are shown in Table 2. Data features are extracted in Conv. The results of BN are restored to the original input characteristics by scaling and shifting. Maxpool is short for max-pooling layer, which is used to reduce the number of parameters. Reshape represents the reshape layer, in which the data size is adjusted from  $1 \times 51 \times 512$  to  $1 \times 26, 112$ . Dropout layer is used after each fully connected layer. Softmax is a classifier that is used to classify four types of tobacco cultivation regions. PReLU is the activation function existing in each layer of the neural network. The fully connected layers use dropout to avoid overfitting. Dropout makes the neurons of the fully connected layers stop working at a probability of 0.5 in each batch of training, which means that the value of the activation function becomes zero at a probability of 0.5. In this paper, the probability is set to 0.5, which means half of nodes in the fully connected layer stop working.

#### 3.3. Network training

During the network training, the batch size of training data is set to 16, and the total number of training iterations is set to 450,000. The initial learning rate is set to 0.001. The learning rate of attenuation coefficient is set to 0.99 and is updated every 6000 iterations. The initial weights of the convolutional layer and the fully connected layer are set by using the "msra" method [35].

The exact recognition rate and the loss value of the training set are shown in Fig. 6. The black and red curves represent the trend of accuracy and the loss value respectively. The ordinates on the left and right sides represent the accuracy rate and the loss value respectively.

**Table 2**

Detailed structure of 1D-ResNet.

Stage	Detailed Structure	Convolution Kernel		Output size
		Size	Number	
First stage	Conv1 + BN + PReLU	1 × 9	32	1 × 1, 609 × 32
	Maxpool1	2	-	1 × 805 × 32
Second stage (also called Residual block1)	Conv2 + BN + PReLU	1 × 3	64	1 × 805 × 64
	Conv3 + BN + PReLU	1 × 3	64	1 × 805 × 64
	Conv4 + BN + PReLU	1 × 3	64	1 × 805 × 64
Third stage	Conv5 + BN + PReLU	1 × 3	128	1 × 805 × 128
	Maxpool2	2	128	1 × 403 × 128
Fourth stage (also called Residual block2)	Conv6 + BN + PReLU	1 × 3	128	1 × 403 × 128
	Conv7 + BN + PReLU	1 × 3	128	1 × 403 × 128
	Conv8 + BN + PReLU	1 × 3	256	1 × 403 × 256
Fifth stage	Conv9 + BN + PReLU	1 × 3	512	1 × 403 × 512
	Maxpool3	2	-	1 × 202 × 512
	Conv10 + BN + PReLU	1 × 3	512	1 × 202 × 512
	Maxpool4	2	-	1 × 101 × 512
	Conv11 + BN + PReLU	1 × 3	512	1 × 101 × 512
	Maxpool5	2	-	1 × 51 × 512
Output stage	Reshape	-	-	26,112
	FC1 + PReLU + Dropout	-	-	262,144
	FC2 + PReLU + Dropout	-	-	262,144
	FC3 + PReLU + Dropout	-	-	4
	Softmax	-	-	4

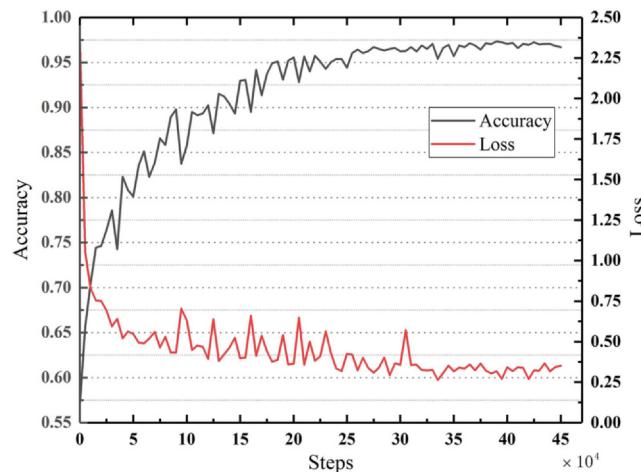


Fig. 6. Network convergence on the training set.

The abscissa represents the number of training iterations which has a value of 450,000.

In Fig. 6, the accuracy of the network tends to rise rapidly and then gradually slows down until the final stabilization. After 250,000 iterations of network training, the training accuracy rate is 0.9442. Then the training accuracy rate increases to 0.9669 and starts to converge at 360,000 iterations. The training accuracy rate reaches the highest value 0.9694 at 450,000 iterations. The loss value shows a rapid downward trend at the beginning of network training, and then slowly decreases until convergence. The training loss value is 0.3600 at 350,000 iterations, and finally decreases to 0.3506 at 450,000 iterations.

#### 3.4. Ablation study and activation function

As an important part of a neural network [35,36], the activation function adds nonlinearity to the network that makes the network more expressive. As the default activation function of the proposed 1D-

ResNet neural network, PReLU is compared with three other activation functions including hyperbolic tangent (tanh) [52], ReLU [36], and exponential linear units (ELU) [53] to evaluate the effectiveness.

