Detection of endogenous foreign bodies in Chinese hickory nuts by hyperspectral spectral imaging at the pixel level

Zhe Feng[1,2†], Weihao Li[1,2†], Di Cui[1,2*]

(1. *College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China*;
2. *Key Laboratory of On Site Processing Equipment for Agricultural Products, Ministry of Agriculture and Rural Affairs, Hangzhou 310058, China*)

Abstract: It is difficult to differentiate small, but harmful, shell fragments of Chinese hickory nuts from their kernels since they are very similar in color.   Including shell fragments of Chinese hickory nuts by mistake may create safety hazards for consumers.   Therefore, there is a need to develop an effective method to differentiate the shells from the kernels of Chinese hickory nuts.   In this study, a deep learning approach based on a two-dimensional convolutional neural network (2D CNN) and long short-term memory (LSTM) integrated with hyperspectral imaging for distinguishing the shells and kernels of Chinese hickory nuts at the pixel level was proposed.   Two classical classification methods, principal component analysis-*K*-nearest neighbors (PCA-KNN) and the support vector machine (SVM), were employed to establish identification models for comparison.   The results showed that the 2D CNN-LSTM model achieved the best performance with an overall classification accuracy of 99.0%.   Moreover, the shells in mixtures of shells and kernels were detected based on the proposed deep learning method and visualized for subsequent operations for the removal of foreign bodies.
Keywords: Chinese hickory nut, endogenous foreign body, hyperspectral spectral imaging, pixel level, detection
DOI: 10.25165/j.ijabe.20221502.6881

1   Introduction

Chinese hickory (*Carya cathayensis* Sarg.) in the genus *Carya* (Juglandaceae) is an important commercially cultivated nut tree[1]. Its nuts are popular in eastern China and are well known for their daintiness and nutritional content[2].   Compared to pecan, Chinese hickory nuts have a smaller size and harder shell.   Furthermore, the difference between the shell and kernel is also small.   The physical and structural characteristics of Chinese hickory nuts create difficulties for shell breaking and shell-kernel separation. Although most shells are removed by airflow shell-breaking machines, a few small shell fragments occasionally remain.   The presence of shell fragments not only affects product quality but also creates safety hazards for consumers.   Therefore, there is a need to develop an effective method to detect small shell fragments produced during the processing of Chinese hickory nuts for quality control and safety assurance.

The shell fragments of Chinese hickory nuts, which are intrinsic foreign bodies different from the food product itself, are very similar in color to kernels[3].   Currently, small shell fragments are manually removed, which is labor-intensive and subjective[4].   Since the last century, many noninvasive techniques have been developed for the detection of foreign bodies in food,

such as X-ray, computer vision, thermal imaging, spectroscopy, hyperspectral imaging, ultrasonication, and terahertz.   Among these methods, hyperspectral imaging technology, which is sensitive to minor components, provides spatial and spectral information about objects[5].   Therefore, researchers have attempted to employ this technique to develop objective methods for the discrimination of shells and kernels of nuts.   Jiang et al.[6] acquired hyperspectral images of shells and kernels of walnuts under UV fluorescent lamps at 365 nm and classified walnut shells and kernels using principal component analysis and a Gaussian mixture model (PCA-GMM)-based Bayesian approach, an overall 90.3% recognition rate was achieved.   To improve the overall recognition rate, Jiang et al.[7] adopted a Gaussian-kernel-based support vector machine (SVM) algorithm to analyze the hyperspectral fluorescence images of walnut shells and kernels, the overall recognition rate increased to 95.6%.

Although hyperspectral imaging technology has shown great potential for distinguishing shells from nut kernels, the major challenge in applying this technology in food processing is handling the huge amount of data in real time.   In recent years, deep learning has achieved remarkable success in big data analysis. Many attempts have been made to analyze hyperspectral images using deep learning.   Mahmoud et al.[8] succeed in detecting adulteration in different states of red meat products using hyperspectral imaging and deep learning.   The spectral-spatial features in hyperspectral images were extracted by a deep convolution neural network (CNN) to establish the detection models.   The developed deep-learning approach was able to extract robust features from raw hyperspectral images independently of the states of meat products and was more suitable for real-time applications.   Jin et al.[9] applied a deep neural network algorithm to classify the pixels of hyperspectral images to identify the diseased area of wheat heads.   The one-dimensional

pixel spectral data were reshaped into a two-dimensional structure as the input layer of the CNN. The results indicated that the two-dimensional CNN model achieved better performance than the one-dimensional CNN model. Alvaro et al.[10] used deep learning based detector for real-time recognition of tomato plant diseases and pests. Kang et al.[11] used deep learning to realize fruit detection with 28 ms inference time. Zhu et al.[12] identified seven varieties of cotton seeds using near-infrared hyperspectral imaging combined with deep learning. A classification model was established based on a self-designed CNN and a residual network (ResNet), which achieved better performance than the classification models based on partial least squares discriminant analysis (PLS-DA), the logistic regression (LR), and the support vector machine (SVM) with full spectra as inputs. The results indicated that deep learning provided an effective solution for analyzing hyperspectral imaging data.

