Rapid on fast resolution of multi-grain mixed broken particles combined with optimization algorithm and VIS-hyperspectral imaging technology

Yu Lei
School of Mechanical Engineering, Sichuan University of Science and Engineering, Zigong 643000, China
Xinjun Hu (suse2021@126.com)
School of Mechanical Engineering, Sichuan University of Science and Engineering, Zigong 643000, China
Jiahong Zhang
School of Mechanical Engineering, Sichuan University of Science and Engineering, Zigong 643000, China
Songcai Yan
School of Mechanical Engineering, Sichuan University of Science and Engineering, Zigong 643000, China
Qinyuan Xue
Xiaoyan Ma
School of Mechanical Engineering, Sichuan University of Science and Engineering, Zigong 643000, China
Jianping Tian
School of Mechanical Engineering, Sichuan University of Science and Engineering, Zigong 643000, China
Manjiao Chen
School of Mechanical Engineering, Sichuan University of Science and Engineering, Zigong 643000, China
Dan Huang
College of Bioengineering, Sichuan University of Science and Engineering, Zigong 643000, China

Research Article

Keywords: multi-grain broken grain, hyperspectral imaging technology, quickly distinguish, mixture ratio

Posted Date: April 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1503561/v1

Abstract

In this study, the optimization algorithm was combined with the visible light hyperspectral imaging technology to distinguish the multi-grain broken grain varieties and determine their mixing ratio. The Otsu method combined with the watershed algorithm was used to extract the ROI spectral reflectance of the sample particles, DBSCAN-MD was used to remove outliers, IVSO-CARS was used to extract the characteristic wavelength, establish four identification models (PNN, GRNN, BPNN, RBFNN) based on full wavelength and characteristic wavelength. Then input three sets of mixed samples from the outside to verify the effect of the BPNN model trained by 7 different kernel functions. The results show that the BPNN model trained with the One Step Secant function has the best effect, and the average recognition accuracy of the three groups of samples is 99%. Therefore, this method can realize the rapid identification and research of multigrain and multigrain, and determine the mixing ratio.

1. Introduction

China has a long history of liquor production, rich varieties of famous liquors, multi-grain strong-flavor liquor is one of them. The raw materials used in its production are sorghum 36%, rice 22%, glutinous rice 18%, wheat 16%, corn 8%, and then brewed by traditional techniques. In order to improve the quality of the liquor, a variety of food raw materials are usually crushed and then evenly mixed together, the uniformity of not less than 90%. The purpose of crushing grains is to expose more starch and increase the contact surface between grain grains and koji powder, which is conducive to saccharification fermentation(Zheng et al., 2018). The crushed particles can't be too fine, to pass 20 mesh sieve fine powder is not more than 20% shall prevail. If it is too fine, it will lead to easy pressure when distilling, sticky liquor grains and lumps, which will increase the amount of filler and reduce the quality of finished products. On the contrary, it is easy to lead to the inpenetration of cooking paste and incomplete function of koji powder, which will leave many available starch residue, resulting in low liquor yield(Qian et al., 2021). Because of the different sizes and shapes of grain particles after crushing, it is difficult to distinguish each variety by naked eye, and it is impossible to judge whether the grains of multiple varieties are evenly mixed. Therefore, finding a simple, rapid, efficient and accurate method to distinguish the varieties of grain grains is the key to ensure the quality of liquor.

Hyperspectral imaging technology is a popular technology in the field of non-destructive testing and classification in recent years, and has achieved good results in many studies. For example, use of hyperspectral imaging for rapid detection of sorghum adulteration and classification of cake shelf life(Bai et al., 2020; Sricharoonratana et al., 2021). In addition to this, there are also spectral techniques combined with optimization algorithms for classification. For example, combining hyperspectral imaging techniques with combinatorial optimization strategies to achieve classification and identification of food residues(Chen et al., 2021). Combining deep convolutional neural networks with hyperspectral techniques to classify food residues, both have achieved good results(B. Liu et al., 2021). Although HIS has achieved satisfactory results in the above studies, it has not yet been applied to particle classification, detection and detection. visualization.

Therefore, the main purpose of this study is to combine hyperspectral imaging technology, image processing and optimization algorithm to distinguish the mixed grains, and provide a method reference for the high-quality production of multi-grain strong-flavor liquor. The specific steps of this study are as follows :(1) prepare samples of five kinds of grain particles, and use hyperspectral data acquisition system to collect spectral data of five kinds of grain particles in the range of visible light (397-1004.5mm); (2) Otsu method was combined with watershed

algorithm to find the image segmentation algorithm suitable for different varieties, different sizes and different shapes of grain particles, and extract the spectral data of each grain particles; (3) The method of combining density clustering (DBSCAN) and mahalanobis distance (MD) was used to remove the abnormal particle samples of each grain broken particle; (4) Combined iterative variable subset optimization algorithm (IVSO) with adaptive reweighting algorithm (CARS) to select characteristic wavelengths; (5) The back propagation neural network model (BPNN) was established based on the optimization of different kernel functions of full wavelength and characteristic wavelength, and the network was trained with seven different kernel functions to determine the optimal recognition model, so as to realize the visualization of single particle samples and mixed particle samples and detect their mixing ratio.