Prediction accuracy ( $P_a$ ) as an important metric is used to evaluate the overall classification performance of tobacco cultivation regions, which is defined as follows.

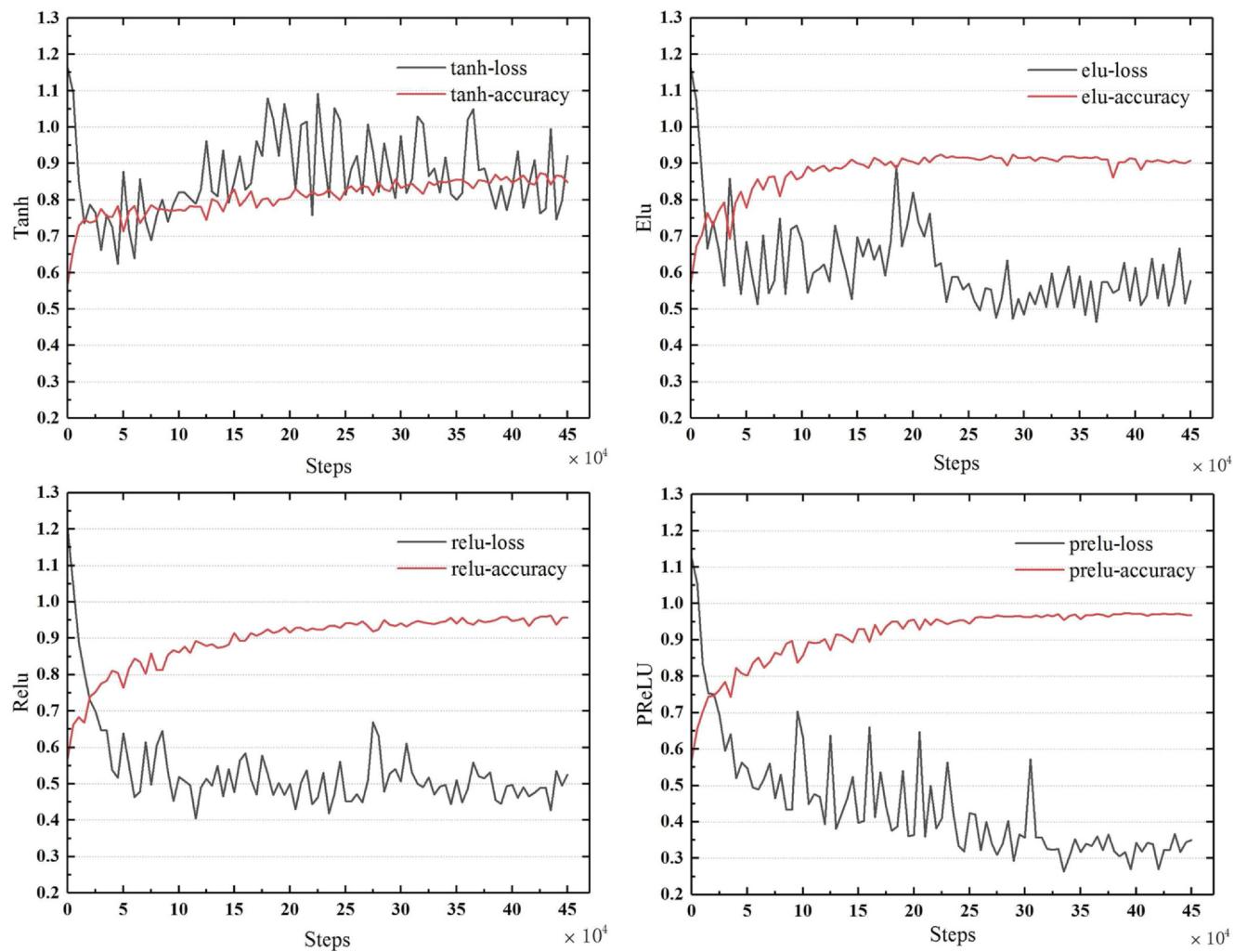
$$P_a = \frac{n_r}{N_t} \quad (9)$$

where  $n_r$  is the number of correctly predicted samples and  $N_t$  is the number of samples for prediction. The comparison results of accuracy and loss value are shown in Fig. 7. Both prediction accuracy and running time of different activation functions are shown in Table 3.

In Fig. 7, the loss value of the tanh function increases during the training, and the prediction accuracy is 0.8613 at 450,000 iterations, which is worse than ELU, ReLU, and PReLU. The tanh function has a saturation issue. Its output interval is  $[-1, 1]$  and is centered at 0. Its sensitivity is lost as the value exceeds the corresponding interval boundaries.

For the less than zero part of exponential linear unit, the ELU function adopts a negative exponential form. ELU has soft saturation characteristics when the input is less than zero, which improves the robustness to noises. However, according to the experiment results, the identification accuracy of ELU varies. The prediction accuracy of the ELU function is 0.9142, which is slightly worse than ReLU. Additionally, ELU is slow with approximately 1.2688 s per round.