Chinese hickory nuts are a special local product of Lin'an in Zhejiang Province, China. Local manufacturers need an objective method to detect shell fragments to replace manual inspection. Therefore, the goal of this research was to develop a method using hyperspectral imaging and deep learning for Chinese hickory nut manufacturers to discriminate shells from kernels. This goal was accomplished by 1) establishing a classification model based on deep learning to distinguish the shells and kernels of Chinese hickory nuts at the pixel level; 2) evaluating the performance of the proposed model by comparing its performance with those of classic classification models; 3) visualizing shell fragments.

2   Materials and methods

2.1   Sample preparation

As shown in Figure 1, the samples were divided into 4 categories according to the structure and composition of Chinese hickory nuts: the inner shell, outer shell, light kernel, and dark kernel. Among these components, the light kernel and dark kernel are the inside and outside of the kernel, respectively. The color of the light kernel is ivory, which is different from the colors of the other categories. The dark kernel, inner shell, and outer shell are similar in color, especially the dark kernel and the inner shell.

Figure 1   Four categories based on the structure and composition of Chinese hickory nuts

Almost all Chinese hickory nuts were produced in Lin'an, Zhejiang Province, China; purchased at a local supermarket, and broken manually in the laboratory. A total of 213 samples including 53 dark kernel fragments, 55 light kernel fragments, 62 outer shell fragments, and 43 inner shell fragments were obtained.
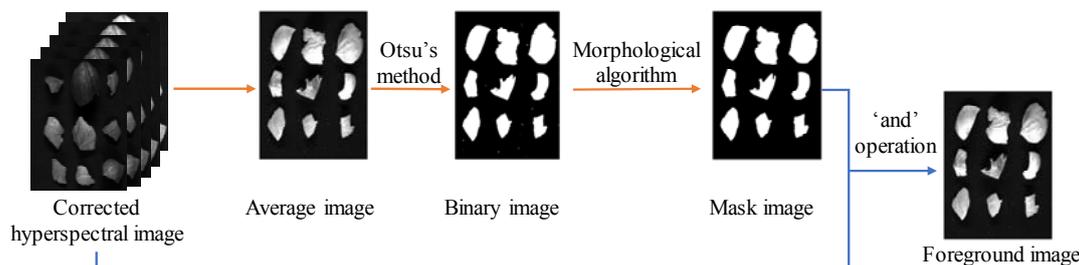
Each category was placed on black hardboards (TB5, Thorlabs Inc., USA) to take hyperspectral images separately for establishing and testing the detection models.

2.2   Experimental system setup

An array charge-coupled device (CCD) camera (C8484-05G01, Hamamatsu Photonics, Japan), a line scan spectrometer with a spectral range of 400-1000 nm (ImSpector V10E-QE, Spectral Imaging Ltd, Finland) and a 150 W halogen light source (2900, Illumination Technologies, Inc., USA) were the main components of the hyperspectral imaging system in this study. The spectral resolution was 2.8 nm. Since there was considerable noise above 900 nm, the spectra in the range of 400-900 nm were used for subsequent analysis (400 wavebands in total). The samples were spread manually on black hardboards. Furthermore, an electric displacement platform (TSA200-B, Beijing Zhuoli Instrument Co., Ltd., China) carried the samples of each class on black hardboards to perform line scanning. The speed of the electric displacement platform in this experiment was 1.9 mm/s. The distance between the camera and the samples was 50 cm, and the exposure time was set at 8.5 ms. A dedicated computer (Intel RcoreTM2 4400 @ 2.00 GHz, ACER, China) was used to collect the hyperspectral images through commercial software (SpectralCube_v2_75, Spectral Imaging Ltd., Finland).

2.3   Hyperspectral image correction and pixel size acquisition

To obtain the reflectance and eliminate the noise in the spectral images, the raw hyperspectral image was first corrected by the following equation[13].

$$R = \frac{R_{\text{sample}} - R_{\text{dark}}}{R_{\text{reference}} - R_{\text{dark}}} \qquad (1)$$

where, R_{sample} is the raw hyperspectral image; R is the corrected hyperspectral image; R_{dark} is the black image with the camera lens covered and the light source turned off; $R_{\text{reference}}$ is the white image with a 99.9% reflectance Teflon panel (Isuzu Optics Corp., Shanghai, China).

Moreover, a printed checkerboard was used to acquire the size of a pixel in the hyperspectral image collected in this study. The length of each grid on the checkerboard was 89.0 mm. The printed checkerboard was also placed on the electric displacement platform to acquire its hyperspectral image. By counting the number of pixels representing the length of each grid in the hyperspectral image, the pixel size was acquired by calculating the ratio of the actual length (89.0 mm) to the number of pixels.

2.4   Background removal and pixel spectrum extraction

After the hyperspectral images were corrected, the background of the resulting images was removed by calculating the average image of the whole band images, segmenting the average image using the Otsu thresholding method for binarization, denoising using morphological filtering, and using the hole filling algorithm to obtain a mask image. Subsequently, the 'and' operation was performed between the mask image and the calibrated hyperspectral images. The above procedure is shown in Figure 2.

