CNN CLASSIFICATION OF SOYBEANS WITH STORAGE TIME BASED ON NEAR INFRARED SPECTROSCOPY

Yan He[1], Kai Kang[2], Qiwei Yin[1], Yang Peng[1], Wei Zhang[1*]

[1*]Corresponding author. Engineering College, Heilongjiang Bayi Agricultural University, China.
Email: zhang66wei@126.com | ORCID ID: https://orcid.org/0000-0001-6569-4945

KEYWORDS

near infrared spectroscopy, soybeans, storage time, CNN.

ABSTRACT

In soybeans from different years, the infrared spectroscopy waveforms exhibit similarities, yet variations in aging time lead to significant differences in their fatty acid content. To rapidly discern the year of soybean production, this study employed ten feature extraction algorithms in conjunction with Convolutional Neural Networks (CNN) to establish models for classifying soybeans from different years (2019, 2020, and 2021). The research findings reveal the following: (1) The CNN models, after feature extraction, all demonstrated improved accuracy. Notably, the optimal models for both powdered and granulated states were the Kpca-CNN models, achieving an accuracy of 100%. The corresponding loss function values were 0.0002 and 0.0007, with processing times of 0.19 seconds and 0.20 seconds, respectively. (2) The modeling results of all models suggest that the classification accuracy of soybean powder spectral data is higher compared to soybean particle spectral data. (3) Validation of the optimal Kpca-CNN model confirmed its consistent accuracy of 100% when introducing new data or reducing the size of the training dataset. In conclusion, the fusion of near-infrared spectroscopy analysis and CNN technology is considered an effective method for classifying soybeans from different years. This method provides a practical solution for the rapid and precise determination of seed production years.

INTRODUCTION

Soybean is one of the important sources of vegetable oil and is also processed into soy products and animal feed. It is a significant global economic crop and the largest imported agricultural product in China (Zhichao et al., 2021). As a high-fat and high-protein crop, soybeans are not easy to store and have a short lifespan, hence being regarded as short-lived seeds. The storage time significantly affects the seed viability of soybeans (Huifan et al., 2023), as the seed viability rapidly decreases with the extension of the storage years. In practical production, seeds with longer storage periods have issues such as low germination rate and low yield. Traditional methods for determining soybean crop year include empirical identification and chemical analysis (Lin et al., 2021). Chemical analysis methods can accurately detect the chemical composition and content of soybeans (Menglin et al., 2023). However, these methods are cumbersome to operate, can damage samples, pollute the

environment, are costly, and cannot be used for large-scale detection. However, these methods are cumbersome, can damage samples, pollute the environment, are costly, and cannot be used for large-scale detection. Therefore, there is an urgent need to find a scientifically effective and rapid method for identifying the storage year of soybeans.

Spectral analysis is a non-destructive testing method that offers advantages such as rapidity, non-destructiveness, and no pollution. It has been widely applied in seed testing (Kangmu et al., 2022). Li Juan, Huang Yawei, and others (Li et al., 2012) utilized full spectrum detection and established models for assessing rice freshness by analyzing the peak characteristics of spectra using the "first derivative + smoothing + least squares" processing method. This approach makes full use of effective spectral feature information but does not eliminate redundant features. Xie Ding and others (Xie et al., 2010) used principal component analysis and clustering to extract storage time-sensitive

spectral bands and achieved accurate identification of rice freshness. This method can effectively locate spectral features related to storage years but does not perform deep semantic feature extraction. Convolutional Neural Networks (CNN) is a novel artificial neural network method that combines artificial neural networks with deep learning techniques, possessing the ability for deep feature extraction and model representation (Kumar, 2022). Data and others (Emrah, 2022) improved the accuracy of maize seed classification using deep feature selection methods with CNN models, achieving a classification accuracy as high as 96.7%. Xin Zhao and others (Xin et al., 2022) utilized CNN for wheat seed variety classification, with an average classification accuracy of 96.4%. By combining spectral analysis with CNN, it is possible to effectively identify spectral features related to storage years and accurately classify seed storage years.

In summary, this study used soybeans of different years as experimental materials, collected near-infrared spectra of soybean powder and soybean granules, employed ten feature extraction algorithms for initial data dimension reduction, and utilized the CNN algorithm to extract deep features and establish a soybean crop year classification model. The reliability of the optimal model was tested by both increasing and decreasing the number of samples. This method combines the advantages of near-infrared spectroscopy and convolutional neural networks to effectively extract deep features related to storage time in soybean samples and achieve accurate classification. Additionally, even when the number of samples was decreased or increased, this method was able to maintain the reliability of the model.

The Impact of Storage Time on Soybeans

During the storage process of soybeans, the fat is prone to hydrolysis, leading to the occurrence of free fatty acids in the grains. The content of free fatty acids in grains is usually related to the fat content, moisture, temperature, storage time, and storage conditions of the grains (Statsenko Ekaterina et al., 2021). Under the same conditions, compared to other grains, soybeans show a rapid increase in fat content and the production of hydrolyzed fatty acids (Table 1). From Table 1, it can be observed that the values of free fatty acids in soybeans stored for one and two years are 11.64 and 16.42, respectively. However, after three years of storage, the content of free fatty acids in soybeans sharply rises to 21.59, indicating signs of deterioration. After four years of storage, the value further increases to 26.95, and the quality of soybeans significantly deteriorates, losing their characteristic luster and exhibiting peeling phenomenon.

TABLE 1. Changes in crude fat and fatty acid content in soybeans of different years

Storage Year	Crude Fat Content（%）	Free Fatty Acid Value（%）
1	21.06	11.64
2	20.19	16.42
3	20.14	21.59
4	20.02	26.95

The germination rate of soybeans is closely related to their edible quality (Anita et al., 2020). Soybeans that have lost their germination ability cannot be used for sprouting or tofu production, and their planting yield decreases (Chen et al., 2019). Generally, soybeans with normal moisture content have difficulty maintaining a normal germination rate when the temperature rises to 25°C, especially for yellow soybeans, which have loose seed coat tissue and strong metabolic activity. In the first four years of soybean storage, for every year extended, the germination rate decreases by 20.5% (Table 2). Therefore, the germination rate of soybeans is one of the indicators for assessing their freshness and edible quality (Yimiao et al., 2019).