As shown in Fig. 7 and Table 3, the ReLU function has a relatively high prediction accuracy 0.9561. The positive data from ReLU is not saturated, while negative data becomes zero. Therefore, it makes many neurons "dead", which causes the network to become sparse. The PReLU function gives a negative slope to the negative value of the training session, which effectively alleviates the issue of neuron "death" in ReLU function. As shown in Fig. 7 and Table 3, the activation function PReLU achieved the most stable performance in terms of both accuracy and loss values. Its training accuracy reaches the highest value 0.9694 at 450,000 iterations. Meanwhile, it only takes 0.8088 s per training iteration, which is the most efficient among all four functions.



**Fig. 7.** Variations in accuracy and loss values for different activation functions.

**Table 3**  
Prediction accuracy and time under different activation functions.

Activation Function	$P_a$	Time (s)
tanh	0.8613	1.8212
ELU	0.9142	1.2688
ReLU	0.9561	0.9644
PReLU	0.9694	0.8088

The PReLU activation function only adds a small number of parameters, so its computation complexity is low. In summary, compared with three other activation functions, PReLU makes the proposed network perform the best.

### 3.5. Ablation study on loss function and training result analysis

In this study, tobacco samples are imbalanced, which causes difficulty in classifying different cultivation regions of tobacco leaves. The balanced focal loss function is applied to solve the classification issue.

To evaluate the performance of FL, CE is used in the proposed solution for comparison. The initial number of iterations is 10,000 in the comparative experiments. The comparison results are shown in Fig. 8. For classification accuracy, FL is higher than CE. For the training loss, FL is lower than CE.

According to multiple experimental attempts, the balance parameter  $\alpha_t$  and the suppression parameter  $\gamma$  as the penalty item are set to

0.2 and 1.5 respectively in the balanced FL function. The training accuracy and loss of FL are 0.9694 and 0.3506 respectively, which are 0.0201 higher and 0.1189 lower than the corresponding result of CE. Therefore, the optimized FL function can adjust the weight distribution effectively and alleviate the issue of imbalanced tobacco samples.

To further investigate the impact of different loss functions, the training performance of each individual region (class) is explored. As shown in Table 1, 2325, 1137, 6116, and 1118 samples from north, northeast, northwest, and external regions respectively are used to train four classes of tobacco leaves. In Fig. 9, four sub-figures from top left, top right, bottom left to bottom right compare the training accuracy of north, northeast, northwest, and external samples obtained by CE and FL respectively. The abscissa represents the number of training iterations and the ordinate represents the accuracy. As shown in all four sub-figures, every class demonstrates that FL has better prediction accuracy than CE.

As previously explained, the number of samples affects the network accuracy. If the number of a certain class is large, this class is relatively easy to train. For example, according to Table 1, the northwest has the largest number of samples (2325) among the four cultivation regions. So the northwest region has much higher training accuracy than the other three. On the other hand, the number of training samples from the external region (1118) is the smallest, and its training accuracy is obviously the worst by using either CE or FL loss function, especially during the iteration between 10,000 and 100,000. However, as shown in the bottom-right sub-figure of Fig. 9, the training difficulty of small-

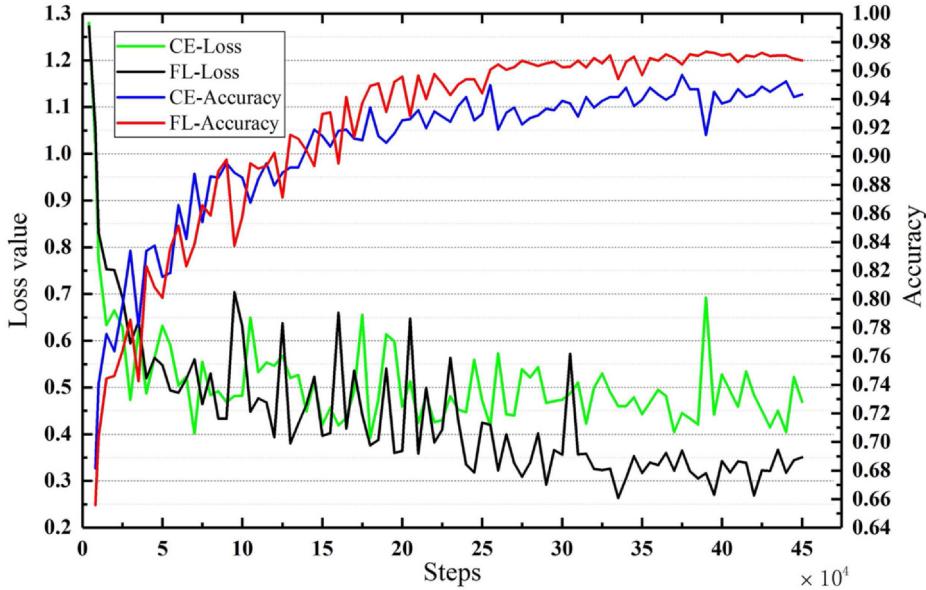


Fig. 8. Comparison of cross entropy and focal loss on the training set.