Figure 2   Hyperspectral image background segmentation process

Each hyperspectral image was processed by the above steps. Finally, the shell and kernel pixels in the hyperspectral images of 4 categories were isolated, and the spectra were extracted as the inputs of the classification models.   Finally, 10 000 spectra were acquired.   Each category has 2500 spectra, of which 2000 were used as the training set and the rest were used as the testing set. The classification models were established using the proposed deep learning method and classic multivariate analysis methods with the training sets of four categories as inputs (Section 2.4).   The above background removal procedure and pixel spectrum extraction were performed in Matlab R2020a.   The deep learning model and classical classification models were established by Keras (v.2.1.2) with the TensorFlow (v.1.4.1) backend and Scikit-Learn (v.0.23.1) in Python (v.3.6.2), respectively.   The software was installed in a high-performance computer equipped with an Intel i7-8700K CPU, an NVIDIA GeForce GTX 1080ti Graphics card, 32 GB RAM, and 500 GB SSD.

2.5   Multivariate analysis

The spectra of pixels belonging to four categories were analyzed using deep learning, principal component analysis-K-nearest neighbor (PCA-KNN), and support vector machine (SVM) methods, respectively.   The classification models based on the PCA-KNN and SVM are compared with the deep learning model.

2.5.1   Deep learning

A convolutional neural network (CNN) is a feed-forward neural network typically containing a convolution layer, pooling layer, and classification layer[14].   It is widely used for object recognition and classification.   The convolution layer extracts high-level abstracted feature representations from the input data. The pooling layer merges semantically similar features into one feature[15].   Furthermore, the classification layer classifies the input data.   A recurrent neural network (RNN), which is designed to recognize the sequential characteristics in data, is another effective neural network[16].   An RNN produces its output by considering not only its current input but also the history of its previous inputs[17].   Since the basic RNN has the problem of long-term dependencies, a special type of RNN, long short-term memory (LSTM) networks, is introduced to solve this problem[18].

Each pixel spectrum of hyperspectral images can be regarded as a data sequence since different wavelengths of the spectrum are correlated with each other[19].   In this study, a CNN was integrated with LSTM to take advantage of the characteristics of the convolutional and recurrent layers.   As shown in Figure 3, the proposed deep learning architecture was composed of 4 two-dimensional convolutional layers, 2 pooling layers, and a 3-layer stacked LSTM.   The high-level features of spectra were extracted in the convolutional layers, semantically similar features were merged in the pooling layers, and the contextual information of the features was obtained from stacked LSTM.   First, the spectral vector of the hyperspectral image pixel (1×400) was normalized and reshaped into two dimensions (20×20).   The normalized spectral vector was one-dimensional and had 400 elements.   When the spectral vector was reshaped, the first line of the new two-dimensional matrix (20×20) consisted of elements 1-20 of the spectral vector, and the second line consisted of elements 21-40.   By analogy, the normalized pixel spectrum was reshaped into two dimensions.   Four two-dimensional convolutional layers were used to extract deep features from the spectra.   Every 2 convolutional layers were followed by a max pooling layer to acquire the compressed feature representation. The convolution kernel sizes of the convolutional layers and max pooling layers are 3×3 and 2×2, respectively.   Each convolutional layer of the architecture generates 64 feature maps.   Before entering the stacked LSTM, the data were flattened into a one-dimensional vector.   The softmax activation function was adopted to acquire multiple scores over all the categories for an input spectrum and output the category with the highest score. The L2 regularization method was used to prevent overfitting[20]. The batch size and the regularization parameter λ were set as 128 and 0.01, respectively.   The cross-entropy loss function was used to measure the errors between the predicted outputs and real outputs[21].   The Adam optimizer was adopted to optimize the weights of the proposed model[22].   The configuration of the proposed deep learning architecture based on 2D CNN-LSTM is listed in Table 1.
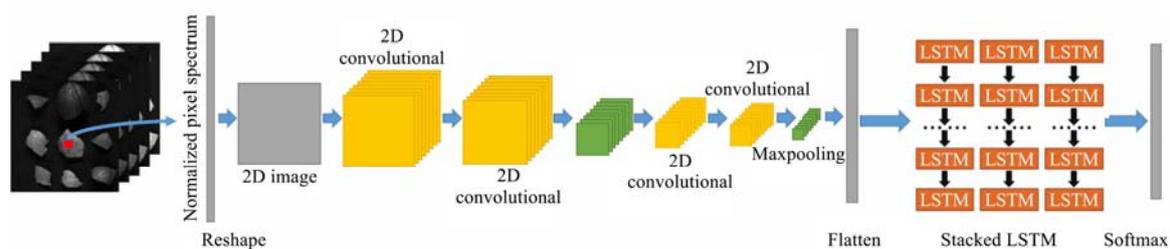
Note: LSTM: Long short-term memory.