2. Materials And Methods

2.1. Sample preparation

The five food crops selected for this study are all from Sichuan, namely red glutinous sorghum from Luzhou, long-grain sorghum rice from Nanchong, long-round glutinous rice from Zigong, Xikemai No. 3 wheat from Mianyang and home-grown corn from Guang'an. To selected 0.5kg grains with complete grains and no insect erosion to place in a blender, and crushed for 5s one times. Each variety was weighed and crushed separately, and then the crushed grain particles were placed in a 20-mesh sieve, the final standard is not more than 20% fine powder through 20-mesh sieve. The sieved fine powder is set aside and the remaining coarse particles are reserved for sample preparation. Each variety prepared for 23 samples, and each sample contained 100 grains of different sizes and shapes, totaling 11500 samples. Three mixed samples were prepared according to the proportion of sorghum 36%, rice 22%, glutinous rice 18%, wheat 16% and 8%. In addition, a total of 1500 samples with 300 grains in each of five varieties were selected as the external input validation set of the model to test the model generalization effect.

2.2 Hyperspectral imaging system and data acquisition

All sample data for this study were collected by the visible hyperspectral imaging system (FX10E, Specim, Finland), which includes: electronic control platform, two sets of 150W halogen lamps (OSRAM, Germany), computer equipped with special LUMO-scanner software (DELL, USA) and auxiliary stand. The FX10E hyperspectral camera with a field of view is 38 degrees, a 12-bit camera output, a spatial resolution of 1 024 × 628 pixels, and a spectral range of 397 ~ 1004.5nm, generating 448 bands in total. Before data collection, the parameters of the system were set as follows: peak lighting 3616, exposure frequency 50HZ, exposure time 8ms, platform movement speed 10.84ms. After parameters are set, the system starts to collect data. The samples were tiled in the order of numbering in a petri dish with a diameter of 90 mm and a height of 10 mm to ensure that the broken particles did not stick to and overlap with each other, and then placed on the mobile platform for data collection.

In the process of sample data collection, the intensity distribution of light source with different wavelength is not uniform, which will affect the collected image signal. Secondly, the dark current inside the camera will generate noise in the process of data collection and affect the accuracy of data. Therefore, black-white correction should be carried out on the collected data, and the correction formula is shown in equation:

$$D = \frac{DO - D\mathrm{d}}{Dw - D\mathrm{d}}$$

1

Where D is corrected spectral image, Do is covered the dark reference image collected by the lens, Dd is original hyperspectral image, Dw is collected standard whiteboard image.

2.3 Image segmentation and spectral data extraction

The spectral data collected by the hyperspectral imaging system contains the information of grain broken particles and background information. In order to effectively remove the background information and retain the information of grain broken particles, the Otsu method combined with watershed algorithm are used to segment images of grain broken particles with different sizes and shapes in the samples(Xi et al., 2021). The specific steps are :(1) After black and white correction, remove image noise; (2) Image gray processing, using Otsu method to find the optimal threshold of different varieties of samples, to determine the connected domain(Nandhini & Porkodi, 2021); (3) Adjust the filter size to perform edge erosion and watershed transformation on the binarized sample image; (4) Extracting effective region of Interest (ROI) spectral data values; The formula for calculating the success rate of spectral data extraction is shown in Equation:

$$ACU\text{ex}(\%) = \frac{E}{Tr} \times 100$$

2

Where ACUex is the success rate of data extraction, E is the number of samples extracted, Tr is the total number of samples;

2.4 Eliminate outliers

After extracting valid spectral data, it is necessary to clean the data to remove outliers that affect the modeling results. Here, a combination of density clustering (DBSCAN) and Mahalanobis distance (MD) is used to remove outliers. Density clustering can divide regions with high enough density into clusters, and can find clusters of arbitrary shapes in noisy spatial datasets, and then remove noise points that affect the modeling results(Alireza & Negin, 2021). Mahalanobis distance considers points to be anisotropic, and the specific parameters of anisotropy can be represented by a covariance matrix. The covariance matrix can be regarded as a multidimensional normal distribution covariance matrix, then the contour line describing the distribution density function is a common ellipse. The Mahalanobis distance from the center of the ellipse to each point on the ellipse is equal. If it is not satisfied, it will be removed as an abnormal point. This method can effectively remove points with abnormal distribution(Jiayou et al., 2021).

2.5 Preprocessing and feature wavelength screening

In order to effectively eliminate the spectral differences caused by different scattering levels, a region of interest (MSC) is used to preprocess the spectral data to enhance the correlation between the spectra and the data(Haoping, Xinjun, & Jianping, 2021). The data is divided into training set and test set according to the ratio of 4:1 using the random partition method. Then 10 bands before and after are removed to improve the signal-to-noise ratio.

The data collected by the hyperspectral imaging system contains a lot of complicated and redundant information, and the images of adjacent bands are highly similar. In order to select representative band data, it is necessary to

dimension the spectral data. Therefore, this study combines Iteratively Variable Subset Optimization (IVSO) and Competitive adaptive re-weighting algorithm (CARS) to select characteristic wavelengths. IVSO is a highly stable variable selection method, which can eliminate variables with less information and extract effective variable information(Wang et al., 2015). CARS can remove the bands with smaller weights and extract important variables related to the detection target. IVSO-CARS extracts a total of 41 characteristic wavelengths, effectively removing redundant information(Bonah et al., 2020), saving modeling time and improving modeling efficiency.