size samples is mitigated with the help of FL function. The prediction accuracy of the external regions (shown in the black curve) obtained by FL increases quickly in a consistent pace overall, while the accuracy trend of CE loss function (shown in the red curve) fluctuates a lot

between the iteration number 100,000 and 300,000. When the training reaches 450,000 iterations, the accuracy of FL function becomes stable reaching its best value which is higher than CE function by a large margin. Compared with regular CE loss function, FL function is able to

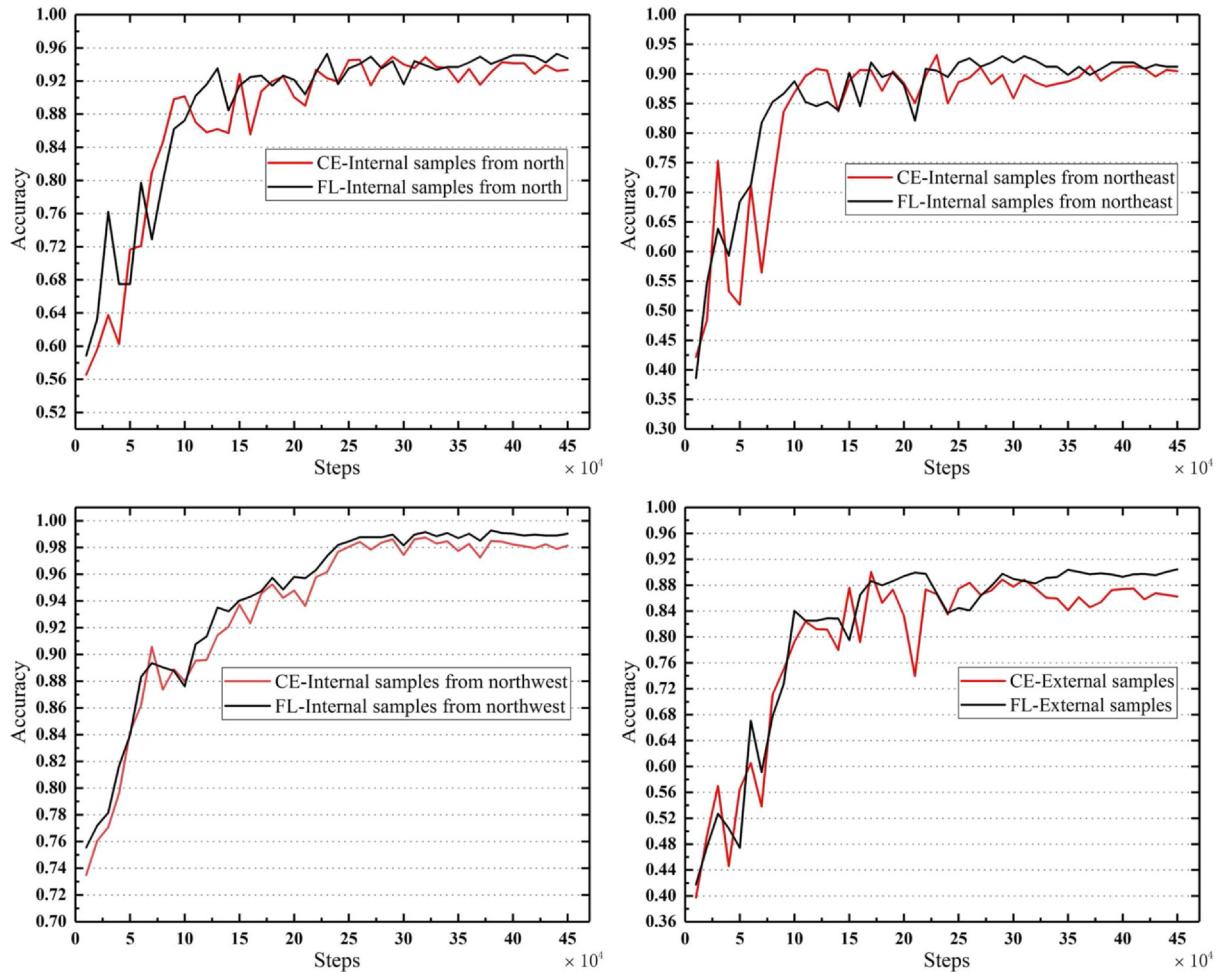


Fig. 9. Training results of the proposed network.

**Table 4**  
The definition of given and predicted labels.

	Predicted label			
	Positive	Negative		
Given label	Positive	$n_{TP}$	Negative	$n_{FN}$
	Negative	$n_{FP}$		$n_{TN}$

effectively adjust the loss weights for the regions with imbalanced data samples.

