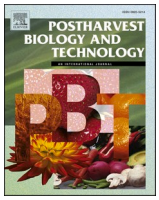
Postharvest Biology and Technology 213 (2024) 112960

Contents lists available at ScienceDirect

Postharvest Biology and Technology

journal homepage: www.elsevier.com/locate/postharvbio



Dual-aspect attention spatial-spectral transformer and hyperspectral imaging: A novel approach to detecting *Aspergillus flavus* contamination in peanut kernels

Zhen Guo [a,b,c], Jing Zhang [a], Haifang Wang [d], Shiling Li [a,b,c], Xijun Shao [a,b,c], Haowei Dong [a,b,c], Jiashuai Sun [a,b,c], Lingjun Geng [a,b,c], Qi Zhang [e], Yemin Guo [a,b,c,*], Xia Sun [a,b,c,*], Lianming Xia [a,b,c], Ibrahim A. Darwish [f]

[a] School of Agricultural Engineering and Food Science, Shandong University of Technology, No. 266 Xincun Xilu, Zibo, Shandong 255049, China
[b] Shandong Provincial Engineering Research Center of Vegetable Safety and Quality Traceability, No. 266 Xincun Xilu, Zibo, Shandong 255049, China
[c] Zibo City Key Laboratory of Agricultural Product Safety Traceability, No. 266 Xincun Xilu, Zibo, Shandong 255049, China
[d] Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China
[e] Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China
[f] Department of Pharmaceutical Chemistry, College of Pharmacy, King Saud University P.O. Box 2457, Riyadh 11451, Saudi Arabia

ARTICLE INFO

Keywords:
Attention fusion mechanisms
Deep Learning
Convolutional neural network
Graph convolutional network
Peanut kernels

ABSTRACT

In this study, an innovative dual-aspect attention spatial-spectral transformer (DAASST) was introduced to advance postharvest quality control by the detection of *Aspergillus flavus* contamination and the accurate identification of contamination times in peanut kernels. The critical importance of maintaining postharvest quality and safety in nuts was recognized, with hyperspectral imaging technology being leveraged due to its great potential in non-destructive testing and quality assessment of nuts. At the heart of DAASST's innovation, an enhanced transformer architecture that incorporated an attention fusion mechanism was employed for the effective integration of the extracted features. This sophisticated integration not only improved the model's performance but also was significantly surpassed by the capabilities of traditional machine learning methods in the context of postharvest biology and technology. Exceptional accuracy was demonstrated in testing, with 99.40% achieved in detecting *Aspergillus flavus* contamination and a remarkable 100% in distinguishing between different contamination times. Significant contributions to the field of postharvest biology and technology were made by merging cutting-edge feature extraction techniques, attention mechanisms, and transformer architecture to refine hyperspectral image analysis for postharvest quality control. The proven effectiveness of the DAASST in accurately detecting *Aspergillus flavus* and determining contamination times in peanut kernels highlighted its potential as a valuable tool for ensuring the safety and quality of postharvest nuts.

1. Introduction

Peanut kernels are rich in protein, fat, vitamins and dietary fiber, which represent valuable nutritional resources suitable for direct consumption or processing into peanut butter and edible oil (Sun et al., 2020). However, these kernels are susceptible to contamination by toxigenic fungi including *Aspergillus flavus* and *Aspergillus parasiticus* after harvest (Tao et al., 2020). These fungi are known as producers of aflatoxin B_1 (AFB_1), which is a highly toxic and carcinogenic substance (Chu et al., 2017). Consequently, a critical task involves identifying and selecting peanut kernels that are similarly contaminated to prevent AFB_1 from entering the food chain (Qiao et al., 2017). Investigation into alterations in the tissue structure and chemical composition of *Aspergillus flavus*-contaminated peanut kernels facilitates the discrimination between healthy peanut kernels and contaminated peanut kernels.

Hyperspectral imaging integrates imaging with spectroscopy to simultaneously furnish physical and geometrical features, along with the chemical compositions of the product through spectral analysis

* Corresponding authors at: School of Agricultural Engineering and Food Science, Shandong University of Technology, No. 266 Xincun Xilu, Zibo, Shandong 255049, China.
E-mail addresses: gym@sdut.edu.cn (Y. Guo), sunxia2151@sina.com (X. Sun).

https://doi.org/10.1016/j.postharvbio.2024.112960
Received 3 March 2024; Received in revised form 9 April 2024; Accepted 9 April 2024

(Feng et al., 2018). Targeted regions within the samples are selected from the hyperspectral images, and the corresponding spectra are extracted for subsequent chemometric calibrations (Femenias et al., 2022). Hyperspectral imaging has been widely used to identify fungal contamination in maize kernels and wheat grains. One research reports the inoculation of sterilized maize kernels with *Aspergillus parasiticus* daily from day 1 to day 7, while non-inoculated sterilized kernels serve as controls. Experimental results substantiate the feasibility of utilizing near-infrared hyperspectral imaging to verify the detection of *Aspergillus parasiticus* contamination in maize kernels, and a support vector machine (SVM) model achieves classification accuracies of 97.92% and 91.67% on the calibration dataset and validation dataset, respectively (Zhao et al., 2017). Another study employs near-infrared hyperspectral imaging (895–1728 nm) to detect *Fusarium*-damaged wheat grains, achieving an 85.8% classification accuracy using an artificial neural network model (Femenias et al., 2020). Additionally, a feature pre-extraction method for hyperspectral data and a multi-feature fusion block are employed to detect moldy peanuts among 1066 samples, and then eventually the model is constructed with the highest average accuracy of 92.07% (Liu et al., 2020). These investigations collectively demonstrate the feasibility and challenges of using hyperspectral imaging for identifying *Aspergillus*-contaminated peanut kernels, emphasizing the difficulty in achieving 100% accuracy in identifying contaminated kernels. Furthermore, few studies explore the discrimination of contamination time, making it challenging to assess the degree of kernel moldiness and implement proper treatment.