Figure 3   Proposed deep learning architecture integrated 2D CNN with LSTM

Table 1   Structural parameters of 2D CNN-LSTM

Layer	Kernel size	Stride	Output size	Activation function
2D Convolutional	3×3	1×1	18×18×64	ReLU
2D Convolutional	3×3	1×1	16×16×64	ReLU
Max pooling	2×2	--	8×8×64	--
2D Convolutional	3×3	1×1	6×6×64	ReLU
2D Convolutional	3×3	1×1	4×4×64	ReLU
Max pooling	2×2	--	2×2×64	--
LSTM	--	--	64×1	ReLU
LSTM	--	--	64×1	ReLU
LSTM	--	--	64×1	ReLU
Classification	--	--	4×1	Softmax

Note: Stride is the number of jumps required when a filter scans the original image.   ReLU: Rectified Linear Unit; CNN: Convolution Neural Network; LSTM: Long short-term memory.

2.5.2   PCA-KNN

Principal component analysis (PCA) is a classical statistical method that can reduce the dimensionality of a dataset by converting the original variables to a new set of orthogonal variables called principal components[23].   To reduce the dimension of the spectra, PCA was performed before modeling based on the K-nearest neighbors (KNN) algorithm.   The KNN, which is a nonparametric classifier, is a supervised machine learning algorithm[24].   The classes of new observation data are predicted according to the majority class of the K-nearest neighbors in the training set.

In this study, the number of components was determined by the cumulative contribution rate of the variance of principal

components.   When the first n principal components retained 90% of the variance of the original data, they were fed as the inputs of the KNN model.   The selection of the neighborhood size K was optimized by the tenfold cross-validation method.

2.5.3   SVM

The support vector machine (SVM) is a supervised machine learning algorithm for linear and nonlinear classification that separates the classes with a decision surface or hyperplane that maximizes the margin between the classes[25].   The SVM has many unique advantages in solving small samples, nonlinear, and high dimensional pattern recognition and is widely used to analyze multispectral and hyperspectral images[22,26].   The effectiveness of an SVM depends on the selection of the kernel, the kernel's parameters, and the soft margin parameter C[27].   In this study, the Gaussian radial basis function (RBF) was used as the SVM kernel function.   The kernel's parameter γ and soft margin parameter C, which were selected using the tenfold cross-validation method, were 0.9 and 0.8, respectively.

2.6   Model evaluation

Accuracy, precision, recall, and F1 score[28] were used to evaluate the performance of the classification model in each category; and the metrics were calculated using Equations (2)-(5), respectively.   The values of these indicators are in the range of 0-100%.   The closer the value is to 100%, the better the performance of the model.   Equation (5) indicates that the F1 score is a better metric of the model performance since the F1 score combines precision with recall.

$$Accuracy = \frac{Correctly\ classified\ number\ of\ samples}{Total\ number\ of\ samples} \times 100\% \quad (2)$$

$$Precision\ (i) = \frac{TP(i)}{TP(i) + FP(i)} \times 100\% \quad (3)$$

$$Recall\ (i) = \frac{TP(i)}{TP(i) + FN(i)} \times 100\% \quad (4)$$

$$F1\ Score\ (i) = \frac{2*Precision\ (i)*Recall\ (i)}{Precision\ (i) + Recall\ (i)} \times 100\% \quad (5)$$

where, $TP(i)$, $FP(i)$, and $FN(i)$ represent the true positives, false-positives, and false negatives of category i (i=1,2,3,4), respectively. $TP(i)$ represents the correctly classified number of samples in category I; $FN(i)$ represents the misjudged number of samples in category I; $FP(i)$ represents the number of samples in other categories that were misjudged as category i. $TP(i)$, $FP(i)$, and $FN(i)$ were extracted from the confusion matrix.

3   Results and discussion

3.1   Pixel size and spectrum analysis

The length of each grid on the checkerboard corresponded to 996 pixels in the hyperspectral image.   Since the actual length of each grid was 89.0 mm, the pixel size was approximately 0.089 mm.   Therefore, the actual area corresponding to a single pixel in the hyperspectral image was approximately 0.008 mm^2. Figure 4 shows the resulting images of a raw hyperspectral image processed according to the procedure described in Section 2.3.

After an 'and' operation was performed between the mask image and the corrected hyperspectral images, the shell and kernel pixels in the hyperspectral images were isolated.   Then, the pixel spectra of the four categories were extracted.   Their mean reflectance spectra are shown in Figure 5.   It was observed that the light kernel had higher reflectance in the range of 400-900 nm than other categories.   The average spectral reflectance of the dark

kernel was lowest in the range of 400-650 nm, which then increased faster than other categories.   In general, the mean reflectance spectra of the outer shell, the inner shell, and the dark kernel overlapped one another; and their waveforms were similar. Therefore, it was difficult to distinguish four categories of Chinese hickory nuts using their mean reflectance spectra.