### 3.6. Evaluation of the proposed 1D-ResNet classification model

As shown in Table 4,  $n_{TP}$ ,  $n_{TN}$ ,  $n_{FP}$ , and  $n_{FN}$  are defined in the evaluation criteria of the models.  $n$  is the number of samples.  $n_{TP}$  (TP is short for true positive) is the number of true positive samples that are predicted as positive by the proposed model.  $n_{TN}$  (TN is short for true negative) is the number of true negative samples that are labeled as negative by the proposed model.  $n_{FP}$  (FP is short for false positive) is the number of false positive samples that are labeled as positive by the proposed model.  $n_{FN}$  (FN is short for false negative) is the number of false negative samples that are labeled as negative by the proposed model. The formulas of the evaluation criteria are given as follows.

$$\gamma_{TP} = \frac{n_{TP}}{n_{TP} + n_{FN}} \quad (10)$$

$$\gamma_{TN} = \frac{n_{TN}}{n_{TN} + n_{FP}} \quad (11)$$

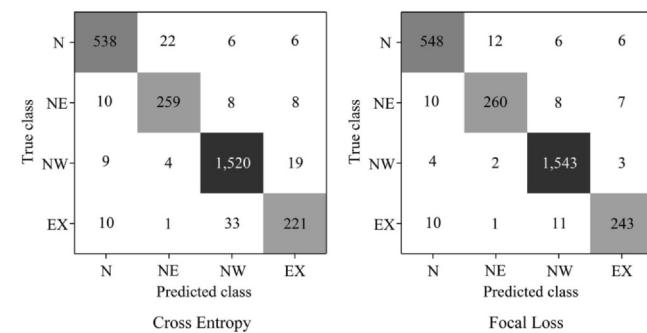
$$\gamma_{PP} = \frac{n_{TP}}{n_{TP} + n_{FP}} \quad (12)$$

$$\gamma_{F1-score} = \frac{2n_{TP}}{2n_{TP} + n_{FP} + n_{FN}} \quad (13)$$

where  $\gamma_{TP}$  is the sensitivity rate, which measures the detection ability of positive patterns;  $\gamma_{TN}$  is the specificity rate, which is the ability to specify the negative patterns;  $\gamma_{PP}$  is the precision rate, which is the ability to predict the positive patterns; and  $\gamma_{F1-score}$  is F1-score [18], which considers both the precision and sensitivity of tests.

As shown in Table 5, 572, 285, and 1552 tobacco samples from north, northeast, and northwest of Guizhou Province and outside Guizhou Province respectively are used as the testing data. The comprehensive assessment of 1D-ResNet classification model on the testing dataset was evaluated by the metrics in Eqs. (9)–(13). The results are shown in Table 5, where N, NE, NW, and EX represent north, northeast, northwest, and external samples respectively. For each evaluation parameter, a larger value indicates a better prediction performance.

As shown in Table 5, all the results of FL marked in bold are better than the corresponding results of CE. For FL, the prediction accuracy  $P_a$  of north, northeast, northwest, external samples are 0.9818, 0.9848, 0.9870, and 0.9855 respectively, which are more than 0.98. For each evaluation indicator, the corresponding result of each region fluctuates. The larger-size samples do not always have the better result. All the



**Fig. 10.** Two confusion matrices of the testing set obtained by using CE and FL respectively.

evaluation indicators are affected by the number of samples and the quality of samples. If sample features are obvious and the same region samples have high similarity, the identification performance of each evaluation indicator is relatively better. Comparing with CE, FL can effectively improve the identification performance, when the sample quality is not stable.

### 3.7. Analysis of testing results

To further evaluate the FL performance in each type of data, the proposed 1D-ResNet uses CE and FL respectively. As shown in Fig. 10, two confusion matrices list the classified results of four tobacco regions obtained by using CE and FL respectively. The values on the diagonal line are the number of correctly predicted samples. The values on the off-diagonal cells are the number of wrongly predicted samples. As shown in Table 6, FL has better testing accuracy than CE in all four tobacco regions and in the overall result.

To evaluate the prediction accuracy, the proposed 1D-ResNet is compared with four existing learning models: SVM, GA-SVM, ANN, and CNN. As shown in Fig. 11, the prediction accuracy of the ANN model is lower than SVM and GA-SVM. ANN is a simple neuron network, which cannot interpret the high-dimensional data from NIR tobacco information. SVM seems more suitable for a small sample size of high-dimensional data. The proposed optimized 1D-ResNet network has the best prediction accuracy 0.9701 in tobacco sample classification, which is 0.1064, 0.0633, 0.1167, and 0.0385 higher than SVM, GA-SVM, ANN, and CNN respectively.