One potential reason for the inability to accurately identify contaminated peanut kernels may be attributed to insufficient extraction of information from hyperspectral images. Retaining key information and eliminating redundant details has been a focal point of related research. Most studies have focused on exploring spectral information within hyperspectral images, extracting average spectral information from the region of interest (ROI) in the samples. Subsequently, chemometric methods are applied to preprocess the spectral data, extract feature wavelengths, and establish discriminant analysis models (Wu & Sun, 2013). A minority of studies simultaneously consider image information and spectral information of hyperspectral images. In one study, texture features and spectral features are combined to differentiate AFB_1-contaminated peanuts, with a SVM model achieving accuracies of 93% and 94% on the calibration set and validation set, respectively (He et al., 2021). Another research integrates spectral information and spatial information for identifying fungal-contaminated peanuts (Qiao et al., 2017). In recent years, deep learning algorithms have provided additional methods for fully exploiting hyperspectral image information. A 1-dimensional convolutional neural network (CNN) is employed for pixel-level classification of aflatoxin contamination (Gao et al., 2021), and CNN-based pixel-spectral reshaping methods are also used to assess aflatoxin contamination in peanut kernels (Han, Gao, 2019). In a recent study, a 3-dimensional-convolutional neural network (3D-CNN) and a 2-dimensional-convolutional neural network (2D-CNN) are combined to simultaneously capture spectral features and spatial features of the hyperspectral images, in which Gaussian-weighted feature tokenizers are used to convert extracted features, achieving an overall accuracy of 97% in identifying the growth years of Kudzu root (Xu et al., 2023). The results highlight the unique advantages of combining 3D-CNN and 2D-CNN for extracting spatial-spectral features from hyperspectral images. Additionally, transformers are applied in this study, which are effective deep neural networks previously established for natural language processing tasks (Islam et al., 2024). Since the establishment of Vision-Transformers, transformers have been proven to be effective alternatives to CNNs in various tasks such as image recognition and object detection (Azad et al., 2024). Combining transformers and hyperspectral imaging research for detecting *Aspergillus flavus* and determining contamination times in peanut kernels will show promising results.

Therefore, the objectives of this study are as follows, (1) thoroughly explore the hyperspectral image information of peanut kernels, achieving accurate identification of *Aspergillus flavus* contamination and differentiation of contamination time, (2) design multi-scale 3D-CNN combined with 2D-CNN sub-networks, capturing both spatial features and spectral features of individual peanut kernels' hyperspectral images, (3) design unique Laplacian-weighted (LW) feature tokenizers and graph convolutional network (GCN) feature extractors, facilitating a more nuanced and effective representation of the peanut kernels' characteristics, (4) innovatively design a dual-aspect attention spatial-spectral transformer (DAASST), incorporating attention-weighted mechanisms to merge 2 high-level features, enhancing expressive and generalization capabilities for classification tasks.

2. Materials and methods

2.1. Peanut kernel samples preparation

Peanut kernels of the 'Weihua' and 'Baisha' varieties were sourced from a Zibo-based supermarket in China. The peanut (Arachis hypogaea) varieties 'Weihua' and 'Baisha' were identified as belonging to the Fabaceae family, genus Arachis. The selection of peanut kernels from these varieties was based on uniform size, maturity, and the absence of disease or pest damage to guarantee the repeatability of results and eliminate data variability. A total of 3360 kernels were equally distributed into a contaminated group and a control group. Sterilization involved immersing all samples in a 75% ethanol solution for 1 min, followed by triple rinsing in sterile water and subsequent placement in a sterile environment. *Aspergillus flavus* (ATCC#28539), acquired from China's National Strain Center, was grown on PDA medium at 28 °C for 5 d. Spores from this culture were then collected and adjusted to a concentration of 1×10^6 CFU/mL using sterile water, and used for the contamination of the peanut kernels. Peanut kernels inoculated with sterile water were used as the control group. These kernels were incubated until day 7 under conditions set to 30 °C and 85% relative humidity (Long et al., 2022). For hyperspectral imaging, 240 kernels from each group were selected and imaged daily from day 1 to day 7.

2.2. Acquisition and processing of hyperspectral images of peanut kernels

The hyperspectral images were acquired using a short-wave infrared hyperspectral imaging system (Isuzu Optics Corp., Taiwan, China). To mitigate the effects of ambient light, the system components were housed in a darkened box. The imaging process involved setting the exposure time to 2.8 ms and the mobile platform speed to 14.5 mm/s. Both black reference image and white reference image were utilized to calibrate the raw hyperspectral images, reducing the ambient noise impact (Guo et al., 2023). Each hyperspectral image contained 60 peanut kernels, yielding a total of 56 images, split evenly between the contaminated group and the control group.

To prevent movement during scanning, peanut kernels were affixed to black paperboard using scotch tape. A range of complex image processing techniques was employed to accurately extract the ROI of individual kernels. Initially, a hyperspectral pseudocolor image of peanut kernels containing 288 spectral bands (1000–2500 nm) was obtained (Fig. 1A). The spectral curves of peanut kernels were compared to 2 backgrounds (paperboard and scotch tape) (Fig. 1B). It revealed that at 1133.1 nm (Fig. 1C), the kernels' reflectance was markedly higher than that at 1936.0 nm (Fig. 1D), a distinction not observed in the backgrounds. Consequently, a ratio operation on the gray images of these 2 wavelengths (1133.1 nm and 1936.0 nm) generated the band ratio image (Fig. 1E). Subsequently, a median filter with a kernel size of 9 was applied to the band ratio image, reducing image noise without significantly blurring the edges. This was followed by 2 rounds of morphological operations (erosion and dilation), with square sizes of 5 and 3 respectively, to remove small objects and gaps, further refining the image quality.