TABLE 2. Changes in the germination rate of soybeans with different storage periods.

Storage Year	Germination Rate（%）
1	93
2	73
3	50
4	33

In summary, the hydrolysis of fat leads to an increase in free fatty acids in soybeans, and this process is closely related to the storage time of soybeans. Additionally, the germination rate of soybeans is directly related to their edible quality and serves as an important indicator for assessing the freshness and quality of soybeans. Based on these findings, this study aims to find a rapid method for identifying the storage period of soybeans, providing a theoretical basis for determining the crop year of other agricultural products. This method can be used to assess the quality and storage condition of agricultural products and assist farmers and consumers in making more accurate decisions.

MATERIAL AND METHODS

Instrument and equipment

TANGO desktop NIR spectrometer was used in the experiment, with a resolution of 8 cm^{-1} and a measurement range of 4000 ~ 11520 cm^{-1}, combined with OPUS software to obtain samples spectral curves. The model of the grinder used is SMF2002. The characteristic wavelengths extraction, modeling analysis and graph drawing are implemented in PyCharm 2019.2.4.

NIR spectra collection

The soybean seeds used in the research were sourced from Jianshan Farm in Heilongjiang. The main cultivated soybean varieties for three years were: Qirong 12, Jiuyan 13, Qirong 26, Mengdou 36, Beidou 37, Heihe 43, and Heike 60. The SMF2002 grinder model was used to grind the beans into powder and prepare the samples. Each sample was subjected to three spectral measurements, and then the average spectrum was taken as the representative spectrum. A total of 261 spectra were obtained for powdered soybeans, and 319 spectra were obtained for intact soybeans.
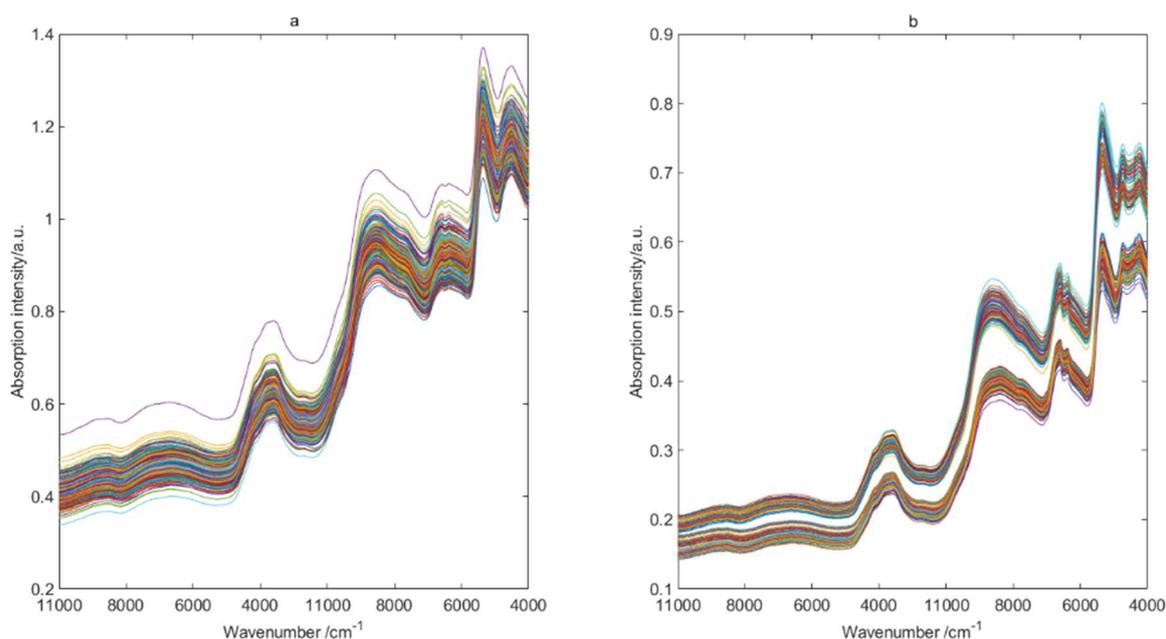
FIGURE 1. 319 original near-infrared spectra obtained from soybean granules (a) and 261 original near-infrared spectra obtained from soybean powder(b).

The overall absorption intensity of soybean granules' spectra is approximately 0.2 higher compared to soybean powder. This can be attributed to the different physical structures of soybean granules and powder. Soybean granules have larger volume and a more dense structure, leading to enhanced scattering and absorption of light, resulting in higher overall absorption intensity in the spectra. Additionally, soybean granules and powder exhibit variations in their chemical compositions. Granules may contain a higher concentration of light-absorbing substances such as proteins, starch, or other compounds, contributing to stronger light absorption and consequently higher absorption intensity in their spectra. While both groups show no significant differences in the corresponding positions of the peaks, the original spectra of soybean powder display clear classification into two categories. Compared to powder, granules have a more complete and uniform structure, with a more balanced distribution of chemical components overall. This results in more consistent spectral reflections and the absence of significant classification phenomena. This classification phenomenon can be attributed to the grinding process and the fragmented particle structure of soybean powder, which can cause significant variations in the fatty acid content between different crop years, leading to distinct spectral patterns. Hence, the classification phenomenon observed in soybean powder may be attributed to the differences in fatty acid content resulting from different crop years.

Spectra feature extraction

Due to the presence of background noise in the raw spectral data (Tong & Xiaochen, 2023), appropriate spectral feature extraction methods were considered to enhance the spectral features (Jian et al., 2023). In this paper, ten spectral feature extraction methods were used, namely PCA, KPCA, LDA, LPP, LLE, LE, MDS, SVD, t-SNE, and Losmap.

The first category is linear methods, which include PCA (Principal Component Analysis) (Liyuan et al., 2023) and LDA (Linear Discriminant Analysis) (Sadaghiyanfam & Kuntalp, 2018). PCA uses linear transformations to project the original data into a lowe-dimensional space, maximizing the variance of the projected data. LDA, on the other hand, is a supervised dimensionality reduction method that projects the data into a lower-dimensional space while maximizing the between-class distance and minimizing the within-class distance, thus improving classification performance. These methods are widely applied in data visualization, feature selection, and classification tasks.