a. Average image   b. Resulting image after   c. Mask image
                           background segmentation
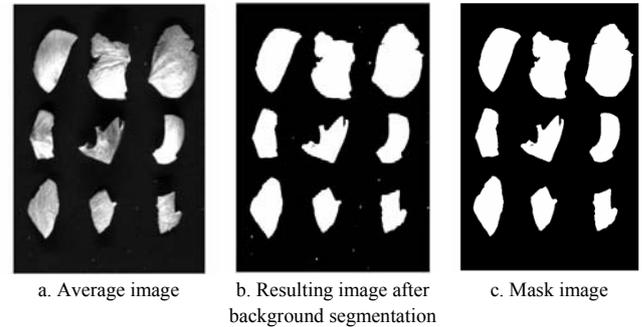
Figure 4   A typical average image, resulting image after background segmentation, and mask image
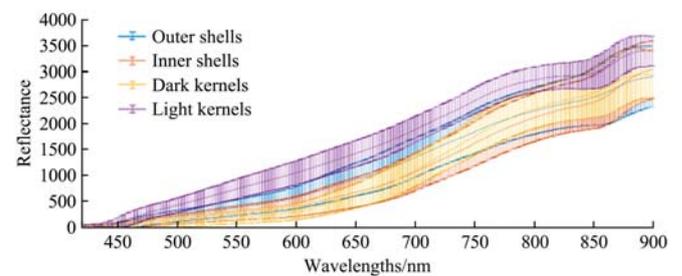
Figure 5   Mean reflectance spectra of the four components of Chinese hickory nuts in the range of 400-900 nm

3.2   Modeling based on CNN-LSTM

Since there was no significant difference between the reflectance spectra of kernels and shells of Chinese hickory nuts, it was difficult to distinguish them using feature wavelengths. Therefore, a model based on CNN-LSTM was established to achieve this goal.   The deep learning architecture integrated a CNN with LSTM.   The extracted spectrum of a pixel was regarded as a one-dimensional vector with a length of 400.   The element in the ith column corresponded to the reflectance at the ith wavelength (i=1, 2, 3, …, 400).   First, this vector was converted into two dimensions (20×20), whose element in the mth row and nth column (m=1, 2, 3, …, 20 and n=1, 2, 3, …, 20) represented the reflectance at the kth wavelength (k=(2m−2)×10+n).   The new matrix was input into the CNN-LSTM model.   The proposed architecture contained 4 convolutional layers, 2 pooling layers, and 3-layer stacked LSTM.   Figure 6 shows the loss curves acquired in the training process of the proposed CNN-LSTM model.   The cross-entropy loss function was used, which learned quickly when the model was poor and learned slowly when the model was good; therefore, the loss function first decreased rapidly and then decreased slowly.   When the model effect gradually improves, the decline speed of the loss function will also slow down.   It was observed that the loss decreased rapidly at the beginning and then slowly with local fluctuations.   The experimental results showed that the model loss function decreased to 0 after the 150th iteration when the parameters were set.   Therefore, the number of iterations in training the model was determined based on the results of the loss values.   Then, the model training process ended.   The trained CNN-LSTM model and its weights were saved.

Figure 7 and Table 2 show the confusion matrix, precision, recall, and F1 score acquired by the CNN-LSTM model for

classifying the four hickory nut categories of the testing set.   The results showed that the correctly classified numbers of outer shells, inner shells, dark kernels, and light kernels were 480, 500, 500, and 500, respectively.   According to Equation (2), the accuracy of the CNN-LSTM model for the testing set equaled (480+500+500+500)/2000×100%=99.0%.   Figure 7 also shows that all samples of the inner shells, dark kernels, and light kernels were judged correctly.   Twenty outer shell pixels were misjudged, but only four of them were classified as kernels.   The precision and recall values of the four categories were all higher than 96.9%.   The F1 scores of the four categories were greater than 98.0%, but the outer shells and inner shells had lower F1 scores than the dark kernels and light kernels.   The results indicated that the proposed CNN-LSTM model achieved satisfactory performance in classifying four categories of Chinese hickory nuts and was better at identifying kernels than shells.
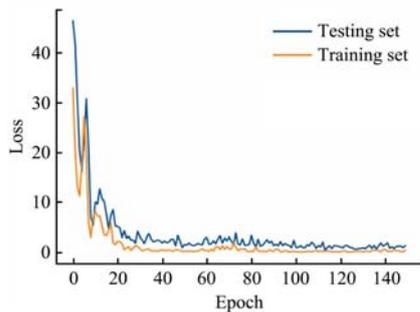
Figure 6   Loss change curves in CNN-LSTM model training

Figure 7   Confusion matrix acquired by the CNN-LSTM model for classifying the four categories of the testing set

Table 2   Precisions, recalls, and F1 scores acquired by CNN-LSTM model for classifying four categories of the testing set

Indicators	Outer shells	Inner shells	Dark kernels	Light kernels
Precision	96.0%	100.0%	100.0%	100.0%
Recall	100.0%	96.9%	99.2%	100.0%
F1 score	98.0%	98.4%	99.6%	100.0%