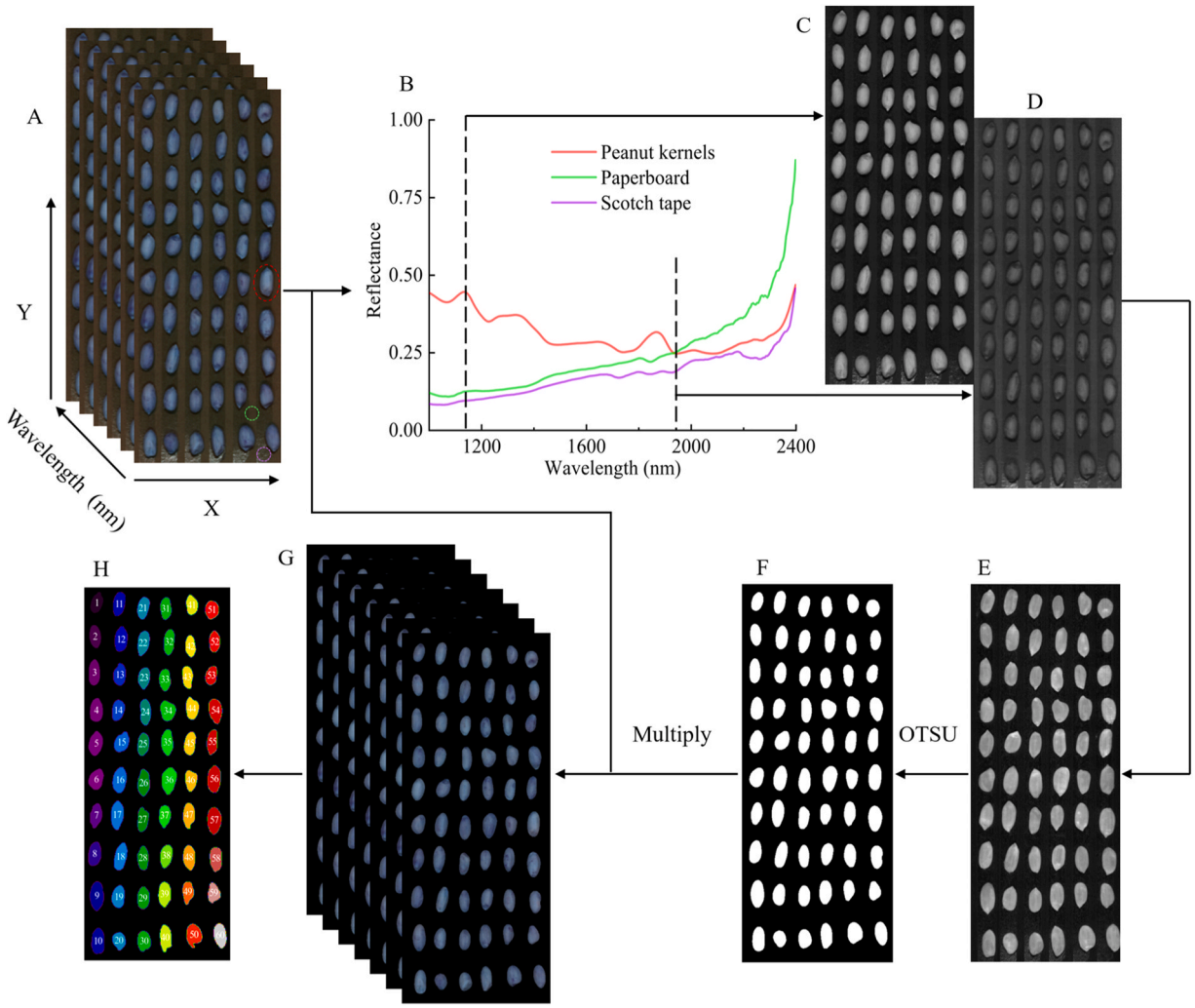


Fig. 1. Hyperspectral image processing, (A) hyperspectral pseudocolor image of peanut kernels, (B) spectral curves of peanut kernels and backgrounds, (C) 1133.1 nm gray image, (D) 1936.0 nm gray image, (E) band ratio image, (F) mask image, (G) mask hyperspectral image, (H) peanut kernels numbered image.

The Otsu's thresholding method (OTSU) was then applied to determine the optimal segmentation threshold, which was used on the cleaned ratio image to create a mask image (Fig. 1F) (Huang et al., 2021). In this mask image, the background and peanut kernels were represented by 0 pixel value and 1 pixel value, respectively. This mask image was multiplied by each band in the raw hyperspectral image, resulting in a mask hyperspectral image showing only the peanut kernels (Fig. 1G). Finally, the peanut kernels were numbered according to their ROI centroids (Fig. 1H), arranged from top to bottom and left to right, enabling precise location information of each kernel.

2.3. Individual peanut kernel hyperspectral image extraction and principal component analysis

In order to obtain individual peanut kernel hyperspectral images, it was imperative to precisely segment each peanut sample from the hyperspectral image. This necessitated the calculation of each peanut's pixel size and the dimensions of its minimum bounding rectangle. Table 1 indicated that there were size differences between the 'Weihua' and 'Baisha' peanut varieties. The 'Weihua's average pixel size of 1178.15 was larger than that of 'Baisha's 918.60. Moreover, the average height of 'Weihua' peanut kernels was greater than that of 'Baisha', although their average widths were similar. The average height and the average width of the smallest bounding rectangles of the ROI for all peanut samples were 47.32 pixel and 29.14 pixel, respectively. To facilitate the extraction of spatial-spectral features of individual peanut samples, the size was set at 50×25×288 (Fig. 2A).

Table 1
Peanut sample ROI size.

Sample	Index	Mean	Max.	Min.	Std.
'Weihua'	Pixel size	1178.15	1979	405	220.70
	Height	53.96	74	25	6.78
	Width	28.70	47	15	3.34
'Baisha'	Pixel size	918.60	1706	353	235.11
	Height	40.74	66	20	7.15
	Width	29.57	42	14	3.73
Total	Pixel size	1047.88	1979	353	262.37
	Height	47.32	74	20	9.60
	Width	29.14	47	14	3.57

Principal component analysis was implemented (PCA) as a method for dimensionality reduction (Fig. 2B), where the number of components was fixed at 10 (Xu et al., 2023). This PCA process preceded the CNN feature extraction phase. Employing PCA in this sequence was aimed at augmenting the CNN's efficiency and potentially bolstering its capacity to discern significant features pertinent to classification tasks. This enhancement could be attributed to the diminished noise and redundancy in the data following PCA processing.