The second category is kernel methods, with KPCA (Kernel Principal Component Analysis) as an example (Kai et al., 2023). KPCA performs a nonlinear transformation to project the data into a high-dimensional feature space, followed by PCA analysis. This approach utilizes kernel techniques to map the data into a high-dimensional space and can handle nonlinear relationships in the data. KPCA is particularly relevant in image processing, bioinformatics, and signal processing. A comparison was made among four kernel functions: linear, polynomial, radial basis function (RBF), and sigmoid (Fig. 2). Among these, the sigmoid kernel function exhibited the poorest clustering performance, while the other three kernel functions—linear, polynomial, and RBF—had relatively similar clustering effects. Therefore, in the comparison of classification accuracy, the RBF kernel function was selected due to its superior performance in terms of accuracy.
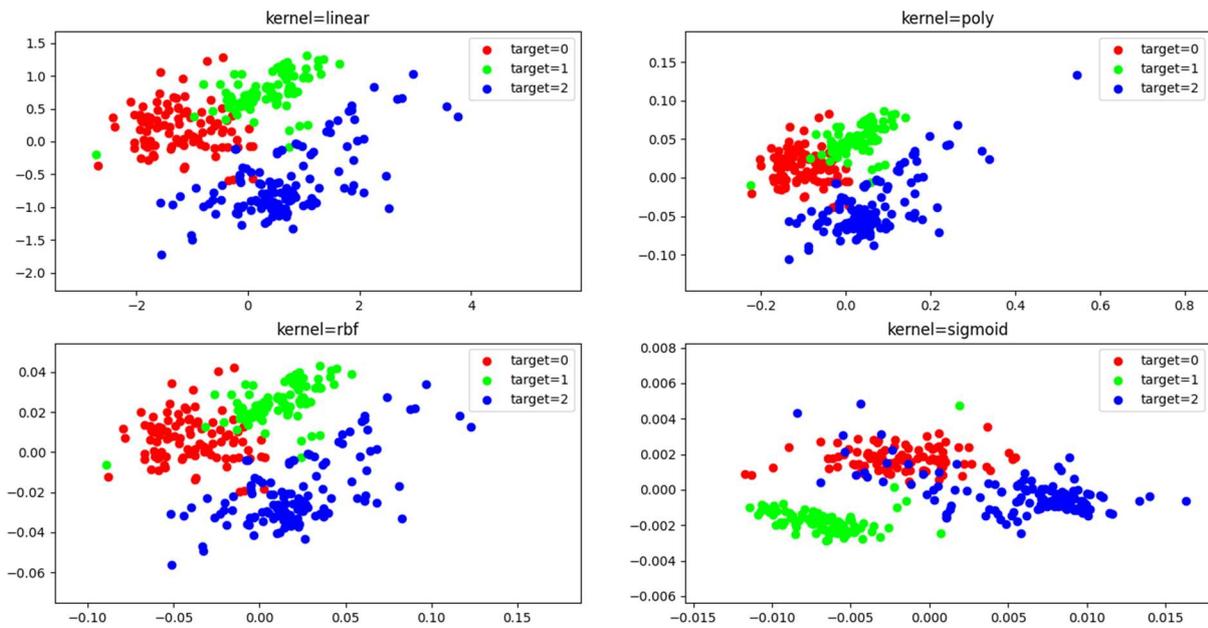
FIGURE 2. Clustering performance plots of the four kernel functions in KPCA: linear, polynomial, radial basis function (RBF), and sigmoid.

The third category is manifold learning methods. These methods consider the local structure among data points and project the data into a lower-dimensional manifold space. LPP (Locality Preserving Projection) achieves projection by preserving the local neighborhood relationships (Shu et al., 2023), while LLE (Locally Linear Embedding) embeds the data into a low-dimensional manifold space by preserving the neighborhood relationships (Liu et al., 2023). LE (Laplacian Eigenmaps) utilizes the Laplacian matrix to determine the low-dimensional representation of the data (John, 2023), preserving the smoothness of the data on the manifold. These methods find extensive applications in image recognition, pattern recognition, and data visualization.

The fourth category is nonlinear dimensionality reduction methods. MDS (Multidimensional Scaling) adjusts the positions of data points in a low-dimensional space by computing the distance matrix among them while preserving the relative distances of the original data (Anna et al., 2023). SVD (Singular Value Decomposition) decomposes the data matrix into singular values, extracting the principal components of the data (Xu et al., 2023). T-SNE (t-Distributed Stochastic Neighbor Embedding) maps high-dimensional data to a low-dimensional space by considering the similarities between data points (Shi et al., 2023), emphasizing the preservation of local structures. Losmap is a method similar to LLE, a locally linear embedding method used to discover local structures in high-dimensional data. These methods play important roles in nonlinear dimensionality reduction, data visualization, and high-dimensional data analysis.

TABLE 3. Results of feature extraction algorithms on near-infrared spectra of soybean granules and powder.

Algorithm	Soybean Granules	Soybean Powder
Raw	1845	1845
Kpca	145	170
Losmap	260	318
Lle	260	200
Tsne	200	319
Svd	261	200
Lda	200	311
Le	234	318
Lpp	258	319
Pca	261	200
Mds	200	200

From Table 3, it can be observed that different feature extraction algorithms yield varying numbers of features for soybean granules and powder in the spectral data. Different algorithms exhibit different effectiveness in capturing different aspects of the data. There are noticeable differences in feature dimensions between soybean granules and powder, particularly in the case of the Kpca algorithm, where the feature dimension for soybean granules is significantly lower than that for soybean powder. This discrepancy may be attributed to the distinct physical structures and spectral characteristics of the granules and powder. Therefore, selecting an appropriate feature extraction algorithm is crucial for accurately establishing models and interpreting the data.

