**Classification of Sugar Beet Based on Hyperspectrum and Extreme Learning Machine**

**Abstract** In this paper, the classification of sugar beet varieties based on hyperspectral technology and Extreme Learning Machine (ELM) algorithm was realized. And the influences of seven kinds of pretreatment methods, namely Savitzky-Golay smoothing (SG), first derivative (FD), Savitzky-Golay smoothing combined with first derivative (SG-FD), logarithmic transformation (LT), logarithmic transformation combined with first derivative (LT-FD), standard normal variate (SNV), standard normal variate combined with first derivative (SNV-FD), on the recognition performances of ELM model were analyzed. In order to simplify the input variables, the standard deviation peak method was used to extract the feature bands under different preprocessed spectral data. Experimental results show that under different pretreatment methods, the recognition rate of ELM models to sugar beet varieties is over 80% with good learning performance and generalization performance, and it is found that the combination of different pretreatment methods and FD can effectively improve the signal-to-noise ratio and enhance the accuracy and stability of spectral models; as a whole, the recognition accuracy of the ELM models established by the feature bands is better than that established by the whole bands, showing that the feature bands extracted by the standard deviation peak method are effective; based on the pretreatment of SG-FD, the ELM models established by the full bands and feature bands both achieve the best recognition effect, showing that the recognition rate of the training sets both reaches 98.13% and the recognition rate of the prediction sets is 93.94% and 95.45% respectively.

**Key words** hyperspectral; ELM; sugar beet variety

**1. Introduction**

Hyperspectrum detection is a rapid, non-destructive and comprehensive detection technology emerging in the 1980s, which possesses the characteristics of high spectral resolution, multiple and continuous bands and overcomes the limitation of wide band and low resolution of multi spectral technology. Hyperspectral technology can reflect subtle spectral difference between adjacent bands of target object and provide an effective technical support for the research of the object slight features (Pu and Gong, 2000).

In recent years, a lot of remarkable achievements have been achieved in the study of plant species identification based on hyperspectral technology combined with machine learning algorithm. Zhang et al. (2014) used hyperspectral technology to identify black bean varieties and found that the extreme learning machine (ELM) model based on spectral feature information extracted from wavelet transform possessed the best recognition effect, and the correct recognition rate of its modeling set and prediction set both reached 100%. Ma et al. (2017) realized the rapid identification of apple varieties based on hyperspectral technology and found that the recognition accuracy of the K nearest neighbor (KNN) model reached 100%. Huang et al. (2016) used hyperspectral technology to identify maize seeds of different years and optimized the least squares support vector machine (LSSVM) using incremental support vector data description (ISVDD) algorithm, finding that its recognition accuracy reached 94.4%. In addition, some scholars have identified the other plant varieties based on hyperspectral technology, such as watermelon seeds, Chinese cabbage seeds, red beans, rice, litchi and so on (Zhang et al., 2013; Cheng et al., 2014; Sun et al., 2016; Wang et al., 2014; Liu et al., 2014). However, the research on recognition of sugar beet varieties based on hyperspectral technology combined with machine learning algorithm has not been reported.

Sugar beet is one of the main raw materials for sugar production in China, which plays an important role in the economic income of Heilongjiang, Inner Mongolia and Xinjiang. The variety of sugar beet is closely related to the sugar beet cultivation and sugar industry. And proper selection and reasonable collocation planting of sugar beet varieties can effectively improve the yield and sugar content. Therefore, the rapid identification of sugar beet varieties plays an important guiding role in the selection of sugar beet varieties and production management in the field. This paper proposes a rapid identification model for three sugar beet varieties based on hyperspectral technology and extreme learning machine (ELM) algorithm, so as to provide theoretical basis for sugar beet production and management.

**2. Materials and methods**

2.1 Experimental materials

The experiment was carried out in the way of field plot, and the selected varieties of sugar beet were SD21816, KWS1676 and KWS9147 (SD21816 was planted in sugar beet cultivation base of Tuzuo Banner, Hohhot, Inner Mongolia with a total of 9 plots in 2013; KWS1676 was planted in sugar beet plantation area of Taiping Town, Songshan District, Chifeng City, Inner Mongolia with a total of 28 plots in 2014; KWS9147 was planted in the experimental field of Inner Mongolia Agricultural University with a total of 18 plots in 2015). During the experiment, the sugar beet canopy spectra were collected in the field from mid-late June to mid-late September, and a total of 400 samples were collected from 3 sugar beet varieties. In this paper, the samples were randomly divided into training set (268) and prediction set (132) according to the proportion of 2:1, and the sugar beet varieties SD21816, KWS1676 and KWS9147 were assigned to 1, 2 and 3 respectively using assignment method (table 1).