3.3   Modeling based on PCA-KNN and SVM

For comparison, the detection model was also established based on classical modeling approaches, the PCA-KNN, and the SVM.   The variance contribution rate of the first ten principal components is shown in Figure 8.   The figure shows that the proportion of the first principal component is very large, accounting for 78%; and the second principal component accounts for 13%.   The cumulative variance contributions of the first two principal components exceeded 90%.   Therefore, the first two principal components were used to establish the PCA-KNN model. A tenfold cross-validation method was used to optimize the neighborhood size K of the PCA-KNN model and the kernel parameter r and soft margin parameter C of the SVM model.   In

this study, the optimal K, r, and C were 3, 0.9, and 0.8, respectively.
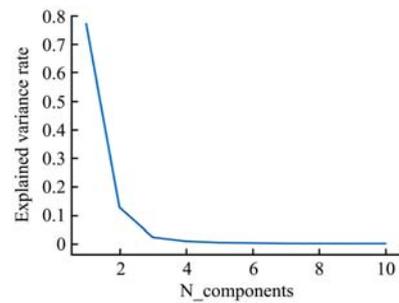
Figure 8   Variance contributions of the first ten principal components of Chinese hickory nut spectra

Figure 9 and Table 3 show the confusion matrices, precision, recall, and F1 scores acquired by the PCA-KNN and SVM models for classifying the four categories of the testing set.   The results showed that the overall accuracy of the PCA-KNN and SVM models was 94.1% and 93.0%, respectively.   For the PCA-KNN model, light kernels had the highest precision (95.0%).   This might be because the spectra of the light kernels were significantly different from those of other components.   The dark kernels had the second-highest precision and the lowest recall, which indicated that a relatively large number of dark kernel pixels were misjudged. The precision of the inner shells was the lowest, but their recall was the highest.   This meant that many other components were classified as inner shells.   The F1 scores of the four categories showed that the PCA-KNN model also had higher accuracies for identifying kernels than shells.   Misjudged samples of shells appeared in the inner shell and outer shell groups, and only 7 of these were classified as kernels.   In contrast, more misjudged samples of shells in the SVM model were classified as kernels. This would increase the occurrence of food safety incidents. Generally, the performance of the PCA-KNN model was better than that of the SVM model but still not as good as that of the CNN-LSTM model.

a. PCA-KNN

b. SVM

Figure 9   Confusion matrices acquired by the PCA-KNN and SVM models for classifying four categories of the testing set

Table 3   Precision, recall, and F1 scores acquired by PCA-KNN and SVM models for classifying four categories of the testing set

Modeling methods	Indicators	Outer shells	Inner shells	Dark kernels	Light kernels
PCA-KNN	Precision	92.4%	97.4%	91.4%	95.0%
	Recall	90.9%	88.5%	97.9%	100.0%
	F1 score	91.6%	92.7%	94.5%	97.4%
SVM	Precision	94.3%	83.2%	96.6%	100.0%
	Recall	86.6%	97.6%	95.2%	92.4%
	F1 score	90.3%	89.8%	95.9%	96.0%

3.4   Foreign body visualization

For subsequent removal operations, foreign bodies need to be visualized.   First, the proposed 2D CNN-LSTM model was used to detect and visualize the shells in the hyperspectral images of mixtures of shells and kernels at the pixel level.   Some color images of the mixtures are shown in Figure 10a.   The 4-category classification results of the 2D CNN-LSTM model are shown in Figure 10b.   The 4 categories of Chinese hickory nuts are marked with different colors.   The figure shows that there were some misjudged pixels.   For the ease of the subsequent removal operations of foreign bodies, the inner and outer shell pixels were further merged into one category (foreign body) marked in red, and the light and dark kernel pixels were also merged into another category (food) marked in blue (Figure 10c).   Although the number of misjudged pixels decreased, a few remained.   Since the actual area corresponding to a single pixel in a hyperspectral image was 0.008 mm^2, a shell with an area of 1 mm^2 corresponds to 125 connected pixels in a hyperspectral image.   The actual area of a shell is usually larger than 1 mm^2.   Therefore, the number of pixels in each connected domain in Figure 10c was calculated.   If the number was less than 125, the color of the pixel was changed to the color of the nearest connected domain.   In this way, all endogenous foreign bodies were accurately detected, and the final visualization results are shown in Figure 10d.