Convolutional Neural Network

Convolutional Neural Network (CNN) is a type of deep learning algorithm that has been widely used for various computer vision tasks. CNN can automatically learn and extract features from spectral data. Through multiple convolutional layers, CNN can capture features of different scales and levels, including local textures, shapes, and spatial information (Zhou et al., 2022). This allows CNN to extract useful features from soybean NIR spectra for the year classification task. The deep architecture of CNN enables hierarchical learning (Bandy Adrian et al., 2023). By stacking multiple convolutional and pooling layers, CNN can gradually abstract and encode higher-level features (Jason et al., 2022). This hierarchical learning enables CNN to model complex patterns and relationships in the task of soybean year classification. CNN incorporates parameter sharing and weight sharing mechanisms, which allow the model to have shared feature representations across different locations and regions. For NIR spectral data, this means that CNN can learn similar feature representations at different positions in the spectrum. This sharing mechanism provides CNN with robustness to variations and noise in the spectral data (Bhuvaneshwari et al., 2022).

In this study, a CNN was employed to build a classifier for the soybean year classification task. Three convolutional layers were introduced, each consisting of 64, 128, and 128 convolutional filters, respectively (Fig. 3). The size of these filters was set to 3. The selection of the number and size of these filters was based on preliminary data analysis and experimentation. In the convolutional layers, the Rectified Linear Unit (ReLU) activation function was used to enhance the model's feature extraction capability. Following the convolutional layers, three MaxPooling layers were introduced to reduce the spatial dimensions of the feature maps. Pooling operations help reduce computational complexity, control overfitting, and enhance the model's translational invariance.
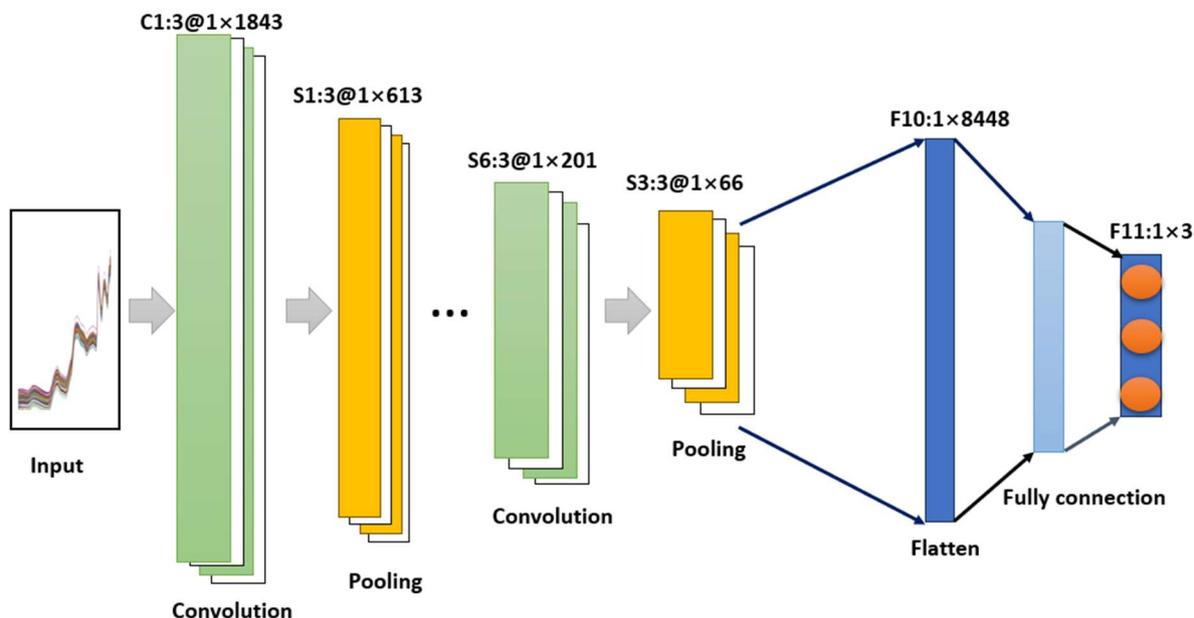

FIGURE 3. The designed 1D-CNN model structure.

Next, a Flatten layer was added to transform the multidimensional data into a one-dimensional vector, facilitating its input into subsequent fully connected layers. This step converted the data from three-dimensional feature maps to a one-dimensional representation. To mitigate overfitting, a Dropout layer was included before the fully connected layers, with a dropout rate set to 0.5. The Dropout layer randomly sets a portion of the neuron outputs to zero, thereby reducing dependencies among neurons and improving the model's generalization ability.

A fully connected layer with three output nodes was added, and the softmax activation function was used for classification predictions. This fully connected layer made classification decisions based on the features learned from the convolutional and pooling layers. Regarding parameter selection, the learning rate was carefully adjusted, and a smaller value (lr=0.01) was chosen to ensure stable convergence during the model training process. The Adam optimizer was employed, and the relevant parameters (beta1=0.9, beta2=0.99, epsilon=1e-08) were set to enhance the optimizer's performance and stability.

For the loss function, mean squared logarithmic error was chosen, as it effectively handles larger differences between predicted and true values and maintains numerical stability when dealing with non-negative data. The optimal parameter combination was selected through cross-validation and performance evaluation on a validation set to ensure good accuracy and robustness of the model in the soybean year classification task.
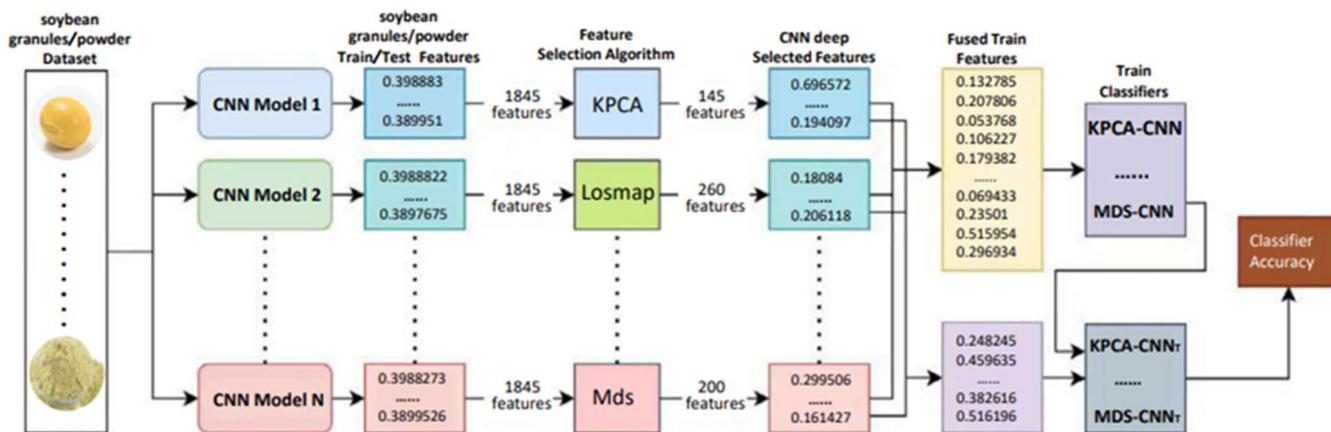
FIGURE 4. The principle diagram depicts the utilization of 10 feature extraction algorithms to extract near-infrared spectroscopic features from soybean particles and soybean powder. These extracted features are then combined with CNN to extract deep-level features, thereby establishing a soybean classification model for different years.