2.6 Building a classification model

In order to select a model suitable for the classification of multi-grain mixed and broken particles, the probabilistic neural network model (Probabilistic Neural Networks, PNN), the generalized regression neural network model (Generalized Regression Neural Network, GRNN), the radial basis neural network model (Radial Basis Function Neural Network, RBFNN) and Back Propagation Neural Network (BPNN) four classification models were compared and analyzed. PNN is a forward-propagating network and does not require back-propagation to optimize parameters. This is because PNN combines the Bayesian minimum risk criterion to determine the sample category. It has the characteristics of fast training and less time consumption(Yin et al., 2021). GRNN is an artificial neural network model based on nonlinear regression theory, which has strong nonlinear mapping ability and learning speed(Hou et al., 2020). RBF is a feedforward neural network with excellent performance, which has global approximation ability, fast convergence speed and good classification ability(Shi et al., 2018). BPNN is a feed-forward neural network model with tutor learning, which has a simple structure and can quickly search for the optimal solution(Qiao et al., 2021). The model is comprehensively evaluated through the training set, test set, external input validation set and running time. The formula for calculating the correct rate is as follows:

$$precision(\%) = \frac{Tp}{Tp+Fp} \times 100 \quad (3)$$

where precision is the classification accuracy, Tp is the number of correct predictions, Fp is the number of prediction errors.

3. Results And Discussion

3.1 Analysis of image segmentation effect and outlier elimination

The combination of the Otsu method and the watershed algorithm to segment the sample images has achieved good results, and the extraction effect is shown in Table 1. According to the extraction results, the best extraction effect in the training set is wheat, with an extraction success rate of 99.45%, and the average extraction success rate of the five varieties is 98.25%; the overall water in the validation set is relatively high, the average extraction success rate is 99.32%.

Principal component analysis (PCA) was used to analyze the spectral data of DBSCAN-MD after removing outliers, taking sorghum and wheat as examples to introduce in detail(Xinna et al., 2021). The black points in Figure.1.(a) and(c) represent outliers that were removed, and the red and purple points represent normal values. From the three-dimensional principal component score map, it can be seen that sorghum PC-1 (60.28%), PC-2 (37.26%) and PC-3 (1.62%), explained a variance of 99.56%, wheat PC-1 (77.02%), PC-3 -2 (22.03%) and PC-3

(0.62%), explained a variance of 99.67%. Due to the different sizes and shapes of broken grains, there are many outliers and noise points, and the distribution of three-dimensional principal component data points of each grain variety is also different. Compared with the spectral curves without outliers (Figures.1. a and b), the spectral curves after outlier removal (Figures.1.e and f) are more compact, and the correlation between the curves is enhanced. Therefore, the DBSCAN-MD method can effectively eliminate abnormal particles that affect the modeling results.

The overall analysis was carried out on the spectral data of five kinds of grain broken grains after removing outliers is shown in Fig. 2. A total of 10808 sample data were extracted, and 9631 sample data remained after DBSCAN-MD processing. Five kinds of grain granules PC-1 (62.9%), PC-2 (31.74%) and PC-3 (4.42%), explained the variance of 99.06%. It can be seen from the figure that there is overlap between different varieties of grain granules, mainly because they contain the same components such as starch, protein, tannin, fiber, etc.(Cabrera-Ramirez et al., 2020; Lyu et al., 2021; Wenting et al., 2021; Zhao et al., 2022). After removing outliers, the scatterplot of broken particles is very close and the correlation is enhanced. Therefore, in order to better distinguish different varieties of broken particles, it is necessary to select an algorithm with better classification performance, conduct comparative analysis, and determine the optimal algorithm for modeling.

3.2 Model performance evaluation

The 9631 groups of full bands after removing outliers and the characteristic bands selected by the IVSO-CARS method were substituted into the PNN, GRNN, BPNN and RBFNN models respectively. Among them, there were 7700 groups in the training set, 1931 groups in the test set, and 1490 groups in the validation set. The identification results are shown in Table 2. Ten-fold cross-validation is used in the GRNN model to find the optimal spread value for model training(Qiao et al., 2020). From the results of the full band and the characteristic band, both the training set and the test set have achieved good results, but there is over-fitting in the validation set of external input, and the validation effect is extremely poor, which is caused by the poor generalization performance of the model(Warey et al., 2021). In GRNN and RBFNN, the classification and recognition effect of the model based on the full wavelength is much better than that of the model based on the characteristic wavelength. This is because the broken grains have different sizes, different shapes and no regularity. The characteristic wavelength is representative, but it can only reflect part of the characteristics of broken particles, and cannot comprehensively summarize the overall situation of broken particles. From the modeling results, the BPNN model has the best effect. Both the model established by the full wavelength and the model established by the characteristic wavelength have achieved good results. However, because the modeling results of the two wavelength datasets are close (the difference in accuracy is 1%) within), so it is impossible to determine which effect is better.