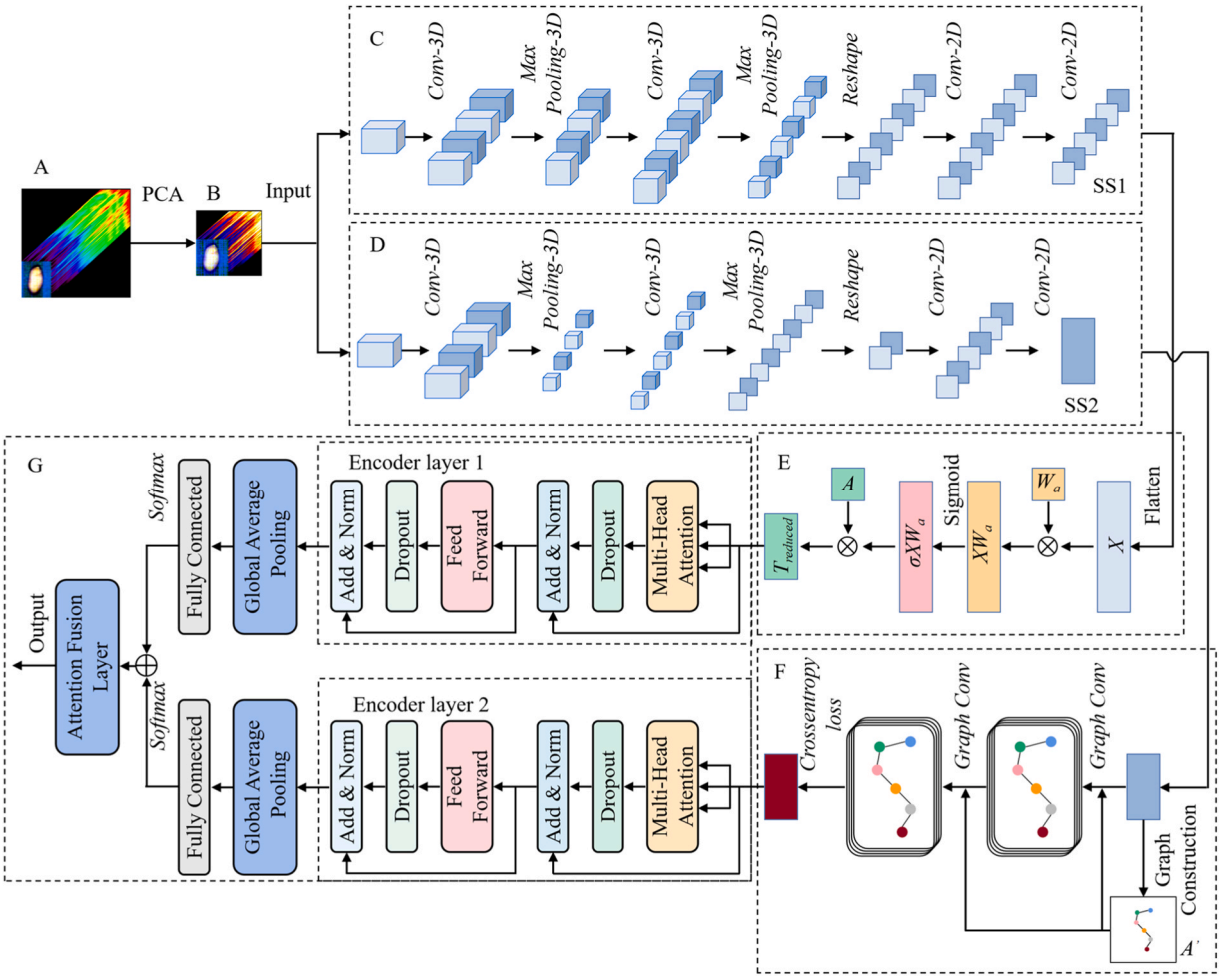


Fig. 2. Feature extraction and DAASST architectures, (A) individual peanut kernel hyperspectral image, (B) individual peanut kernel hyperspectral image processed by PCA, (C) the 1st CNN model, (D) the 2nd CNN model, (E) Laplacian-weighted feature tokenizer, (F) GCN feature extractor, (G) DAASST architectures.

2.4. Spectral-spatial feature extraction of peanut kernels

The 1st CNN model's intricate structure was composed of a succession of layers that included 3D convolutional layers, max pooling layers, a reshape operation layer, and 2D convolutional layers (Fig. 2C). Initially, the model processed the individual peanut kernel hyperspectral image $X_0 \in \mathbb{R}^{50\times25\times10}$. The initial 3D convolutional layer contained 16 filters of dimensions of $3\times3\times3$, and employed a rectified linear unit (ReLU) activation function. It utilized same padding to preserve spatial dimensions, yielding an output tensor $X_1 \in \mathbb{R}^{50\times25\times10\times16}$. The model then employed a 3D max pooling operation, strategically reducing spatial resolution by halving the 1st and 3rd spatial dimensions, and producing an output tensor $X_2 \in \mathbb{R}^{25\times25\times5\times16}$. This was followed by another 3D convolutional layer, now with 32 filters, maintaining kernel size and strides, again using same padding, leading to an output tensor $X_3 \in \mathbb{R}^{25\times25\times5\times32}$. Subsequent to this, another max pooling layer reduced the 3rd dimension by half, resulting in an output of $X_4 \in \mathbb{R}^{25\times25\times2\times32}$. This tensor was then reshaped in preparation for the 2D convolutional layers, merging depth and filter dimensions to construct $X_5 \in \mathbb{R}^{25\times25\times64}$. In the next stages, two 2D convolutional layers were applied, each utilizing filters sized 3×3,. The 1st 2D convolutional layer used 64 filters, and the 2nd employed 32, both with ReLU activation and same padding, yielding a final output tensor $X_6 \in \mathbb{R}^{25\times25\times32}$. The model utilized an Adam optimizer and a cross-entropy loss function, and included an accuracy metric for performance evaluation.