Therefore, this study utilizes 10 feature extraction algorithms to extract near-infrared spectroscopic features from soybean particles and soybean powder. These features are combined with CNN to extract deep features and establish a soybean classification model for different years. The principle of this approach is illustrated in Fig. 4.

RESULTS AND DISCUSSION

Establishing model for soybean granules

A dataset containing 319 spectral data points was divided into training and prediction subsets. The training set comprises 259 samples, while the prediction set comprises 60 samples (20 samples per year category). Utilizing the complete spectral dataset, a Raw+CNN model was established, along with 10 CNN discriminant analysis models that underwent processing with ten different feature extraction algorithms. The KerasClassifier was employed to execute the model training process. Each training batch consisted of 64 samples, and the training was conducted for a total of 100 epochs. At the conclusion of each training epoch, the accuracy (Fig.5) and loss (Fig.6) were recorded for subsequent analysis and visualization.
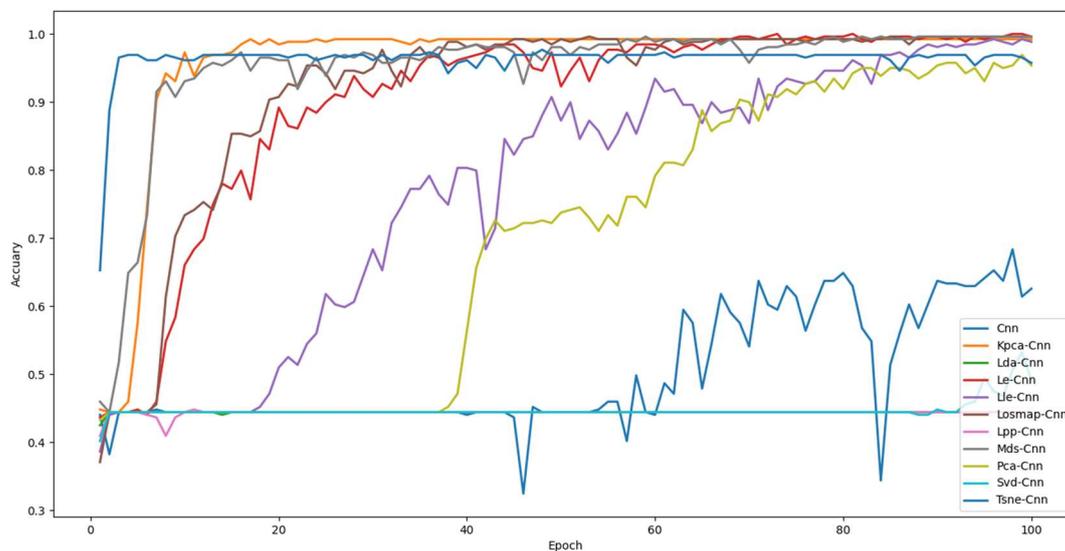

FIGURE 5. The training accuracy of 11 CNN soybean particle classification models.
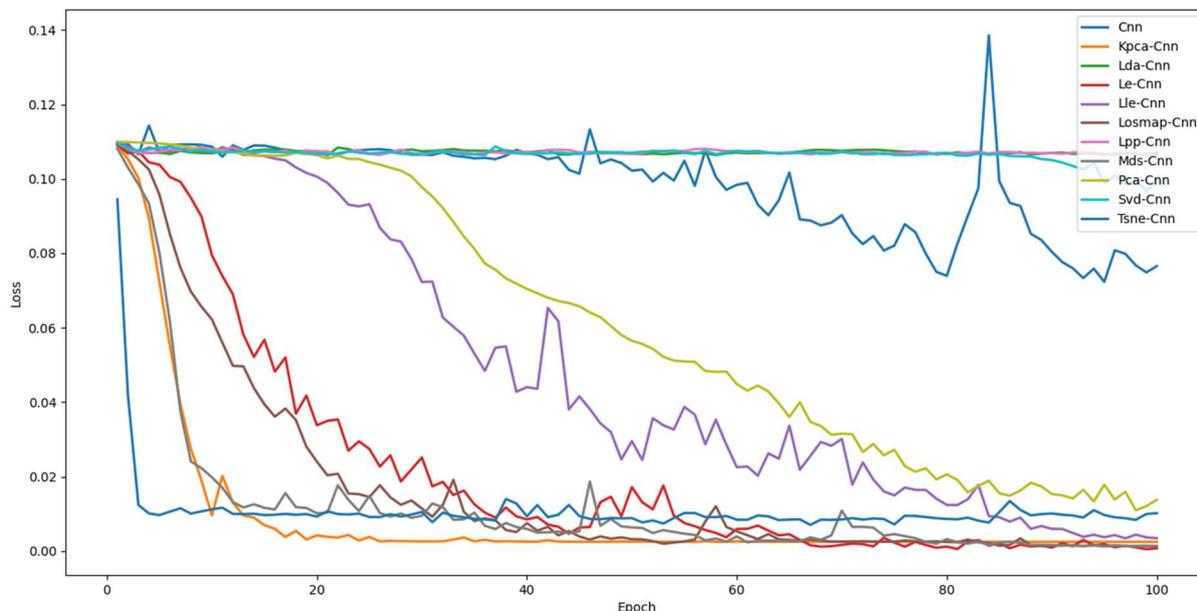
FIGURE 6. The training loss rates of 11 CNN soybean particle classification models.