**Table 1 Classification of sugar beet samples**

|  |  |  |  |
| --- | --- | --- | --- |
| varieties | assignment | training set | prediction set |
| SD21816 | 1 | 72 | 35 |
| KWS1676 | 2 | 112 | 56 |
| KWS9147 | 3 | 84 | 41 |

2.2 Canopy spectral data acquisition

In this experiment, the sugar beet canopy spectra were collected by the field portable ASD Qualityspec hyperspectral spectrometer manufactured by American ASD Company. The wavelength range of the spectrometer is 350-1830 nm, of which the sampling interval of 350-1000 nm is 1.4 nm and the spectral resolution is 3 nm; the spectral sampling interval of 1000-1830 nm is 2 nm and the spectral resolution is 10 nm. The experiment was carried out at Beijing time 10:00-14:00 in the sunny, windless and cloudless weather. When collecting spectra, the fiber probe is vertically downward, and the vertical height from sugar beet canopy is determined according to the sample canopy size and the angle (25 °) of probe view, so that the sample canopy can be located within the field of probe view. A representative sample point was selected in each plot and its canopy spectra were collected for four times continuously, using average value as the sample spectra. And reference whiteboard calibration was performed before collecting canopy spectra in each plot.

2.3 Spectral data pretreatment

In addition to the information related to the sample itself, the spectral data collected by the spectrometer also contains noise signals, such as stray light and soil background. These noise signals can affect the authenticity of spectral information, thus influencing the accuracy and stability of model. Therefore, it is necessary to preprocess the spectral data. However, not every pretreatment method can achieve the desired effect.

This study utilized Savitzky-Golay smoothing (SG), first derivative (FD), Savitzky-Golay smoothing combined with first derivative (SG-FD), logarithmic transformation(LT), logarithmic transformation combined with first derivative (LT-FD), standard normal variate (SNV) and standard normal variate combined with first derivative (SNV-FD) to preprocess hyperspectral data, and the effects of the 7 kinds of pretreatment methods were compared and analyzed.

2.4 Feature band selection

Hyperspectral data possesses the characteristics of large number of bands, large amount of data and strong correlation between adjacent bands, which affects the processing speed of data, brings difficulty to research, and also limits the wide application of hyperspectral technology to a certain extent (Liu et al., 2005; Wang et al., 2006). Therefore, it is necessary to reduce the dimension of hyperspectral data under the premise of ensuring the model accuracy. Researchers usually use the methods of spectral feature extraction and feature band selection to reduce the dimension of hyperspectral data (Qiu et al., 2013; Hsu, 2007; Du and He, 2008).

This paper utilized the standard deviation peak method to select the feature bands of sugar beet canopy spectra. Its basic idea is that the standard deviation of spectral data of all samples in a certain band is greater, which indicates that the spectral difference between different sugar beet varieties is more significant under this band, and the ability of this band to distinguish beet varieties is stronger. The peak value on the standard deviation curve is the maximum of the standard deviation between adjacent bands, which contains the largest spectral information of the different sugar beet varieties between adjacent bands.

The screening process of this method is as follows:

(1) Calculating the standard deviation of all sugar beet samples in each band. The calculation formula is as follows:



--standard deviation at the i band; *N*--number of samples; --spectral mean value at the I band

(2) Screening out the peaks on the standard deviation curve and calculating the average value of all the peaks.

(3) Taking the bands corresponding to the peak value above on the standard deviation curve as the feature bands.

2.5 Recognition model

Extreme learning machine (ELM) is a new machine learning algorithm proposed by Prof. Huang Guangbin on the basis of single hidden layer feedforward neural networks (SLFNs), which is composed of input layer, hidden layer and output layer (Huang et al., 2006). This algorithm randomly generates the input weight (w) and the hidden layer bias (b), which is not required to be adjusted during the training. Just setting the activation function g (x) and the number (L) of the hidden layer neurons, the output weight (β) between the hidden layer and the output layer can be uniquely determined so as to obtain the unique optimal solution. ELM algorithm possesses the advantages of fast learning speed and good generalization performance, which overcomes disadvantages of adjusting the network parameters unceasingly during the training process and improves the problems of slow training speed, sinking into local optimization rather than global optimization and sensitivity of learning rate selection in traditional neural networks algorithm (Huang et al., 2015).

In this paper, the sigmoid function was used as the activation function of ELM algorithm, and the number of hidden layer neurons was set to 60.

2.6 Spectral data analysis software

In this paper, the various pretreatment methods of spectral data, the feature bands selection and the establishment of recognition models were completed in MATLAB R2014a software.