The 2nd CNN model had the same input $X_0^{'} \in \mathbb{R}^{50\times25\times10}$, and included a 3D convolution layer (Fig. 2D). It preserved the original dimensions and enhanced the depth to 16 channels, resulting in a tensor $X_1^{'} \in \mathbb{R}^{50\times25\times10\times16}$. This layer was succeeded by a 3D max pooling layer, adeptly reducing the image dimensions to $X_2^{'} \in \mathbb{R}^{10\times5\times2\times16}$, thus effectively condensing the spatial information. A subsequent 3D convolution layer further refined the data, augmenting the depth to 32 channels, and yielding $X_3^{'} \in \mathbb{R}^{10\times5\times2\times32}$. Following this, a 2nd max pooling layer compressed the spatial dimensions to a dense tensor $X_4^{'} \in \mathbb{R}^{5\times5\times1\times32}$. The data then underwent a reshaping process, and produced an output tensor $X_5^{'} \in \mathbb{R}^{10\times10\times8}$. In the next stages, a 2D convolution layer upheld the spatial dimensions while continuing with a depth of 32, resulting in $X_6^{'} \in \mathbb{R}^{10\times10\times32}$. The final phase of the model involved a reshaping step that flattened the spatial dimensions, transforming the data into a 2-dimensional feature vector $X_7^{'} \in \mathbb{R}^{100\times32}$, comprising 100 elements, each with 32 channels. This structured sequence of convolutional layers and pooling layers was strategically formulated to incrementally refine the input image into a nuanced feature representation.

This approach allowed the model to extract hierarchical features from the input data. The 3D convolutional layers were tasked with capturing spectral-spatial features, while the 2D layers were focused on spatial feature extraction from the reshaped maps. This methodology aligned with the principle of extracting spatial-spectral information from hyperspectral images (Roy et al., 2020; Sun et al., 2022). The spectral-spatial features extracted by the 1st CNN model and the 2nd CNN model were named SS1 and SS2, respectively.

2.5. Laplacian-weighted feature tokenizer

In order to augment the characterization of peanut kernel attributes, the extracted feature maps SS1 were converted into semantic tokens. These tokens were specifically designed to encapsulate and process advanced semantic constructs. This transformation facilitated a more nuanced and effective representation of the peanut kernels' characteristics.

In the Laplacian-weighted feature tokenizer function, each input SS1 from the feature map batch underwent a transformation into semantic tokens with a target shape (Fig. 2E). This process was initialized by generating a weight matrix $W_a \in \mathbb{R}^{D\times D}$ with a Laplacian distribution, characterized by a diversity parameter b=1.0. The semantic group matrix $A \in \mathbb{R}^{D\times A_{\dim}}$ was then defined, aiming to project the flattened feature maps into a semantic space upon application. For each feature map in the batch, the transformation proceeded as follows: The feature map was flattened, resulting in $X \in \mathbb{R}^{uvw\times D}$, which correlated to the reshaping process from a 3D tensor of shape (u, v, z) to a 2D matrix where u, v were the total number of spatial features, and z was the number of channels. The number of the u, v, z was 25, 25, 32, respectively, based on the shape of SS1. The number of D and $A_{\dim}$ were also 32. X was then subjected to a pointwise product with W_a, followed by a sigmoid activation function to yield semantically mapped features:

$$\text{semantic mapped} = \sigma(XW_a) \tag{1}$$

where σ denoted the sigmoid function. These features were further transformed into tokens T through multiplication with the transposed semantic group matrix A:

$$T = \sigma(XW_a)A^T \tag{2}$$

The resulting tokens were averaged along the spatial dimension to reduce the feature representation to the target height, resulting in reduced $T_{\text{reduced}} \in \mathbb{R}^{\text{target height}\times A_{\dim}}$. Where, the target height was the number of token vectors selected after the averaging process which was 100. Finally, a subset of these reduced tokens was selected to match the desired output tensor $X \in \mathbb{R}^{100\times 32}$. Therefore, SS1 was transformed into a more abstract representation that preserved essential information while discarding redundant or irrelevant details, and the transformed feature was named LW features.

2.6. GCN feature extractor

A GCN feature extractor was composed of 2 graph convolutional layers, each initialized to channel high-level feature representations through ReLU activation functions (Fig. 2F). The input to the graph convolutional layer was twofold including the feature matrix SS2 and an adjacency matrix. The adjacency matrix $A' \in \mathbb{R}^{100\times 100}$ was generated by a spatial construction mechanism, where neighboring nodes in the feature space were presumed to be linearly connected, fostering the exchange of information between immediately adjacent nodes in a structured manner reminiscent. The crux of the GCN depended on its iterative engagement of graph convolutional layers, where each layer took the preceding output and the adjacency matrix to propagate and transform feature information across the graph's structure. As each graph convolutional layer operated, it infused the input features with contextual information gleaned from their graph neighbors, modulating the feature matrix through learned weight parameters and non-linear activation. The culmination of this sequential convolution was the output of a transformed feature matrix $X' \in \mathbb{R}^{100\times 32}$, which was named GCN features.

2.7. DAASST

The DAASST integrated the improved transformer architecture with an attention fusion mechanism for processing dual inputs including LW features and GCN features (Fig. 2G). The transformer encoder layers consisted of a multi-head attention component and a feed-forward neural network. Each encoder layer incorporated layer normalization and dropout for regularization, leveraging skip connections to facilitate gradient flow and mitigate the vanishing gradient problem. Each encoder layer was followed by a global average pooling layer and a fully connected layer, the pooled features were classified using softmax-activated dense layers to generate LW classifications and GCN classifications. These classifications were then amalgamated by the attention fusion layer, which applied learned attention weights to integrate the LW features and GCN features, producing a unified output that benefits from both input types. This layer captured the essence of the model's attention fusion, where softmax-activated dense layers determine the contribution of each input type to the final prediction. The attention mechanism's adaptability allowed the model to emphasize more informative features, which was particularly beneficial to scenarios.