From the observations in the graph, it can be seen that the CNN, Lpp-CNN, and Svd-CNN models exhibit relatively stable training accuracy and loss rates, maintaining around 0.45 and 0.11, respectively. In contrast, the Tsne-CNN model shows significant fluctuations, with a maximum training accuracy of 0.67, which is relatively low. Regarding the Pca-CNN and Lle-CNN models, the improvement in training accuracy is relatively gradual, accompanied by slight fluctuations. The training accuracy of these two models starts to increase around the 38th and 18th epochs, respectively, and continues to fluctuate before reaching their peaks. The Pca-CNN model achieves a maximum accuracy of 0.99, while the highest accuracy of the Lle-CNN model is 0.93. Both the Le-CNN and Losmap-CNN models rapidly increase their training accuracy from the 8th epoch and reach peaks of 0.99 by the

40th epoch. In comparison, the training accuracy and loss rates of the Kpca-CNN, Mds-CNN, and Tsne-CNN models experience rapid initial increases and decreases. Notably, the Kpca-CNN model demonstrates the highest stability. It achieves a maximum training accuracy of 1 and a minimum loss rate of 0.0007.

Establish model for soybean powder

Dividing a dataset of 261 spectral data into training and prediction sets, the training set encompasses 201 samples, while the prediction set comprises 60 samples (20 samples per each annual category). Establishing the Raw+CNN model using the entire spectral data and the CNN discriminant analysis model processed with ten different feature extraction algorithms, the training results are shown in the figure.
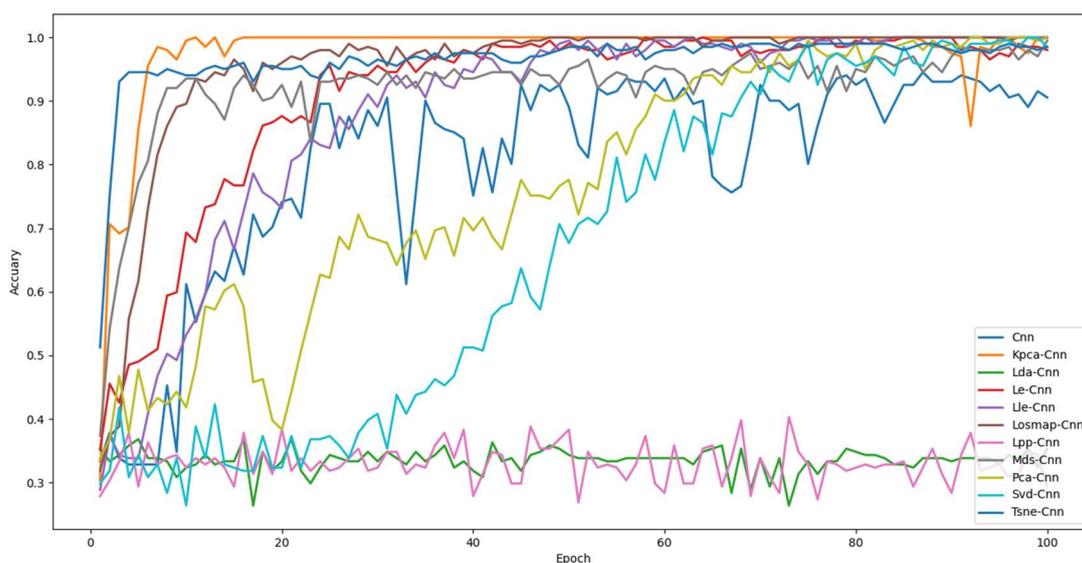

FIGURE 7. The training accuracy of 11 CNN soybean particle classification models.
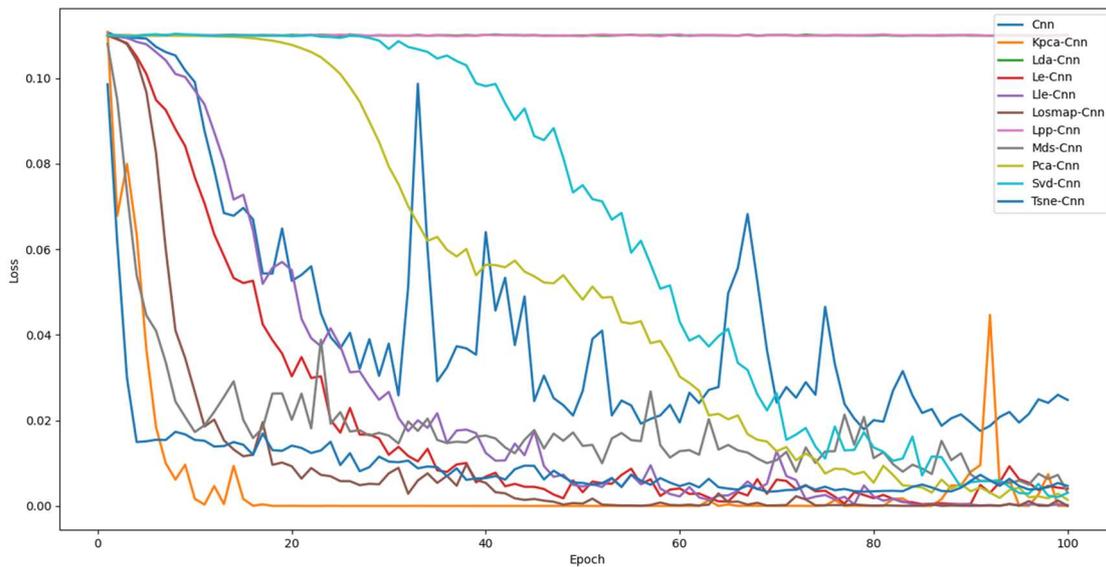
FIGURE 8. The training loss rates of 11 CNN soybean particle classification models.