a. Color images   b. 4-category classification results of the 2D CNN-LSTM model (outer shells in red, inner shells in green, dark kernels in blue and light kernels in yellow)   c. binary classification results (food in blue and foreign bodies in red)   d. Final visualization results for foreign body removal
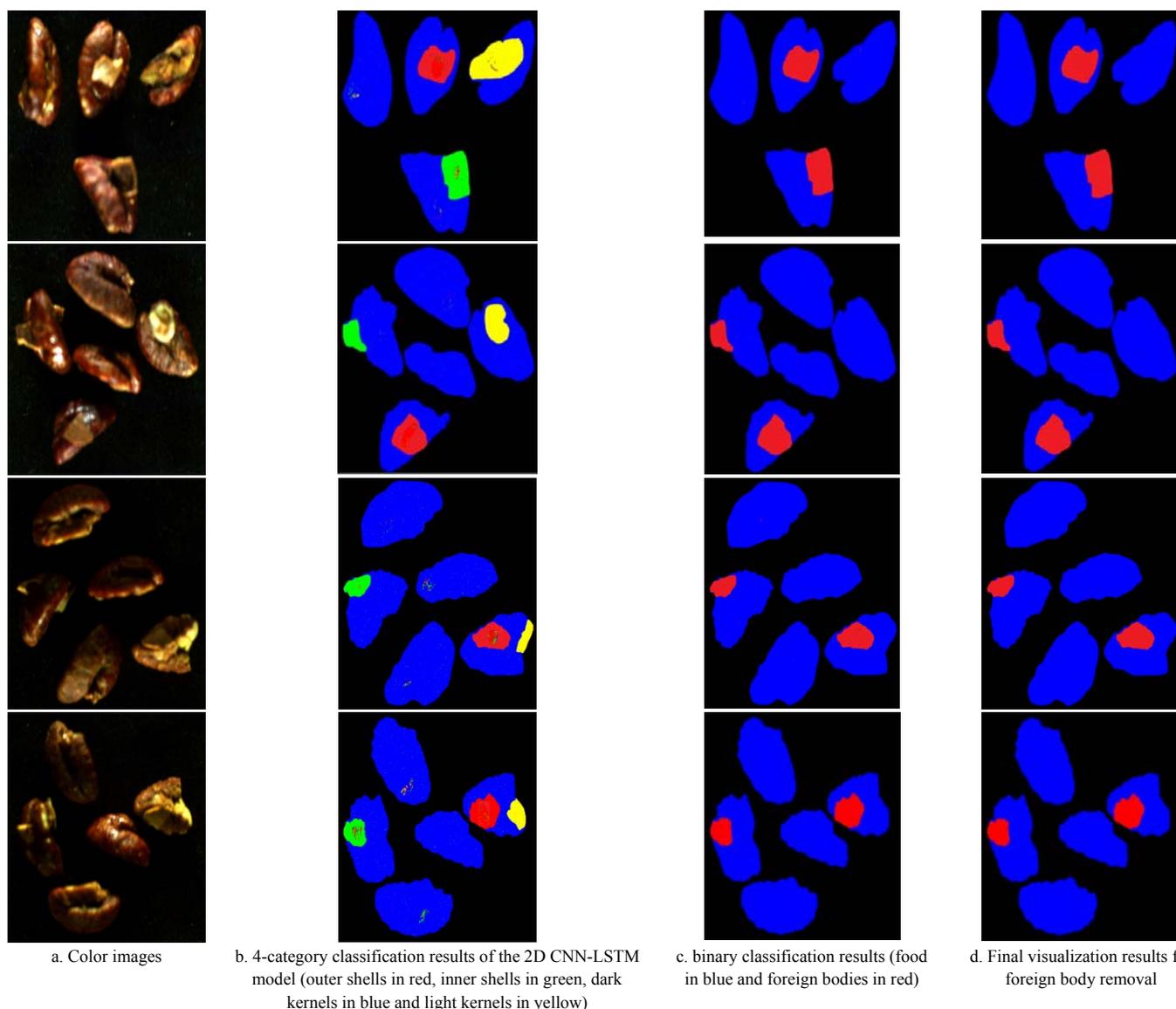
Figure 10   Visualization of endogenous foreign bodies in hyperspectral images of mixtures of shells and kernels

4   Conclusions

In this study, a deep learning approach was proposed for the detection of endogenous foreign bodies in Chinese hickory nuts based on hyperspectral spectral imaging and 2D CNN-LSTM. The mixtures of shells and kernels were classified into 4 categories. The spectra of each category at the pixel level were extracted after removing the background of corrected hyperspectral images, which

were reshaped from one dimension into two dimensions for inputting the 2D CNN-LSTM model. The overall accuracy of 99.0% was achieved by the trained 2D CNN-LSTM model for the testing set. For comparison, the models based on the PCA-KNN and SVM were also established and achieved accuracies of 94.1% and 93.0%, respectively. Although the overall accuracies of the PCA-KNN and SVM models were acceptable, their F1 scores for the inner and outer shell categories were significantly lower than those of the 2D CNN-LSTM model. This result indicated that the occurrence rate of food safety incidents caused by the accidental ingestion of shells would increase. Therefore, the 2D CNN-LSTM model was more promising than the PCA-KNN and SVM models. Moreover, the shells in mixtures of shells and kernels were visualized for ease of subsequent operations for the removal of foreign bodies.

Acknowledgements

[References]

[1]   Huang Y J, Xiao L H, Zhang Z R, Zhang R, Wang Z J, Huang C Y, et al. The genomes of pecan and Chinese hickory provide insights into *Carya* evolution and nut nutrition. GigaScience, 2019; 8(5): giz036. doi: 10.1093/gigascience/giz036.

[2]   Hu L S, Du M H, Zhang J P. Physicochemical properties and nutritional ingredients of kernel oil of *Carya cathayensis* Sarg. American Journal of Plant Sciences, 2018; 9(12): 2494–2503.

[3]   Edwards M C, Stringer M F. Observations on patterns in foreign material investigations. Food Control, 2007; 18(7): 773–782.

[4]   Jiang L, Zhu B, Rao X Q, Berney G, Tao Y. ICA based band selection for black walnut shell and meat classification in hyperspectral fluorescence imagery. In: 2007 ASAE Annual Meeting, 2007; Paper No. 073088. doi: 10.13031/2013.22947.