## 4. Conclusion

This paper proposed a novel 1D-ResNet classification model to classify tobacco cultivation regions by using NIR spectroscopy data of tobacco leaves. The standard samples of tobacco leaves were collected from three different regions in Guizhou Province. The proposed 1D-ResNet can effectively solve or alleviate the issue of gradient disappearance caused by the increase in network depth, and improve the network recognition effect. In this paper, the data of tobacco samples was imbalanced, and the identification difficulty of internal and

**Table 5**  
Identification capacity of cross entropy and focal loss in 1D-ResNet for the Testing Set in Terms of Five Evaluation Parameters. ( $P_a$ :prediction accuracy;  $\gamma_{TP}$ : sensitivity rate;  $\gamma_{TN}$ : specificity rate;  $\gamma_{PP}$ : precision rate;  $\gamma$ : F1-score)

Region	Sample #	$P_a$		$\gamma_{TP}$		$\gamma_{TN}$		$\gamma_{PP}$		$\gamma$	
		CE	FL	CE	FL	CE	FL	CE	FL	CE	FL
N	572	0.9757	<b>0.9818</b>	0.9405	<b>0.9580</b>	0.9857	<b>0.9884</b>	0.9488	<b>0.9580</b>	0.9447	<b>0.9580</b>
NE	285	0.9795	<b>0.9848</b>	0.9087	<b>0.9123</b>	0.9882	<b>0.9936</b>	0.9056	<b>0.9454</b>	0.9071	<b>0.9285</b>
NW	1,552	0.9698	<b>0.9870</b>	0.9793	<b>0.9942</b>	0.9559	<b>0.9768</b>	0.9700	<b>0.9841</b>	0.9747	<b>0.9891</b>
EX	265	0.9705	<b>0.9855</b>	0.8339	<b>0.9170</b>	0.9859	<b>0.9932</b>	0.8701	<b>0.9382</b>	0.8516	<b>0.9275</b>

**Table 6**

Testing accuracy comparison of cross entropy and focal loss in 1D-ResNet.

Region	Accuracy of CE	Accuracy of FL
N	0.9406	<b>0.9580</b>
NE	0.9088	<b>0.9123</b>
NW	0.9794	<b>0.9942</b>
EX	0.8340	<b>0.9170</b>
Overall	0.9491	<b>0.9701</b>

Testing dataset	SVM	GA-SVM	ANN	CNN	1D-ResNet
2,674	<b>0.8637</b>	<b>0.9068</b>	<b>0.8534</b>	<b>0.9316</b>	<b>0.9701</b>

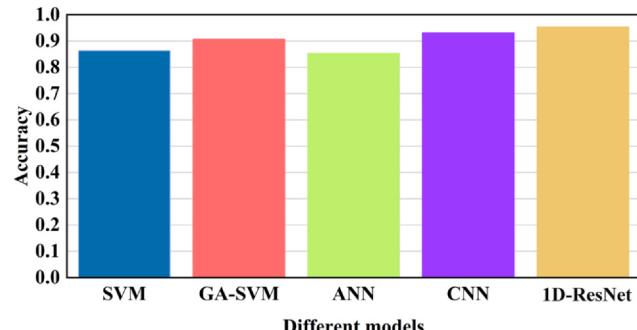


Fig. 11. Prediction accuracy comparison of five models.

external tobacco samples was different. The balance and suppression factors were added to the loss function to solve the issue of imbalanced tobacco samples. The optimized focal loss function can adjust the weight distribution of internal and external tobacco samples and alleviate the imbalanced data of tobacco leaves. In both the convolutional and max-pooling layers, PReLU function was used to add linear factors to the negative input, which can adaptively learn parameters in the network. In the network, BN operation was used to speed up the network training and improve the generalization ability of the network. In addition, the exponential decay learning rate was applied to the control of learning speed. As a deep learning optimization method, dropout was used to avoid overfitting, improve the generalization performance of network, and make the model training converge faster. Compared with the current mainstream recognition methods, such as SVM, GA-SVM, ANN, and CNN, 1D-ResNet exhibited a good adaptability in classification of tobacco cultivation regions. The experimental results showed that the model can accurately recognize different cultivation regions of tobacco leaves. In conclusion, the proposed solution was shown through experimental results to have better overall performance than the four mainstream learning models that have been used for tobacco classification.

In future, the spectra of each chemical components will be analyzed, and the relationships among different components in the classification of tobacco cultivation regions will be explored. The related deep learning-based spectrum analysis methods will be extended to various physical and chemical applications.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgment

This research was funded by the National Natural Science Foundation of China under Grants 61803061, 61906026, and

51705056; the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN201800603); the Chongqing Natural Science Foundation Grant csc2018j-cyjAX0167; the Common Key Technology Innovation Special of Key Industries of Chongqing Science and Technology Commission under Grant Nos. csc2017z-dcy-zdyfx0067, csc2017zdcy-zdyfx0055, and csc2018jszx-cydzd0634; the Artificial Intelligence Technology Innovation Significant Theme Special Project of Chongqing Science and Technology Commission under Grant No. csc2017rgzn-zdyfx0014 and No. csc2017rgzn-zdyfx0035.