2.8. Experiment settings

The processing of hyperspectral images and the running of models were conducted in a Python 3.7.8 environment, utilizing TensorFlow 2.10.0. For training, experiments employed the Adam optimizer with an initial learning rate set at 0.001. In the classification experiments of 8-class (a control group and 7 contaminated groups with contamination time from 1 to 7 d), the categorical crossentropy loss function was used. For the experiments involving 2-class classification (a control group and a contaminated group), the binary crossentropy loss function was applied. The batch size and training epochs were set to 256 and 50, respectively. Samples were randomly divided into a training set and a testing set at a quantity ratio of 4:1, the training set and the test set contained 2688 and 672 peanut kernel individual hyperspectral images, respectively. Five-fold cross-validation was implemented during the training process. Five quantitative metrics were used to assess and compare performance including overall accuracy, precision, recall, F1-score and model run time. To more objectively evaluate and validate the models, each experiment was repeated 10 times, with the mean and standard deviation of each metric being reported.

3. Results and Discussion

3.1. Spectral analysis

A total of 3360 spectral curves were used in this work, with 1680 from the control group and 240 from each of the 7 contaminated groups. Initially, the contaminated groups' spectral reflectance was higher than that of the control group within the 1000–2400 nm range (Fig. 3A), which was in agreement with the others' work (Yuan et al., 2020; Qiao et al., 2017). Subsequently, the spectral reflectance curves of the control group and 7 contaminated groups showed similar patterns of change (Fig. 3B). As the days of contamination increased, there was a corresponding rise in the reflectance of peanut kernels. This increase was likely attributed to changes in the chemical and physical properties within the peanuts due to *Aspergillus flavus* contamination. These alterations led to varying absorption and reflection characteristics of light at different wavelengths on the surface or within the peanut kernels. Additionally, 'Weihua' exhibited a higher spectral reflectance than 'Baisha' in the range of 1000–2216 nm (Fig. 3C). Moreover, beyond 2216 nm, the spectral reflectance of 'Baisha' exceeded that of 'Weihua'. This illustrated the differences in material composition between the 2 peanut kernel varieties. The spectral curves of the 3360 peanut kernel samples, with the spectra identified principal reflectance peaks at 1114, 1302 and 1861 nm (Fig. 3D). The peak at 1114 nm was linked to the 2nd overtone of C-H stretching, while the peak at 1302 nm aligned with combined C-H stretching (Stuart, 2004). The peak at 1861 nm was associated with the combination of C-O and O-H stretching (Kimuli et al., 2018; Berardo et al., 2005). These spectral variances suggested

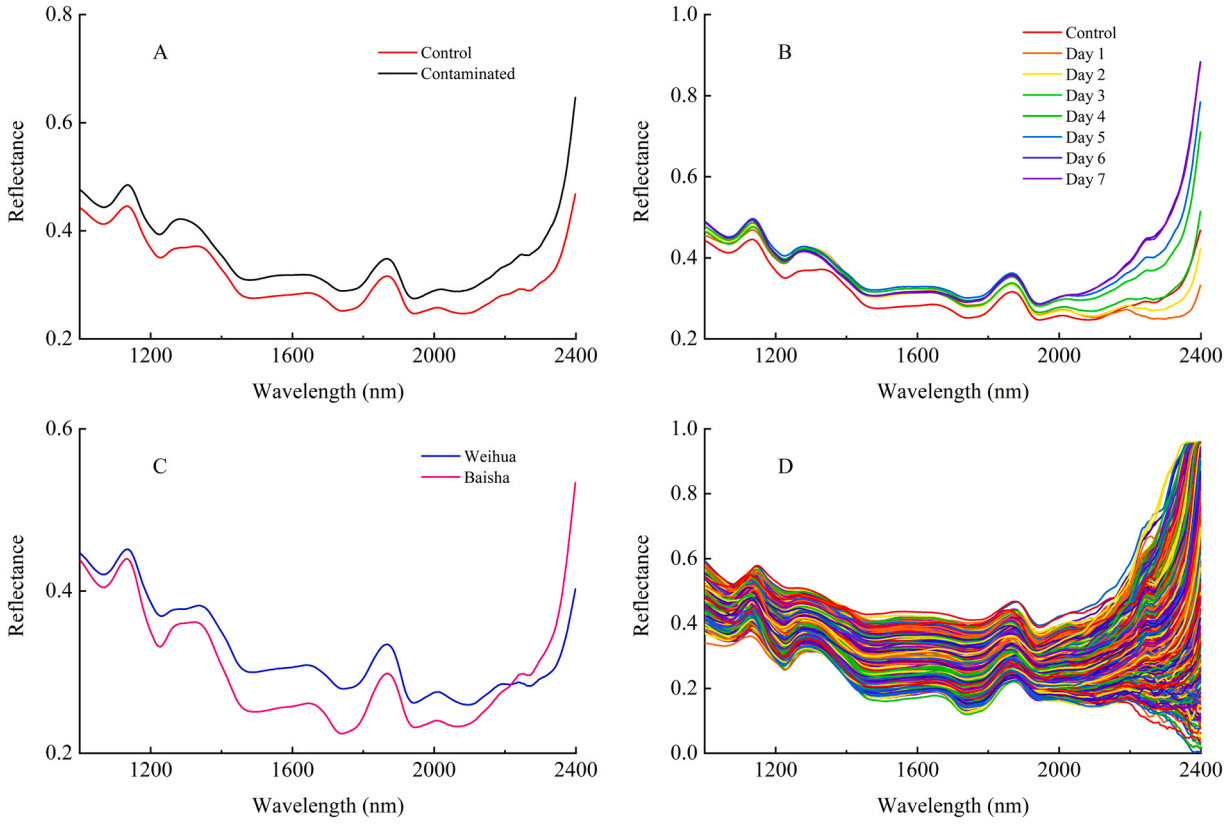


Fig. 3. Spectral curves of peanut kernels, (A) spectral curves of the contaminated group and control group, (B) spectral curves of the 7 contaminated groups and control group, (C) spectral curves of the 'Weihua' and 'Baisha', (D) spectral curves of the 3360 peanut kernel samples.

that the interaction between *Aspergillus flavus* and peanut kernels, marked by fungal contamination, led to distinct physical and chemical changes (Kaya-Celiker et al., 2015).