In order to explore the classification effect of BPNN, find the difference between modeling based on full wavelength and characteristic wavelength, and improve the generalization performance of the model, different kernel functions are selected to train BPNN. The results are shown in Table 3. The BPNN network layer is set to two layers, the maximum number of iterations of the model is set to 500, and the accuracy is 0.1(Bo et al., 2021). From the results in Table 3, the classification and recognition results of the full-wavelength models trained by Resilient training function, Scaled conjugate gradient training function, One Step Secant training function; Fletcher-Powell connection gradient training function and Polak-Ribiere connection gradient training function are all within 99%. Above, the Resilient training function takes the shortest training time, only 32s. Full-wavelength and eigen wavelength models trained with Scaled conjugate gradient training function, Fletcher-Powell connection gradient training function, Gradient descent training function for adaptive learning rate, Gradient descent training

function with momentum for adaptive learning rate, and Polak-Ribiere connection gradient training function The effect is obviously different, and the classification and recognition accuracy differs by more than 10%. However, the results of the Resilient training function and the One Step Secant training function; the two models trained are still close and need further verification.

In order to further verify the results of the Resilient training function and the One Step Secant training function training model, find the BPNN trained by the best function, three groups of multi-grain mixed broken grain samples composed of different broken grains are used for verification. The mixing ratio is: sorghum 36%, rice 22%, glutinous rice 18%, wheat 16% and corn 8%, and the total number of kernels is 100. The verification results are shown in Table 4. In general, the classification and recognition effect of the model based on the full wavelength is better than that of the model based on the characteristic wavelength. From the verification results, the results of the BPNN trained by the One Step Secant training function are better. The average classification and recognition accuracy of the characteristic wavelength models of the three groups of samples is 96%, which is 3% higher than that of RPROP-BPNN; the average of the full-wavelength models of the three groups of samples The classification recognition accuracy is 99%, which is 4% higher than RPROP-BPNN. From this, it can be determined that the full-wavelength BPNN model trained with the One Step Secant training function has the best effect.

3.3 Visual verification

The trained model was visually verified with single-variety broken grain samples and multi-grain mixed broken grain samples. The visual colors were defined as: sorghum red, rice green, glutinous rice blue, wheat purple and corn yellow. The verification results of single-variety broken particles are shown in Supplementary figure. 1. I. Except for one piece of rice that was misjudged as glutinous rice, the verification results of broken grains of other varieties were all correct. Compared with glutinous rice, the broken grain samples of other varieties are easily misjudged because the carbohydrate (mainly starch) content of glutinous rice and rice is not much different, both of which are about 75%. Secondly, the distribution of starch in rice granules is uneven, and some broken granules have high starch content after breaking them, so it is easy to be misjudged as glutinous rice. The visual verification results of multi-grain mixed broken particles are shown in Supplementary figure. 1 II. The varieties corresponding to the proportions are: sorghum, rice, glutinous rice, wheat and corn. Picture a has the best verification effect, with a verification correct rate of 100%, and b picture has the worst recognition effect, with a verification correct rate of only 97%, and there are 3 misjudgments. There is 1 misjudgment in picture c, and the average verification correct rate is 99%. Therefore, the BPNN model trained with the One Step Secant function can accurately classify and identify the mixed grains of multiple grains, and the identification accuracy rate is 99%. Compared with the previous classification and identification of intact particles, broken particle samples are of different sizes and shapes, and the collected spectral data are irregular and have serious mutations, adding a lot of uncertainty (Q. Liu et al., 2021; Nie et al., 2019; Weng et al., 2021). The DBCAN-MD method used in this study can effectively eliminate the outliers caused by this reason and then establish a full-wavelength-based BPNN model, which further improves the classification accuracy of the BPNN model in the visual verification.

4 Conclusion

In this study, the DBSCAN-MD algorithm and BPNN were combined to establish a multi-grain mixed and broken particle resolution model. The optimal segmentation threshold of different varieties of broken particles was determined by the Otsu method, and then the optimized watershed algorithm was used to extract the spectral data

of each broken particle sample, and the extraction success rate was over 98%. The DBSCAN-MD algorithm was used to eliminate outliers, and the characteristic wavelengths were screened by IVSO-CARS, and four models of PNN, GRNN, BPNN and RBFNN based on full wavelengths and characteristic wavelengths were established for comparative analysis. However, since the results of the established full-wavelength and characteristic wavelength models are very close, the two wavelength models cannot be clearly distinguished. Therefore, seven different kernel functions are used to train the BPNN model, and a comparative analysis is carried out to determine that the BPNN model trained with the RPROP and OSS functions has the best effect. In order to further find the difference between the BPNN models established by the full wavelength and the characteristic wavelength, three groups of mixed samples were used for verification, and it was finally determined that the full wavelength model trained by the OSS function had the best effect, and the accuracy was improved by 3% compared with the characteristic wavelength model. Therefore, the BPNN model established by the combination of the visible light hyperspectral imaging technology and the optimization algorithm studied in this paper can realize the classification and identification of different varieties of grain broken particles and the determination of the mixed ratio. It provides a new reference for the mixed resolution of grain particles, and also provides new technical guidance for the winemaking industry to improve the quality of liquor.

Declarations

CRediT authorship contribution statement

Yu Lei: Writing - original draft. **Xinjun Hu:** Investigation, Supervision, Writing - review & editing. **Jiahong Zhang:** Conceptualization, Supervision, Writing - review & editing. **Songcai Yan:** . **Qinyuan Xue: .Xiaoyan Ma: . Jianping Tian: . Manjioa Chen: . Dan Huang: .**

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

The author appreciate the support of the Graduate Innovation Fund of Sichuan University of Light Chemical Industry (y2021033). Fund: Project supported by the National Natural Science Foundation of China (51701133).