#### References

- [1] W. Wu, X.-P. Tang, C. Yang, H.-B. Liu, N.-J. Guo, Investigation of ecological factors controlling quality of flue-cured tobacco (*nicotiana tabacum L.*) using classification methods, *Ecol. Informatics* 16 (2013) 53–61.
- [2] Y. Zhang, Q. Cong, Y. Xie, B. Zhao, et al., Quantitative analysis of routine chemical constituents in tobacco by near-infrared spectroscopy and support vector machine, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 71 (4) (2008) 1408–1413.
- [3] S. Mendell, E.C. Bourlas, M.Z. DeBardeleben, Factors influencing tobacco leaf quality: an investigation of the literature, *Beitrage zur Tabakforschung International/Contributions to Tobacco Research* 12 (3) (1984) 153–167.
- [4] F. White, R. Pandeya, V. Dirks, Correlation studies among and between agronomic, chemical, physical and smoke characteristics in flue-cured tobacco (*nicotiana tabacum L.*), *Can. J. Plant Sci.* 59 (1) (1979) 111–120.
- [5] G. Xiang, H. Yang, L. Yang, X. Zhang, Q. Cao, M. Miao, Multivariate statistical analysis of tobacco of different origin, grade and variety according to polyphenols and organic acids, *Microchem. J.* 95 (2) (2010) 198–206.
- [6] N. Karaivazoglou, N. Tsotsolis, C. Tsadilas, Influence of liming and form of nitrogen fertilizer on nutrient uptake, growth, yield, and quality of virginia (flue-cured) tobacco, *Field Crops Res.* 100 (1) (2007) 52–60.
- [7] C. Zhu, H. Gong, Z. Li, C. Yu, Application of high dimensional feature grouping method in near-infrared spectra of identification of tobacco growing areas, 2016 3rd International Conference on Information Science and Control Engineering (ICISCE), IEEE, 2016, pp. 230–234.
- [8] Y. Bi, S. Li, L. Zhang, Y. Li, W. He, J. Tie, F. Liao, X. Hao, Y. Tian, L. Tang, et al., Quality evaluation of flue-cured tobacco by near infrared spectroscopy and spectral similarity method, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 215 (2019) 398–404.
- [9] J. Omar, B. Slowikowski, A. Boix, Chemometric approach for discriminating tobacco trademarks by near infrared spectroscopy, *Forensic Sci. Int.* 294 (2019) 15–20.
- [10] F.L. Soares, M.C. Marcelo, L.M. Porte, O.F. Pontes, S. Kaiser, Inline simultaneous quantitation of tobacco chemical composition by infrared hyperspectral image associated with chemometrics, *Microchem. J.* 151 (2019) 104225.
- [11] Y. Qin, H. Gong, Nir models for predicting total sugar in tobacco for samples with different physical states, *Infrared Phys. Technol.* 77 (2016) 239–243.
- [12] M.C. Marcelo, F.L. Soares, J.A. Ardila, J.C. Dias, R. Pedó, S. Kaiser, O.F. Pontes, C.E. Pulcinelli, G.P. Sabin, Fast inline tobacco classification by near-infrared hyperspectral imaging and support vector machine-discriminant analysis, *Anal. Meth.* 11 (14) (2019) 1966–1975.
- [13] Y. Zhang, H. Liyuan, Y. Yingze, Identification of producing area of tobacco leaf based on spectrometric analysis and ls-svm, in: International Conference on Computer Technology and Development, 3rd (ICCTD 2011), ASME Press, 2011.
- [14] H. Li, J. Shen, Y. Kong, Z. Cheng, Screening the effective spectrum features of tobacco leaf based on ga and svm, 2016 International Conference on Sensor Network and Computer Engineering, Atlantis Press, 2016.
- [15] Y. Wang, X. Ma, et al., Tobacco quality analysis of industrial classification of different producing area using near-infrared (nir) spectrum, *Spectrosc. Spectral Anal.* 32 (10) (2012) 2694–2697.
- [16] J. Bin, F.-F. Ai, W. Fan, J.-H. Zhou, Y.-H. Yun, Y.-Z. Liang, A modified random forest approach to improve multi-class classification performance of tobacco leaf grades coupled with nir spectroscopy, *RSC Adv.* 6 (36) (2016) 30353–30361.
- [17] L.-J. Ni, L.-G. Zhang, J. Xie, J.-Q. Luo, Pattern recognition of chinese flue-cured tobaccos by an improved and simplified k-nearest neighbors classification algorithm on near infrared spectra, *Analytica chimica acta* 633 (1) (2009) 43–50.
- [18] D. Wang, L. Xie, S. Yang, F. Tian, Support vector machine optimized by genetic algorithm for data analysis of near-infrared spectroscopy sensors, *Sensors* 18 (10) (2018) 3222.
- [19] J. Duan, Y. Huang, Z. Li, B. Zheng, Q. Li, Y. Xiong, L. Wu, S. Min, Determination of 27 chemical constituents in chinese southwest tobacco by ft-nir spectroscopy, *Ind. Crops Prod.* 40 (2012) 21–26.
- [20] O.O. Olarewaju, L.S. Magwaza, H. Nieuwoudt, C. Pobleteecheverria, O.A. Fawole, S.Z. Tesfay, U.L. Opara, Model development for non-destructive determination of rind biochemical properties of 'marsh' grapefruit using visible to near-infrared spectroscopy and chemometrics, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 209 (2019) 62–69.
- [21] Y. Cai, K. Guan, J. Peng, S. Wang, C. Seifert, B. Wardlow, Z. Li, A high-performance and in-season classification system of field-level crop types using time-series landsat data and a machine learning approach, *Remote Sens. Environ.* 210 (2018) 35–47.
- [22] P. Nie, J. Zhang, X. Feng, C. Yu, Y. He, Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning, *Sensors*