3.2. Model performance based on different parameters

In the DAASST model architecture, the number of layers specified the number of encoding layers used in the model. Increasing the number of layers enhanced the model's complexity and expressive capability, aiding in capturing more intricate features and relationships. However, an excessive number of layers could lead to overfitting. The number of heads determined the count of parallel attention heads. The multi-head attention enabled the model to learn information in different subspaces, thus improving its ability to process diverse types of information. Nevertheless, too many heads also increased the computational burden. The dimension of the feedforward (Dff) set the dimension of the feed-forward network within the encoding layers. A larger Dff enhanced the model's learning capacity but also escalated both computational demands and the risk of overfitting. The dropout rate defined the dropout ratio applied to various layers during training to prevent overfitting. An appropriate dropout rate helped in boosting the model's generalization ability. These 4 parameters were crucial for optimizing model performance, and their correct configuration required experimental determination. Therefore, systematic assessment of model performance within a given experimental parameter space reduced the number of experiments needed while ensuring a comprehensive exploration of the parameter space.

Table 2 established 9 different parameter configurations, setting up 18 distinct experiments for peanut kernels classification into 8-class and 2-class using DAASST, respectively, named DAASST 8 and DAASST 2. Each parameter configuration underwent 10 experiments, totaling 90 trained DAASST 8 and DAASST 2 models, which were then evaluated.

Table 2
Parameter settings.

No.	Number of layers	Number of heads	Dff	Dropout rate
1	2	2	128	0.1
2	2	4	256	0.2
3	2	8	512	0.3
4	3	2	256	0.3
5	3	4	512	0.1
6	3	8	128	0.2
7	4	2	512	0.2
8	4	4	128	0.3
9	4	8	256	0.1

Table 3 showed that in the 9 parameter experiments of DAASST 8, the 4th experiment performed the best, achieving an overall accuracy, precision, recall and F1-score of 99.40%. The number of layers, the number of heads, Dff and the dropout rate were set to 3, 2, 256 and 0.3, respectively. This indicated that the model effectively fitted the data and exhibited robustness. The 3rd experiment and the 4th experiment both achieved good classification effects, but the 4th experiment had a shorter runtime. This was attributed to the lower number of heads and Dff in the 4th experiment, reducing the model's computational load. In the 9 parameter experiments of DAASST 2, the 11th experiment achieved the best classification results, with an overall accuracy, precision, recall and F1-score of 100% and a runtime of 50.46 s. Additionally, the 7th, 8th, 9th, 16th, 17th and 18th experiment did not achieve convergence. This suggested that when the encoding layers were set to 4, the model became overly complex to fit the data. Consequently, the optimal parameter combination for DAASST 8 was identified as 3, 2, 256 and 0.3, and for DAASST 2 as 2, 8, 512 and 0.3. Compared to DAASST 8, DAASST 2 reached 100% classification accuracy in overall accuracy and other metrics, and had a shorter runtime. Its optimal parameters

Table 3
Model performance based on different parameters.

Model	No.	Overall accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Run time (s)
DAASST 8	1	98.96 ±1.03	98.97 ±1.01	99.27 ±0.42	99.12 ±0.71	38.72 ±0.46
	2	97.74 ±2.34	97.80 ±2.17	98.71 ±0.98	98.24 ±1.27	47.58 ±0.70
	3	99.40 ±0.00	99.40 ±0.00	99.40 ±0.00	99.40 ±0.00	67.22 ±0.43
	4	**99.40 ±0.00**	**99.40 ±0.00**	**99.40 ±0.00**	**99.40 ±0.00**	**55.31 ±0.30**
	5	98.96 ±0.73	98.96 ±0.73	99.11 ±0.64	99.03 ±0.65	72.31 ±0.51
	6	99.06 ±0.72	99.07 ±0.72	99.23 ±0.56	99.15 ±0.59	88.85 ±0.32
	7	53.12 ±0.00	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	77.40 ±0.61
	8	53.12 ±0.00	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	82.83 ±1.37
	9	53.12 ±0.00	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	117.82 ±0.47
DAASST 2	10	98.88 ±1.97	98.21 ±3.74	99.59 ±1.31	98.85 ±1.98	41.74 ±2.28
	11	**100.00 ±0.00**	**100.00 ±0.00**	**100.00 ±0.00**	**100.00 ±0.00**	**50.46 ±2.32**
	12	98.72 ±2.44	97.55 ±4.56	100.00 ±0.00	98.71 ±2.43	70.67 ±2.60
	13	98.79 ±1.70	97.60 ±3.35	100.00 ±0.00	98.76 ±1.74	58.86 ±3.60
	14	99.12 ±2.18	100.00 ±0.00	98.13 ±4.66	99.00 ±2.51	76.24 ±3.56
	15	99.40 ±0.96	99.18 ±1.74	99.59 ±1.31	99.37 ±1.02	95.04 ±6.09
	16	53.12 ±0.00	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	85.48 ±9.15
	17	53.12 ±0.00	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	88.13 ±5.38
	18	53.12 ±0.00	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00	124.25 ±5.04

indicated that a larger number of attention heads and feedforward network dimensions were more effective for identifying peanuts contaminated with *Aspergillus flavus*. In summary, DAASST accurately classified peanuts contaminated with *Aspergillus flavus* and was capable of identifying peanut kernels contaminated time.

3.3. Ablation experiments

The ablation experiments helped to elucidate the necessity and effectiveness of each component, providing a more nuanced understanding of the model's functionality. Therefore, a total of 6 combinations were designed for evaluation in the 8-class classification.