**3. Results**

3.1 Original spectral response analysis of sugar beet canopy



**Fig. 1 Canopy original reflectance spectra of sugar beet**

Figure 1 shows the original reflectance spectra of all sugar beet samples. As can be seen, the reflectance spectra curves trend of the three varieties is basically the same with no significant difference, but the reflectivity is different. Therefore, the identification of different sugar beet varieties needs further modeling and analysis. At the same time, it can be seen that there is significant noise at the 1351-1450 nm band and the end of spectra. Considering that the head of the original spectra also contain a large amount of random noise, reflectance spectra between 400-1350 nm were selected for subsequent pretreatment and modeling analysis in this paper (Zhao et al., 2017).

3.2 Identification of sugar beet varieties based on whole bands

The above-mentioned seven preprocessed methods were used to preprocess the original spectral data of 400-1350nm. Then, the ELM identification models were established by using raw data and seven kinds of preprocessed spectral data.

**Table 2 Recognition results of ELM models based on whole bands**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| pretreatment methods | training set | | prediction set | |
| Correct recognition number | Recognition rate /% | Correct recognition number | Recognition rate /% |
| raw | 258 | 96.27 | 117 | 88.64 |
| SG | 259 | 96.64 | 117 | 88.64 |
| FD | 260 | 97.01 | 121 | 91.67 |
| SG-FD | 263 | 98.13 | 124 | 93.94 |
| LT | 250 | 93.28 | 108 | 81.82 |
| LT-FD | 263 | 98.13 | 123 | 93.18 |
| SNV | 244 | 91.04 | 109 | 82.58 |
| SNV-FD | 260 | 97.01 | 122 | 92.42 |

The identification results of sugar beet varieties are shown in Table 2. As can be seen, the recognition rate of ELM models based on the whole band under the different pretreatment methods all reaches more than 80%. Among them, the ELM model that was established using the spectral data preprocessed by SG-FD possesses the best recognition effect, in which the recognition rate of training set is 98.13% and the recognition rate of prediction set is 93.94%. The recognition effect of SG is equivalent to that of the original spectrum, in which the recognition rate of training set is 96.64% and 96.27% respectively and the recognition rate of prediction set is both 88.64%. The recognition effect of LT and SNV was poor, in which the recognition rate of the training set and the prediction set is both lower than that of the original spectra, but the recognition effect of other pretreatment methods is all better than that of the original spectra.

According to the combination of pretreatment methods, the recognition effect of different pretreatment methods (SG, LT and SNV) combined with FD is all better than that of single pretreatment method, which indicates that the combination of different pretreatment methods and FD can effectively improve the signal-to-noise ratio of spectral data and enhance the accuracy and stability of spectral models.

3.3 Results of feature bands selection

The spectral data of 400-1350nm intercepted in this study contains 951 spectral bands. In order to reduce the redundancy of spectral information, this paper used standard deviation peak method to select feature bands of different preprocessed spectral data, and the selection results are shown in Table 3.

**Table 3 The number of feature bands selected based on the standard deviation peak method**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| preprocessed methods | raw | SG | FD | SG-FD | LT | LT-FD | SNV | SNV-FD |
| Number of feature bands | 87 | 35 | 83 | 51 | 29 | 106 | 61 | 98 |

It can be seen from table 3 that the number of feature bands selected under different pretreatment methods is different. And the number of feature bands based on LT is the least with 29 bands; the number of feature bands based on LT-FD is the largest with 106 bands. Compared with the original 951 spectral bands, the spectral information dimension is greatly compressed after the selection of feature bands.

3.4 Identification of sugar beet varieties based on feature bands

The ELM models were established by using the feature bands selected under different pretreatments.

**Table 4 Recognition results of ELM models based on feature bands**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| pretreatment methods | training set | | prediction set | |
| Correct recognition number | Recognition rate /% | Correct recognition number | Recognition rate /% |
| raw | 264 | 98.51 | 126 | 95.45 |
| SG | 263 | 98.13 | 121 | 91.67 |
| FD | 257 | 95.90 | 118 | 89.39 |
| SG-FD | 263 | 98.13 | 126 | 95.45 |
| LT | 259 | 96.64 | 117 | 88.64 |
| LT-FD | 262 | 97.76 | 125 | 94.70 |
| SNV | 260 | 97.01 | 122 | 92.42 |
| SNV-FD | 261 | 97.39 | 124 | 93.94 |

Table 4 shows the recognition results of ELM models based on feature bands. As can be seen, compared with the ELM models established by using the whole bands, in addition to FD, the recognition rate of the ELM models established by using the feature bands selected under the other pretreatment methods is improved in training set and prediction set. Among these models, the ELM model based on SG-FD possessed the best recognition effect, in which the recognition rate of training set is 98.13% and the recognition rate of prediction set is 95.45%. The recognition rate of ELM model based on FD has decreased, but it also achieves the ideal recognition accuracy, in which the recognition rate of the training set is 95.90% and the recognition rate of the prediction set is 89.39%. These results indicate that the feature bands selected by the standard deviation peak method basically cover the effective information of sugar beet varieties, which greatly reduces the number of variables and improves the efficiency of processing data.