- Actuators B: Chem. (2019) 126630.
- [23] S. Zhu, L. Zhou, C. Zhang, Y. Bao, B. Wu, H. Chu, Y. Yu, Y. He, L. Feng, Identification of soybean varieties using hyperspectral imaging coupled with convolutional neural network, Sensors 19 (19) (2019) 4065.
- [24] S. Zhu, L. Zhou, P. Gao, Y. Bao, Y. He, L. Feng, Near-infrared hyperspectral imaging combined with deep learning to identify cotton seed varieties, Molecules 24 (18) (2019) 3268.
- [25] M. Hana, W. McClure, T. Whitaker, M. White, D. Bahler, Applying artificial neural networks: Part ii. using near infrared data to classify tobacco types and identify native grown tobacco, J. Near Infrared Spectrosc. 5 (1) (1997) 19–25.
- [26] D. Wang, F. Tian, S.X. Yang, Z. Zhu, D. Jiang, B. Cai, Improved deep cnn with parameter initialization for data analysis of near-infrared spectroscopy sensors, Sensors 20 (3) (2020) 874.
- [27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [28] H. Gao, Y. Yang, D. Yao, C. Li, Hyperspectral image classification with pre-activation residual attention network, IEEE Access 7 (2019) 176587–176599.
- [29] M.E. Paoletti, J.M. Haut, R. Fernandez-Beltran, J. Plaza, A.J. Plaza, F. Pla, Deep pyramidal residual networks for spectral–spatial hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 57 (2) (2018) 740–754.
- [30] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
- [31] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1492–1500.
- [32] T. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal loss for dense object detection, 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2999–3007.
- [33] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167.
- [34] S. Santurkar, D. Tsipras, A. Ilyas, A. Madry, How does batch normalization help optimization? Advances in Neural Information Processing Systems, 2018, pp. 2483–2493.
- [35] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, in, Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1026–1034.
- [36] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in, Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 807–814.
- [37] T. Bian, F. Chen, L. Xu, Self-attention based speaker recognition using cluster-range loss, Neurocomputing 368 (2019) 59–68.
- [38] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-first AAAI conference on artificial intelligence, 2017, pp. 4278–4284.
- [39] D. Wang, F. Tian, S.X. Yang, Z. Zhu, D. Jiang, B. Cai, Improved deep cnn with parameter initialization for data analysis of near-infrared spectroscopy sensors, Sensors 20 (3) (2020) 874.
- [40] Y. Tang, Deep learning using linear support vector machines, arXiv preprint arXiv:1306.0239.
- [41] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, European Conference on Computer Vision, Springer, 2016, pp. 630–645.
- [42] B. Xu, N. Wang, T. Chen, M. Li, Empirical evaluation of rectified activations in convolutional network, arXiv preprint arXiv:1505.00853.
- [43] L. Zheng, S. Duffner, K. Idrissi, C. Garcia, A. Baskurt, Pairwise identity verification via linear concentrative metric learning, IEEE Trans. Cybernetics 48 (1) (2016) 324–335.
- [44] G. Qi, Z. Zhu, K. Erqinhu, Y. Chen, Y. Chai, J. Sun, Fault-diagnosis for reciprocating compressors using big data and machine learning, Simul. Model. Pract. Theory 80 (2018) 104–127.
- [45] G. Qi, W.-T. Tsai, W. Li, Z. Zhu, Y. Luo, A cloud-based triage log analysis and recovery framework, Simul. Model. Pract. Theory 77 (2017) 292–316.
- [46] Z. Zhu, G. Qi, M. Zheng, J. Sun, Y. Chai, Blockchain based consensus checking in decentralized cloud storage, Simul. Model. Pract. Theory (2019).
- [47] W. Tsai, G. Qi, Dicb: Dynamic intelligent customizable benign pricing strategy for cloud computing, 2012 IEEE Fifth International Conference on Cloud Computing, 2012, pp. 654–661.
- [48] W. Tsai, G. Qi, Y. Chen, A cost-effective intelligent configuration model in cloud computing, 2012 32nd International Conference on Distributed Computing Systems Workshops, 2012, pp. 400–408.
- [49] W. Tsai, G. Qi, Y. Chen, Choosing cost-effective configuration in cloud storage, 2013 IEEE Eleventh International Symposium on Autonomous Decentralized Systems (ISADS), 2013, pp. 1–8.
- [50] Wikipedia, Guizhou Province, <http://en.wikipedia.org/wiki/Guizhou> (2020 (accessed Mar. 30, 2020)).
- [51] F. Tian, Z. Zhu, W. Pan, Automatic prediction of leave chemical compositions based on nir spectroscopy with machine learning, Int. J. Robot. Automat. 34 (4).
- [52] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011, pp. 315–323.
- [53] D.-A. Clevert, T. Unterthiner, S. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), arXiv preprint arXiv:1511.07289.