Overall, when compared to the soybean particle CNN model, the soybean powder CNN model exhibits larger fluctuations in training accuracy and loss rates. In the case of the Lda-CNN and Lpp-CNN models, the training accuracy is relatively low, fluctuating between 0.2 and 0.4, while the loss rate remains relatively high at 0.11 with minimal fluctuations. In contrast, the CNN model demonstrates pronounced fluctuations, reaching a maximum training accuracy of 0.91. For the Pca-CNN and Svd-CNN models, the improvement in training accuracy is gradual, accompanied by slight fluctuations. The training accuracy of these two models starts to increase around the 20th epoch and continues to fluctuate before reaching a peak of 0.99. The Le-CNN and Lle-CNN models also exhibit slow improvement, reaching peaks of 0.99 by the 50th epoch. In comparison to the aforementioned models, the Kpca-CNN, Losmap-CNN, Mds-CNN, and Tsne-CNN models experience rapid initial increases and decreases in training accuracy and loss rates. Notably, the Kpca-CNN model demonstrates the highest stability, reaching its peak accuracy by the 8th epoch with a maximum value of 1, and the lowest loss rate of only 0.0002.

Comparison and Analysis between Soybean Particle and Powder Models

Eleven distinct CNN models were employed to predict soybeans from different years. To accomplish this, a dataset consisting of 60 samples (20 per year) was utilized. The objective was to assess the stability of the models, as shown in Tables 5 and 6. In general, the CNN models for soybean powder exhibited varying degrees of improvement compared to those for soybean particles. The Svd-CNN, Lpp-CNN, and Lda-CNN models yielded the lowest prediction accuracies of 0.33 for both soybean conditions. The CNN model's prediction accuracy increased by 0.33 to 0.66. The Le-CNN and Lle-CNN models saw improvements of 0.11 and 0.03 in their prediction accuracies, respectively. Among them, the Kpca-CNN, Pca-CNN, Tsne-CNN, Losmap-CNN, and Mds-CNN models achieved prediction accuracies of over 0.91 for soybean particles. However, in the case of soybean powder, only the Tsne-CNN model fell short of an accuracy of 1, achieving a value of 0.95. The Kpca-CNN model attained prediction accuracies of 1 for both soybean conditions, while in the case of soybean powder, it managed to reduce prediction time by 0.01 seconds. Considering the overall performance of all models, the Kpca-CNN model showcased the highest accuracy while also achieving the shortest prediction time.

TABLE 4. Prediction results of near-infrared soybean particle CNN model.

Model	Accuracy(%)	Time(s)
CNN	0.33	0.71
Losmap-CNN	0.96	0.28
Lle-CNN	0.77	0.21
Tsne-CNN	0.91	0.28
Svd-CNN	0.33	0.29
Lda-CNN	0.33	0.31
Le-CNN	0.75	0.37
Lpp-CNN	0.33	0.29
Pca-CNN	0.96	0.31
Mds-CNN	0.93	0.21
Kpca-CNN	1	0.20

TABLE 5. Prediction results of near-infrared soybean powder CNN model.

Model	Accuracy(%)	Time(s)
Cnn	0.66	0.74
Losmap-cnn	1	0.29
Lle-cnn	0.88	0.23
Tsne-cnn	0.95	0.27
Svd-cnn	0.33	0.33
Lda-cnn	0.33	0.29
Le-cnn	0.78	0.31
Lpp-cnn	0.33	0.30
Pca-cnn	1	0.33
Mds-cnn	1	0.38
Kpca-cnn	1	0.19

Validation of the optimal model

To validate the optimized soybean powder model, Kpca-CNN, two approaches were employed. The first approach involved augmenting the existing prediction dataset with newly measured spectral data for each year to assess the model's stability (Table 6). The second approach maintained a fixed prediction dataset size of 60 samples while gradually reducing the number of samples in the training dataset, aiming to evaluate the model's accuracy performance under different training dataset sizes (Table 7).

TABLE 6. Translation of the impact of increasing prediction dataset size on Kpca-CNN model.

Model	Test Set Size	Accuracy(%)	Time(s)
Kpca-CNN	90	1	0.26
Kpca-CNN	120	1	0.23
Kpca-CNN	150	1	0.23

TABLE 7. Translation of the impact of reducing training dataset size on Kpca-CNN model.

Model	Training Set Size	Accuracy(%)	Time(s)
Kpca-CNN	171	1	0.26
Kpca-CNN	141	1	0.22
Kpca-CNN	111	1	0.20
Kpca-CNN	81	1	0.20
Kpca-CNN	51	1	0.17

From the table, it can be observed that with the introduction of new data, the accuracy of the Kpca-CNN model remains unchanged at 1. Even when the training dataset size is reduced to 51 samples, the accuracy still remains at 1, thereby solidly confirming the model's stability. This algorithm is not only applicable to soybean year detection tasks but also serves as a valuable reference for other agricultural crop classification tasks.

CONCLUSIONS

This study has achieved accurate identification of soybean storage years through the utilization of near-infrared spectroscopy and Convolutional Neural Networks (CNN) methods. Near-infrared spectral data from soybean particles and powder samples of various years were collected, and 10 feature extraction algorithms were employed to achieve data dimension reduction and feature enhancement. Subsequently, a CNN model was established to construct classifiers for soybean year classification. Throughout multiple rounds of training and validation, the Kpca-CNN model demonstrated the highest stability and accuracy. It not only attained a prediction accuracy of 1 for both powdered and intact soybeans but also maintained its performance stability when introducing new data or reducing the size of the training dataset. This achievement underscores the successful extraction of deep features related to storage years by fully capitalizing on the advantages of near-infrared spectroscopy and CNNs. It provides a robust tool for the accurate classification of soybean storage years.