References

1. Ai, Y.-j., Liang, P., Wu, Y.-x., Dong, Q.-m., Li, J.-b., Bai, Y.,.. . Ni, D. (2018). Rapid qualitative and quantitative determination of food colorants by both Raman spectra and Surface-enhanced Raman Scattering (SERS) *Food Chemistry*, *241*(Feb.15), 427–433. https://doi.org/10.1016/j.foodchem.2017.09.019
2. Alireza, L. P., & Negin, D. (2021). DBHC: A DBSCAN-based hierarchical clustering algorithm *Data* & *Knowledge Engineering*(prepublish). https://doi.org/10.1016/j.datak.2021.101922
3. Bo, K., Hoang, N., Xuan-Nam, B., Hoang-Bac, B., & Trung, N.-T. (2021). Prediction of the sorption efficiency of heavy metal onto biochar using a robust combination of fuzzy C-means clustering and back-propagation neural network Journal of Environmental Management, 293. https://doi.org/10.1016/j.jenvman.2021.112808

4. Bonah, E., Huang, X., Yi, R., Aheto, J. H., & Yu, S. (2020). Vis-NIR hyperspectral imaging for the classification of bacterial foodborne pathogens based on pixel-wise analysis and a novel CARS-PSO-SVM model. Infrared Physics & Technology, 105. https://doi.org/10.1016/j.infrared.2020.103220
5. Cabrera-Ramirez, A. H., Luzardo-Ocampo, I., Ramirez-Jimenez, A. K., Morales-Sanchez, E., Campos-Vega, R., & Gaytan-Martinez, M. (2020). Effect of the nixtamalization process on the protein bioaccessibility of white and red sorghum flours during in vitro gastrointestinal digestion. Food Res Int, *134*, 109234. https://doi.org/10.1016/j.foodres.2020.109234
6. Chen, Y., Xu, Z., Tang, W., Hu, M., Tang, D., Zhai, G., & Li, Q. (2021). Identification of various food residuals on denim based on hyperspectral imaging system and combination optimal strategy. Artificial Intelligence in Agriculture, *5*, 125–132. https://doi.org/10.1016/j.aiia.2021.06.001
7. Delorme, M. M., Guimaraes, J. T., Coutinho, N. M., Balthazar, C. F., Rocha, R. S., Silva, R.,.. . Cruz, A. G. (2020). Ultraviolet radiation: An interesting technology to preserve quality and safety of milk and dairy foods Trends in Food Science & Technology, *102*, 146–154. https://doi.org/10.1016/j.tifs.2020.06.001
8. Haoping, H., Xinjun, H., & Jianping, T. (2021). Rapid and nondestructive prediction of amylose and amylopectin contents in sorghum based on hyperspectral imaging *Food Chemistry*, *359*. https://doi.org/10.1016/j.foodchem.2021.129954
9. Haoping, H., Xinjun, H., Jianping, T., Jiang, X., Luo, H., & Huang, D. (2021). Rapid detection of the reducing sugar and amino acid nitrogen contents of Daqu based on hyperspectral imaging *Journal of Food Composition and Analysis*(prepublish). https://doi.org/10.1016/j.jfca.2021.103970
10. Hou, R., Hu, Y., Zhao, Y., & Liu, H. (2020). Hyperspectral image quality evaluation using generalized regression neural network. Signal Processing: Image Communication, 83. https://doi.org/10.1016/j.image.2020.115785
11. Jiayou, D., Qiaoming, S., Long, L., Ruijin, W., & Zefei, Z. (2021). Evaluation of the influence of aggregation morphology on thermal conductivity of nanofluid by a new MPCD-MD hybrid method *International Communications in Heat and Mass Transfer*, *127*. https://doi.org/10.1016/j.icheatmasstransfer.2021.105501
12. Lili, Z., Petros, S., Erica, P., & N., P. K. (2021). Deep learning and machine vision for food processing: A survey Current Research in Food Science, *4*. https://doi.org/10.1016/j.crfs.2021.03.009
13. Liu, B., Yu, A., Zhang, P., Ding, L., Guo, W., Gao, K., & Zuo, X. (2021). Active deep densely connected convolutional network for hyperspectral image classification International Journal of Remote Sensing, *42*(15), 5915–5934. https://doi.org/10.1080/01431161.2021.1931542
14. Liu, Q., Wang, Z., Long, Y., Zhang, C., Fan, S., & Huang, W. (2021). Variety classification of coated maize seeds based on Raman hyperspectral imaging. Spectrochim Acta A Mol Biomol Spectrosc, *270*, 120772. https://doi.org/10.1016/j.saa.2021.120772
15. Lyu, X., Liu, Y., Li, N., Ku, L., Hou, Y., & Wen, X. (2021). Foliar applications of various nitrogen (N) forms to winter wheat affect grain protein accumulation and quality via N metabolism and remobilization. The Crop Journal. https://doi.org/10.1016/j.cj.2021.10.009
16. Nandhini, K., & Porkodi, R. (2021). A new fusion of mutual information and Otsu multilevel thresholding technique for hyperspectral band selection. Egyptian Informatics Journal, *22*(2), 133–143. https://doi.org/10.1016/j.eij.2020.06.002
17. Nie, P., Zhang, J., Feng, X., Yu, C., & He, Y. (2019). Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning. Sensors and Actuators B: Chemical, 296. https://doi.org/10.1016/j.snb.2019.126630