According to the combination of pretreatment methods, the recognition effect of different pretreatment methods (SG, LT and SNV) combined with FD is still all better than that of single pretreatment method based on feature bands, which is consistent with the results obtained under the whole bands.

**4. Discussion and conclusions**

This paper utilized hyperspectral technology and ELM algorithms to identify three varieties of sugar beet (SD21816、KWS1676、KWS9147), and the recognition performance of ELM models under different pretreatment methods (SG, FD, SG-FD, LT, LT-FD, SNV, SNV-FD) were compared. The following conclusions are drawn:

(1) Under different pretreatment methods, the ELM models have all achieved the ideal recognition accuracy, and the recognition rate of training set and prediction set is above 80%, which indicates that ELM algorithm shows good learning performance and generalization performance in the identification of sugar beet varieties.

(2) Compared with the ELM models established based on the whole bands, the ELM models established by using feature bands have generally improved the accuracy of identifying sugar beet varieties, indicating that the feature bands extracted by standard deviation peak method basically covers the effective information of sugar beet varieties, which streamlines the spectral variables and improves the efficiency of processing information.

(3) Under different pretreatment methods, the ELM models established by respectively using the whole bands and feature bands based on the pretreatment method of SG-FD have both achieved the optimal recognition effect, in which the recognition rate of training set is both 98.13% and the recognition rate of prediction set is 93.94% and 95.45% respectively.

(4) According to the combination of pretreatment methods, the combination of different pretreatment methods and FD can effectively improve the signal-noise ratio of the spectral data and enhance the accuracy and stability of the spectral models.

This paper shows that it is feasible to identify the beet varieties by using hyperspectral technique. However, the ELM models only for three varieties of Sugar beet were established in this study. In the future research, more robust recognition models should be established for wider sugar beet varieties, so as to provide more powerful theoretical support for Sugar beet field management.

**References**

Cheng S X, Kong W W, Zhang C, et al. Variety recognition of Chinese cabbage seeds by hyperspectral imaging combined with machine learning. Spectroscopy and Spectral Analysis, 2014, 34(9): 2519-2522.

Du Q, He Y. Similarity-Based Unsupervised Band Selection for Hyperspectral Image Analysis. IEEE Geoscience and Remote Sensing Letters, 2008, 5(4):564-568.

Hsu P H. Feature extraction of hyperspectral images using wavelet and matching pursuit. ISPRS Journal of Photogrammetry and Remote Sensing, 2007, 62(2):78-92.

Huang F H, Zhang S J, Yang Y, et al. Application of hyperspectral imaging for detection of defective features in nectarine fruit. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(11): 252-259.

Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications. Neurocomputing, 2006, 70(1/2/3): 489－501.

Huang M, Tang J Y, Yang B, et al. Classification of maize seeds of different years based on hyperspectral imaging and model updating. Computers and Electronics in Agriculture, 2016, 122:139-145.

Liu D, Wang L, Sun D, et al. Lychee variety discrimination by hyperspectral imaging coupled with multivariate classification. Food Analytical Methods, 2014, 7(9): 1848-1857.

Liu X Y, Lin H, Xiong J L, et al. Band selection from hyperspectral data of forestry species. Remote Sensing Information, 2005, (4): 41-44+64.

Ma H L, Wang R L, Cai C, et al. Rapid identification of apple varieties based on hyperspectral imaging. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(4): 305-312.

Pu R L, Gong P. Hyperspectral remote sensing and its applications. Beijing: Higher Education Press, 2000.

Qiu L, Lin H, Zang Z, et al. Hyperspectral characteristic band selection for wetland vegetation based on mean confidence interval. Journal of Central South University of Forestry & Technology, 2013, 33(1): 41-45.

Sun J, Lu X Z, Zhang X D, et al. Identification of red bean variety with probabilistic GA-PNN based on hyperspectral imaging. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(6): 215-221.

Wang L, Liu D, Pu H, et al. Use of hyperspectral imaging to discriminate the variety and quality of rice. Food Analytical Methods, 2014, 8(2): 515-523.

Wang X H, Xiao P, Guo J M. Research on dimensionality reduction technology of hyperspectral data. Bulletin of Soil and Water Conservation, 2006, 26(6): 89-91.

Zhang C, Liu F, Kong W W, et al. Fast identification of watermelon seed variety using near infrared hyperspectral imaging technology. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(20): 270-277.

Zhang C, Liu F, Zhang H L, et al. Identification of varieties of black bean using ground based hyperspectral imaging. Spectroscopy and Spectral Analysis, 2014, 34(3): 746-750.

Zhao X T, Zhang S J, Liu J L, et al. Study on varieties discrimination of nectarine by hyperspectral technology combined with CARS-ELM algorithm. Modern Food Science and Technology, 2017, 33(10):281-287.