REFERENCES

Anita Z, Małgorzata K, Magdalena T (2020) A comparison of the composition and contamination of soybean cultivated in Europe and limitation of raw soy seed content in weaned pigs' diets. Animals 10(11). https://doi.org/10.3390/ani10111972

Anna L, Yuying X, Qiang S (2023) An analysis of classical multidimensional scaling with applications to clustering. Information and Inference 12(1). https://doi.org/10.1093/IMAIAI/IAAC004

Bandy Adrian D, Spyridis Y, Villarini B, Argyriou V (2023) Intraclass clustering-based CNN approach for detection of malignant melanoma. Sensors (2). https://doi.org/10.3390/S23020926

Bhuvaneshwari P, Nagaraja Rao A, Harold Robinson Y (2022) Correction to: Top-N recommendation system using explicit feedback and outer product based residual CNN. Wireless Personal Communications (2). https://doi.org/10.1007/S11277-022-10055-Y

Chen S, Yang Charles T, Downs Melanie L (2019) Detection of six commercially processed soy ingredients in an incurred food matrix using parallel reaction monitoring. Journal of Proteome Research (3). https://doi.org/10.1021/acs.jproteome.8b00689

Emrah D (2022) Enhancing classification capacity of CNN models with deep feature selection and fusion: a case study on maize seed classification. Data & Knowledge Engineering 141. https://doi.org/10.1016/j.datak.2022.102075

Huifan L, Shanshan Z, Yanxia G (2023) Effect of storage time on the volatile compounds and taste quality of Meixian green tea. LWT 173. https://doi.org/10.1016/j.lwt.2022.114320

Jason B, Zahra G, Nawin R (2022) CNN based image classification of malicious UAVs. Applied Sciences 13(1). https://doi.org/10.3390/APP13010240

Jian W, Zhongwei L, Junfang Y (2023) A multilevel spatial and spectral feature extraction network for marine oil spill monitoring using airborne hyperspectral image. Remote Sensing 15(5). https://doi.org/10.3390/rs15051302

John AB (2023) Bisecting for selecting: using a Laplacian eigenmaps clustering approach to create the new European Football Super League. Mathematics 11(3). https://doi.org/10.3390/MATH11030720

Kai Z, Ke Z, Rui B (2023) Prediction of gas explosion pressures: a machine learning algorithm based on KPCA and an optimized LSSVM. Journal of Loss Prevention in the Process Industries 83. https://doi.org/10.1016/j.jlp.2023.105082

Kangmu M, Shiying W, Shixing H (2022) Myocardial infarct border demarcation by dual-wavelength photoacoustic spectral analysis. Photoacoustics 26 (prepublish). https://doi.org/10.1016/j.pacs.2022.100344

Kumar KR (2022) A novel approach to unsupervised pattern discovery in speech using Convolutional Neural Network. Computer Speech & Language 71. https://doi.org/10.1016/j.csl.2021.101259

Li J, Li Z, Fu X (2012) Research on rapid non-destructive detection of rice freshness using near-infrared spectroscopy. Spectroscopy and Spectral Analysis 32(08): 2126-2130.

Lin L, Hui H, Daize W (2021) Rheological and textural properties of acid-induced soybean protein isolate gel in the presence of soybean protein isolate hydrolysates or their glycosylated products. Food Chemistry 360. https://doi.org/10.1016/J.FOODCHEM.2021.129991

Liu S, Chen Z, Jiao F (2023) Detection of maize seed germination rate based on improved locally linear embedding. Computers and Electronics in Agriculture. https://doi.org/10.1016/J.COMPAG.2022.107514

Liyuan Y, Kuan Z, Zehua C (2023) Fault diagnosis of WOA-SVM high voltage circuit breaker based on PCA Principal Component Analysis. Energy Reports 9(S8). https://doi.org/10.1016/j.egyr.2023.04.341

Menglin G, Fugang X, Borui W (2023) Study on detection of soybean components in edible oil with ladder-shape melting temperature isothermal amplification (LMTIA) assay. Analytical Methods 15:581-586. https://doi.org/10.1039/D2AY01719A

Sadaghiyanfam S, Kuntalp M (2018) Comparing the performances of PCA (Principle Component Analysis) and LDA (Linear Discriminant Analysis) transformations on PAF (Paroxysmal Atrial Fibrillation) patient detection. Biomedical Imaging, Signal Processing 2018. https://doi.org/10.1145/3288200.3288201

Shi S, Xu Y, Xu Xi, Mo X, Ding J (2023) A preprocessing manifold learning strategy based on t-distributed stochastic neighbor embedding. Entropy (Basel, Switzerland) (7).

Shu L, Zhengguang C, Feng J (2023) Detection of maize seed germination rate based on improved locally linear embedding. Computers and Electronics in Agriculture 204:107514. https://doi.org/10.1016/J.COMPAG.2022.107514

Statsenko Ekaterina S, Korneva Nadezhda Yu, Pokotilo Olesya V, Litvinenko Oksana V (2021) Development of technology for producing wheat bread enriched with soy ingredient. Food science and technology international (2). https://doi.org/10.1177/10820132211062991

Tong L, Xiaochen Y (2023) Paralinguistic and spectral feature extraction for speech emotion classification using machine learning techniques. EURASIP Journal on Audio, Speech, and Music Processing 2023(1). https://doi.org/10.1186/s13636-023-00290-x

Xie D, Wei Y, Ouyang J (2010) Rapid discrimination of storage time for indica rice using principal component discriminant analysis of near-infrared spectroscopy. Grain Science and Technology and Economy 35(05): 41-42.

Xin Z, Haotian Q, Xiulan S (2022) Hybrid convolutional network based on hyperspectral imaging for wheat seed varieties classification. Infrared Physics and Technology 125. https://doi.org/10.1016/j.infrared.2022.104270

Xu W, Tan L, Lin R (2023) Weighted singular value decomposition basis of Szegő kernel and its applications to signal reconstruction and denoising. Journal of Computational and Applied Mathematics. https://doi.org/10.1016/J.CAM.2023.115067

Yimiao X, Fusheng C, Kunlun L, Lifen Z, Xiaojie D, Xin Z, Zhenya Z (2019) Compositional differences between conventional Chinese and genetically modified Roundup Ready soybeans. Crop and Pasture Science (6). https://doi.org/10.1071/cp19006

Zhichao Z, Xiaodan M, Haiou G (2021) A method for calculating the leaf inclination of soybean canopy based on 3D point clouds. International Journal of Remote Sensing 42(15). https://doi.org/10.1080/01431161.2021.1930271

Zhou L, Nandal A, Ganchev T (2022) Breast cancer detection by fusion of deep features with CNN extracted features. International Journal of Intelligent Systems Technologies and Applications 20(6). https://doi.org/10.1504/IJISTA.2022.10053604