[5]   Mohd Khairi M T, Ibrahim S, Md Yunus M A, Faramarzi M. Noninvasive techniques for detection of foreign bodies in food: A review. Journal of Food Process Engineering, 2018; 41(6): e12808. doi: 10.1111/jfpe.12808.

[6]   Jiang L, Zhu B, Rao X Q, Berney G, Tao Y. Discrimination of black walnut shell and pulp in hyperspectral fluorescence imagery using Gaussian kernel function approach. Journal of Food Engineering, 2006; 81(1): 108–117.

[7]   Jiang L, Zhu B, Jing H, Chen X, Rao X Q, Tao Y. Gaussian mixture model-based walnut shell and meat classification in hyperspectral fluorescence imagery. Transactions of the ASABE, 2007; 50(1): 153–160.

[8]   Al-Sarayreh M, Reis M M, Yan W Q, Klette R. Detection of red-meat adulteration by deep spectral-spatial features in hyperspectral images. Journal of Imaging, 2018; 4(5): 63. doi: 10.3390/jimaging4050063.

[9]   Jin X, Jie L, Wang S, Qi H J, Li S W. Classifying wheat hyperspectral pixels of healthy heads and *Fusarium* head blight disease using a deep neural network in the wild field. Remote Sensing, 2018; 10(3): 395. doi: 10.3390/rs10030395.

[10]   Fuentes A, Yoon S, Kim S C, Park D S. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors, 2017; 17(9): 2022. doi: 10.3390/s17092022.

[11]   Kang H W, Chen C. Fast implementation of real-time fruit detection in apple orchards using deep learning. Computers and Electronics in Agriculture, 2020; 168: 105108. doi: 10.1016/j.compag.2019.105108.

[12]   Zhu S S, Zhou L, Gao P, Bao Y D, He Y, Feng L. Near-infrared hyperspectral imaging combined with deep learning to identify cotton seed varieties. Molecules, 2019; 24(18): 3268. doi: 10.3390/molecules24183268.

[13]   Qin J W, Chao K L, Kim M S, Lu R F, Burks T F. Hyperspectral and multispectral imaging for evaluating food safety and quality. Journal of Food Engineering, 2013; 118(2): 157–171.

[14]   Fan S X, Li J B, Zhang Y H, Tian X, Wang Q Y, He X, et al. On line detection of defective apples using computer vision system combined with deep learning methods. Journal of Food Engineering, 2020; 286: 110102. doi: 10.1016/j.jfoodeng.2020.110102.

[15]   Hossain M S, Al-Hammadi M, Muhammad G. Automatic fruit classification using deep learning for industrial applications. IEEE Transactions on Industrial Informatics, 2018; 15(2): 1027–1034.

[16]   LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015; 521(7553): 436–444.

[17]   Nurvitadhi E, Sim J, Sheffield D, Mishra A, Krishnan S, Marr D. Accelerating recurrent neural networks in analytics servers: Comparison of FPGA, CPU, GPU, and ASIC. In: 2016 26th International Conference on Field Programmable Logic and Applications (FPL), Lausanne, Switzerland: IEEE, 2016; pp.1–4. doi: 10.1109/FPL.2016.7577314.

[18]   Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997; 9(8): 1735–1780.

[19]   Zhou F, Hang R L, Liu Q S, Yuan X T. Hyperspectral image classification using spectral-spatial LSTMs. Neurocomputing, 2019; 328: 39-47.

[20]   Schaffer C. Overfitting avoidance as bias. Machine Learning, 1993; 10(2): 153–178.

[21]   Qu Z, Mei J, Liu L, Zhou D Y. Crack detection of concrete pavement with cross-entropy loss function and improved VGG16 network model. IEEE Access, 2020; 8: 54564–54573.

[22]   Kingma D P, Ba J. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 Conference Track Proceedings. International Conference on Learning Representations, ICLR, 2017; arXiv: 1412.6980.

[23]   Farrell M D, Mersereau R M. On the impact of PCA dimension reduction for hyperspectral detection of difficult targets. IEEE Geoscience and Remote Sensing Letters, 2005; 2(2): 192–195.

[24]   Singh S, Kasana S S. Efficient classification of the hyperspectral images using deep learning. Multimedia Tools and Applications, 2018; 77(20): 61–74.

[25]   Mishra S, Datta-Gupta A. Applied statistical modeling and data analytics: A practical guide for the petroleum geosciences. Elsevier, 2018; 250p.

[26]   Qin J, Lu R. Detection of pits in tart cherries by hyperspectral transmission imaging. Transactions of the ASAE, 2005; 48(5): 1963–1970.

[27]   Belousov A I, Verzakov S A, von Frese J. Applicational aspects of support vector machines. Journal of Chemometrics, 2002; 16(8–10): 482–489.

[28]   Zhang R, Shen J, Wei F, Li X, Sangaiah A K. Medical image classification based on multi-scale non-negative sparse coding. Artificial Intelligence in Medicine, 2017; 83: 44–51.