18. Qian, W., Lu, Z. M., Chai, L. J., Zhang, X. J., Li, Q., Wang, S. T.,.. . Xu, Z. H. (2021). Cooperation within the microbial consortia of fermented grains and pit mud drives organic acid synthesis in strong-flavor Baijiu production. Food Res Int, *147*, 110449. https://doi.org/10.1016/j.foodres.2021.110449
19. Qiao, L., Wang, Z., & Zhu, J. (2020). Application of improved GRNN model to predict interlamellar spacing and mechanical properties of hypereutectoid steel. Materials Science and Engineering: A, 792. https://doi.org/10.1016/j.msea.2020.139845
20. Qiao, S., Tian, Y., Wang, Q., Song, S., & Song, P. (2021). Nondestructive detection of decayed blueberry based on information fusion of hyperspectral imaging (HSI) and low-Field nuclear magnetic resonance (LF-NMR) *Computers and Electronics in Agriculture*, *184*. https://doi.org/10.1016/j.compag.2021.106100
21. Shi, C., Qian, J., Zhu, W., Liu, H., Han, S., & Yang, X. (2018). Nondestructive determination of freshness indicators for tilapia fillets stored at various temperatures by hyperspectral imaging coupled with RBF neural networks *Food Chemistry*, *275*. https://doi.org/10.1016/j.foodchem.2018.09.092
22. Wang, Weiting, Yun, Yonghuan, Deng, Baichuan,.. . Yizeng. (2015). Iteratively variable subset optimization for multivariate calibration *RSC advances*, *5*(116). https://doi.org/10.1039/c5ra08455e
23. Warey, A., Gao, J., & Grover, R. O. (2021). Generalization performance of a deep learning based engine-out emissions model. *Energy and AI*, *5*. https://doi.org/10.1016/j.egyai.2021.100080
24. Weng, S., Han, K., Chu, Z., Zhu, G., Liu, C., Zhu, Z.,.. . Huang, L. (2021). Reflectance images of effective wavelengths from hyperspectral imaging for identification of Fusarium head blight-infected wheat kernels combined with a residual attention convolution neural network. Computers and Electronics in Agriculture, *190*. https://doi.org/10.1016/j.compag.2021.106483
25. Wenting, W., Wenpei, C., Ke, X., Hui, G., Haiyan, W., & Hongcheng, Z. (2021). Effects of Early- and Late-Sowing on Starch Accumulation and Associated Enzyme Activities During Grain Filling Stage in Rice. Rice Science, *28*(2), 191–199. https://doi.org/10.1016/j.rsci.2021.01.008
26. Xi, T., Chi, Z., Jiangbo, L., Shuxiang, F., Yi, Y., & Wenqian, H. (2021). Detection of early decay on citrus using LW-NIR hyperspectral reflectance imaging coupled with two-band ratio and improved watershed segmentation algorithm.. Food Chemistry, 360. https://doi.org/10.1016/j.foodchem.2021.130077
27. Xinna, J., Xinjun, H., Haoping, H., Jianping, T., Youhua, B., Dan, H., & Huibo, L. (2021). Detecting total acid content quickly and accurately by combining hyperspectral imaging and an optimized algorithm method Journal of Food Process Engineering, *44*(11). https://doi.org/10.1111/jfpe.13844
28. Yang, L., Su, Q., Si, B., Zhang, Y., Zhang, Y., Yang, H., & Zhou, X. (2022). Enhancing bioenergy production with carbon capture of microalgae by ultraviolet spectrum conversion via graphene oxide quantum dots. Chemical Engineering Journal, 429. https://doi.org/10.1016/j.cej.2021.132230
29. Yin, Z., Jun, G., Qian, Z., & Shuang, W. (2021). Research on damage identification of hull girder based on Probabilistic Neural Network (PNN) Ocean Engineering, 238. https://doi.org/10.1016/j.oceaneng.2021.109737
30. Zhao, K., Tao, Y., Liu, M., Yang, D., Zhu, M., Ding, J.,.. . Li, C. (2022). Does temporary heat stress or low temperature stress similarly affect yield, starch, and protein of winter wheat grain during grain filling? Journal of Cereal Science, 103. https://doi.org/10.1016/j.jcs.2021.103408
31. Zheng, F., Mengyi, W., Weifeng, H., & Siyuan, C. (2021). Potassium di-hydrogen phosphate identification based on wide energy X-ray absorption spectrum and an artificial neural network Computers and Electronics in Agriculture, *183*. https://doi.org/10.1016/j.compag.2021.106062

32. Zheng, J., Zhao, D., Peng, Z., Yang, K., Zhang, Q., & Zhang, Y. (2018). Variation of aroma profile in fermentation process of Wuliangye baobaoqu starter. Food Res Int, *114*, 64–71. https://doi.org/10.1016/j.foodres.2018.07.060

Tables

Table 1 Grain sample extraction results

		Sorghum	Rice	Sticky rice	Wheat	Corn
Models for training and testing	True	2200	2200	2200	2200	2200
	Extracted	2150	2167	2154	2188	2149
	ACUex	97.72%	98.50%	97.90%	99.45%	97.68%
External input validation set	True	300	300	300	300	300
	Extracted	297	296	297	300	300
	ACUex	99.00%	98.60%	99.00%	100.00%	100.00%

Abbreviations used: ACUex: data extraction success rate.

Table 2 Modeling Results

Model	Data sets	Number of wavelengths	Training sets (%)	Testing sets (%)	Validation sets (%)
PNN	All	428	92.4416	90.9891	87.2483
	IVSO-CARS	41	71.2207	70.7405	72.2818
GRNN	All	428	99.7299	96.841	15.4362
	IVSO-CARS	41	93.7273	85.189	19.9328
BPNN	All	428	99.9	99.7	99.2617
	IVSO-CARS	41	99.8	99.2	99.5302
RBF	All	428	95.4286	91.6106	90.6711
	IVSO-CARS	41	81.2467	80.4764	81.0067

Abbreviations used: PNN: Product Neural Network; GRNN: Generalized Regression Neural Network; BPNN: Back Propagation Neural Network; RBF: Radial Basis Function Neural Network;

Table 3 Model results of different kernel function training

KFUN	Number of wavelengths	Time(s)	Epoch	Training set(%)	Testing set(%)	Validation set(%)
RPROP	41	25	500	99.9	99.7	99.2617
	428	32	496	99.8	99.2	99.5302
GDA	41	24	500	74	75	70.7382
	428	31	500	92.1	91.5	90.5369
OSS	41	89	500	99.5	99.2	99.5302
	428	122	500	99.6	99.3	99.4631
CGF	41	87	499	86.5	85.9	89.1946
	428	134	500	99.8	99.4	98.7919
SCG	41	48	500	87.9	88.2	85.5772
	428	61	500	99.9	99.6	99.4631
GDX	41	24	500	79.4	80.5	77.1812
	428	32	500	97.7	97.7	97.7181
CGP	41	99	500	78.7	80.4	79.6644
	428	135	500	100	99.8	99.396

Abbreviations used: KFUN: Kernel Function; RPROP: Resilient training function; GDA: Gradient descent training function for adaptive learning rate; OSS: One Step Secant training function; CGF: Fletcher-Powell connection gradient training function; SCG: Scaled conjugate gradient training function; GDX: Gradient descent training function with momentum for adaptive learning rate; CGP: Polak-Ribiere connection gradient training function; EPOs: number of epochs to train.

Table 4 Validation results of three groups of mixed and broken particles

KFUN	Number of wavelengths	Group	Sorghum	Rice	Sticky rice	Wheat	Corn	ACU_{ex}	ACU_{cr}	AACU
RPROP	41	1	39	18	19	16	8	100%	96%	0.93
		2	39	12	25	16	8	100%	90%	
		3	37	14	25	16	8	100%	92%	
	428	1	36	22	18	16	8	100%	100%	0.95
		2	36	14	26	16	8	100%	92%	
		3	37	16	23	16	8	100%	94%	
OSS	41	1	36	20	19	17	8	100%	98%	0.96
		2	37	15	23	17	8	100%	93%	
		3	36	18	21	17	8	100%	96%	
	428	1	36	22	18	16	8	100%	100%	0.99
		2	35	23	16	18	8	100%	97%	
		3	36	22	17	17	8	100%	99%	

Abbreviations used: RPROP: Resilient training function; OSS: One Step Secant training function; ACUex: data extraction success rate; ACUcr: Classification and recognition accuracy; AACU : Average classification and recognition accuracy of three groups of samples.

Figures

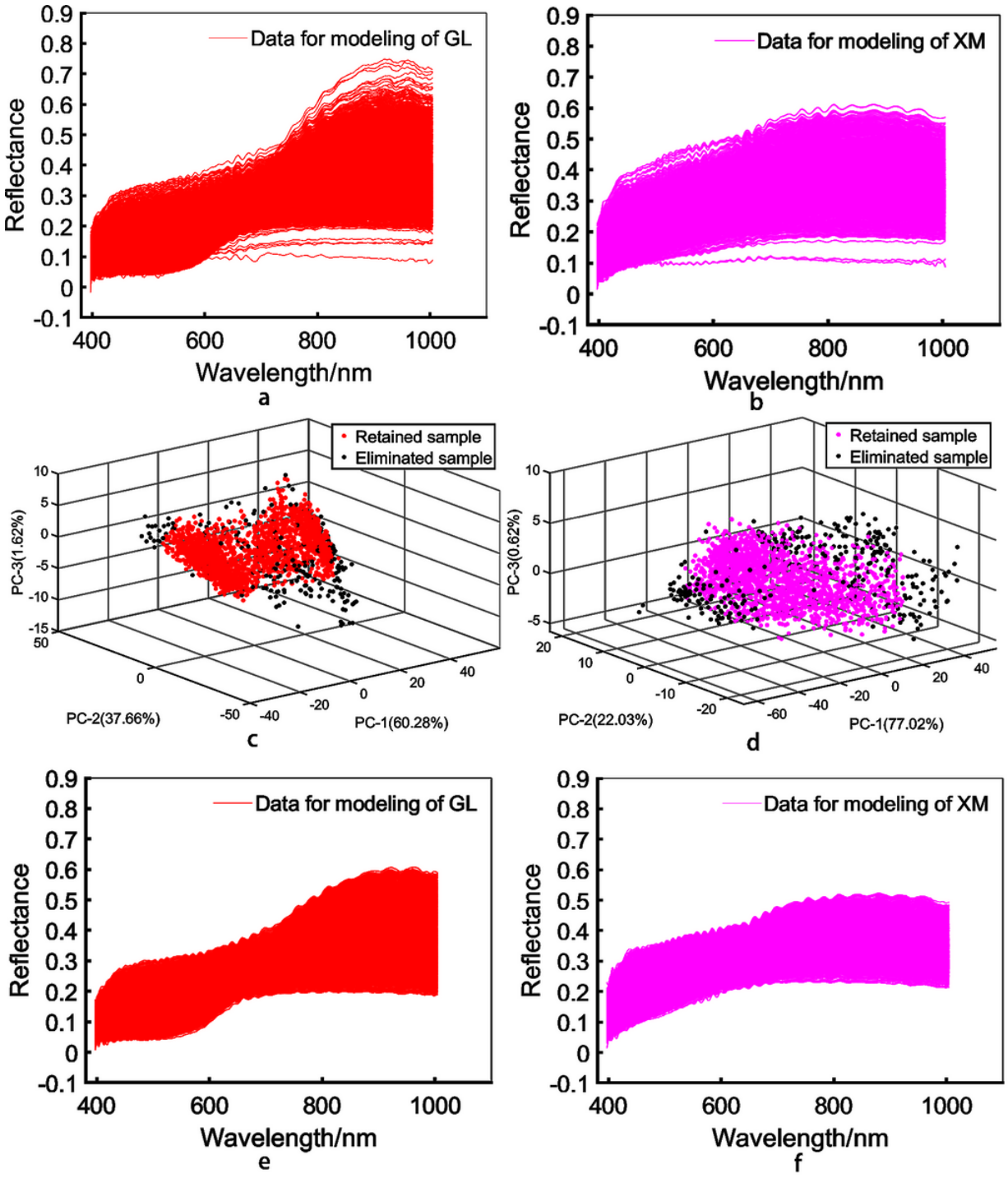

Figure 1

Analysis of the results of removing outliers, (a) original spectral curve of sorghum, (b) original spectral curve of wheat, (c) analysis results of outliers in sorghum, (d) analysis results of outliers in wheat, (e) abnormal removal of sorghum, (f) Spectral curve of wheat after removing outliers.

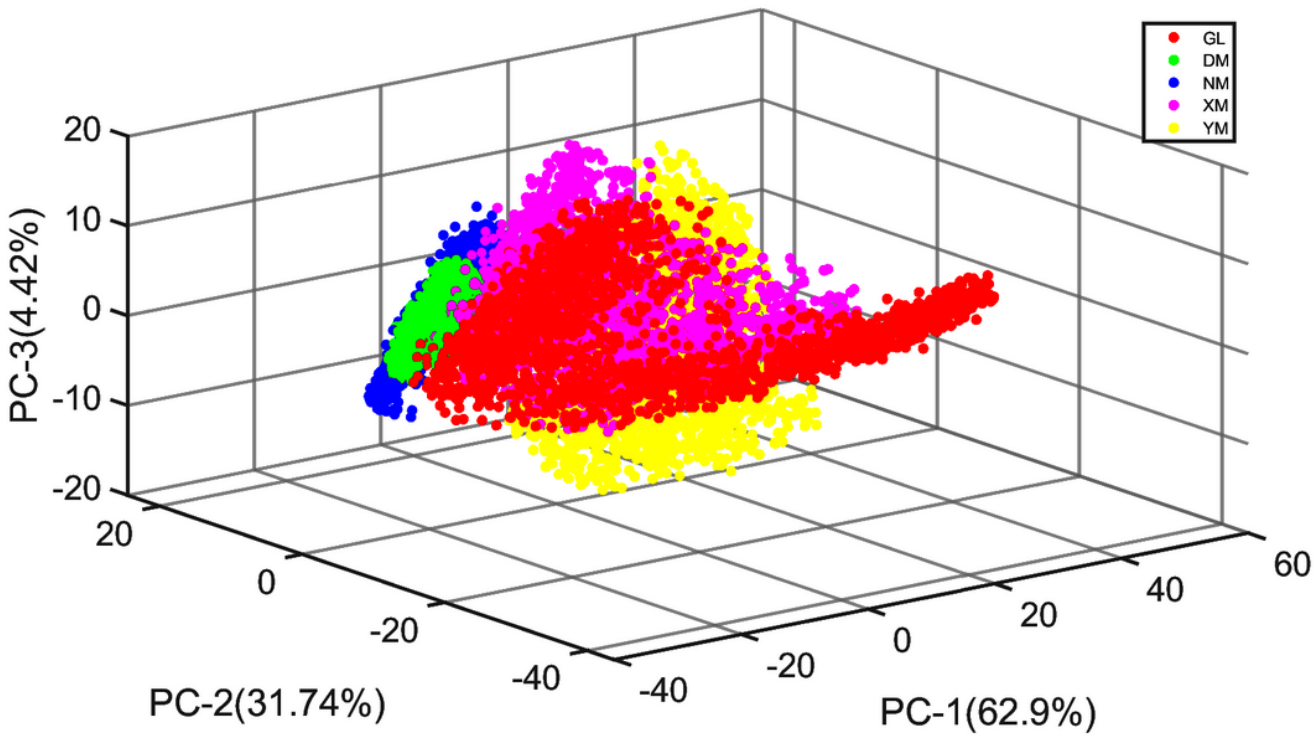

Figure 2

Three-dimensional principal component score map of multi-grain mixed broken particles

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfigure.tif