Initially, the first 4 experiments did not incorporate the attention fusion layer. Instead, only 1 of the following features was utilized including SS1, SS2, LW features, or GCN features. Table 4 revealed that using only LW features yielded superior classification performance compared to exclusively using SS1 features. Similarly, employing only GCN features outperformed using SS2 features. This underscored the effectiveness of the Laplacian-weighted feature tokenizer and GCN feature extractor, indicating their capability to capture more advanced and valuable feature information. Subsequently, in the absence of an attention fusion layer, the 8th experiment utilizing an average fusion layer exhibited the poorest performance, with an overall accuracy of 77.59%. This emphasized the efficacy of the attention fusion layer. Finally, employing the attention fusion layer to merge LW features and GCN features in the DAASST model resulted in the best classification performance. The overall accuracy, precision, recall and F1-score all reached 99.40%. However, it's important to note that this model incurred a longer runtime compared to the first 4 experiments due to the need for consistent experimental conditions and running a fixed 50 epochs. In reality, DAASST achieved data fitting within the first 10 epochs (Fig. 4). This indicated that leveraging the attention fusion layer for the classification results of LW features and GCN features, through attention-weighted fusion, enhanced the model's generalization performance and stability. It allowed the model to make informed predictions by leveraging the complementary strengths of different input types. Furthermore, GCN features and SS2 exhibited faster convergence compared to LW features and SS1, suggesting that SS2 and GCN features capture more crucial data features.

3.4. Other methods for classification

3.4.1. Machine learning algorithms based on spectral information

Five machine learning algorithms were employed to identify *Aspergillus flavus* contamination in peanut kernels and differentiate contamination time including linear discriminant analysis (LDA), principal component analysis-linear discriminant analysis (PCA-LDA), partial least squares-discriminant analysis (PLS-DA), SVM and 1D-CNN (Guo et al., 2024; Zhang et al., 2023). Given that SS1, SS2, LW features and GCN features were high-dimensional, the average spectra of extracted peanut kernel ROIs were utilized as feature inputs for the models. Additionally, the successive projection algorithm (SPA) and competitive adaptive resampling scheme (CARS) were applied for feature wavelength extraction (Guo et al. 2023). A total of 30 models were constructed, as presented in Table 5. SPA and CARS extracted 40 and 58 feature wavelengths, respectively. All models underwent 5-fold cross-validation, with n components set to 10 for PCA-LDA and

Table 4
Model performance in ablation experiments.

No.	SS1	SS2	LW features	GCN features	Attention fusion layer	Overall accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Run time (s)
1	√	×	×	×	×	98.54 ± 1.91	98.64 ± 1.70	98.47 ± 2.00	98.55 ± 1.85	35.87 ± 0.32
2	×	√	×	×	×	98.39 ± 1.14	98.41 ± 1.12	99.03 ± 0.80	98.72 ± 0.88	36.49 ± 1.27
3	×	×	√	×	×	99.27 ± 0.13	99.29 ± 0.12	99.27 ± 0.13	99.28 ± 0.12	35.59 ± 0.33
4	×	×	×	√	×	99.08 ± 1.04	99.12 ± 0.91	99.02 ± 1.22	99.07 ± 1.07	35.61 ± 0.55
5	×	×	√	√	△	77.59 ± 16.47	95.59 ± 7.63	50.16 ± 28.27	61.44 ± 24.73	64.63 ± 0.24
6	×	×	√	√	√	**99.40 ± 0.00**	**99.40 ± 0.00**	**99.40 ± 0.00**	**99.40 ± 0.00**	**55.31 ± 0.30**

Note: △ denotes average fusion layer.

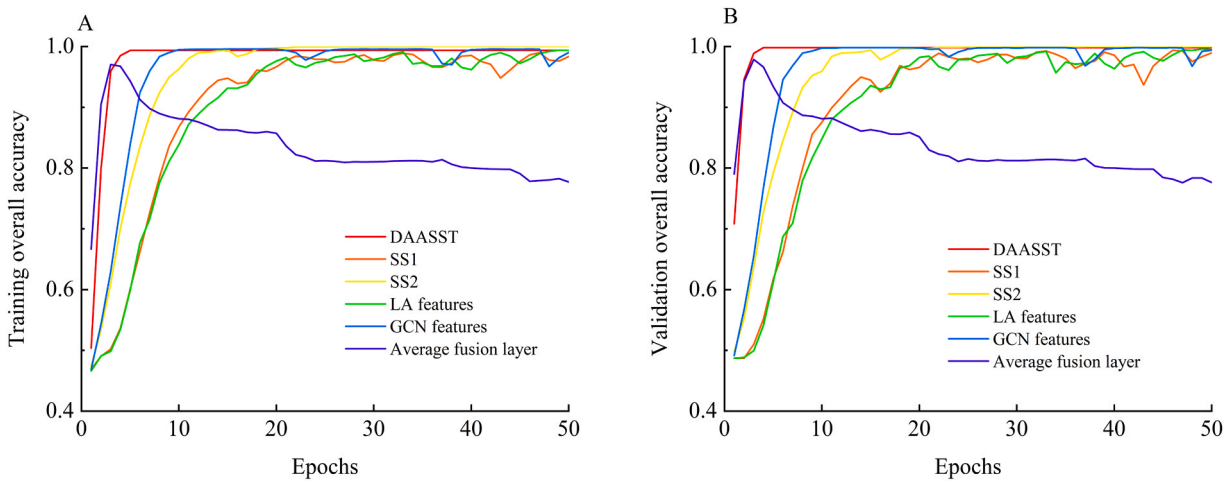


Fig. 4. Overall accuracy curves in ablation experiments, (A) training overall accuracy curves in ablation experiments, (B) validation overall accuracy curves in ablation experiments.

Table 5
Machine learning algorithms performance.

Category	Method	Overall accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
8-class	**LDA**	**92.71**	**86.66**	**86.44**	**86.50**
	PCA-LDA	65.92	43.25	42.56	41.56
	PLS-DA	72.92	49.97	50.17	46.36
	SVM	69.49	51.33	44.64	41.72
	1D-CNN	78.42	61.45	59.80	59.47
	SPA-LDA	82.89	69.25	68.57	68.46
	SPA-PCA-LDA	63.69	38.22	38.30	37.47
	SPA-PLS-DA	68.75	41.68	42.52	39.30
	SPA-SVM	64.14	44.90	38.07	33.97
	SPA-1D-CNN	73.66	54.16	51.51	47.53
	CARS-LDA	87.50	77.15	76.66	76.75
	CARS-PCA-LDA	76.19	59.96	59.13	59.08
	CARS-PLS-DA	72.92	48.44	50.16	46.72
	CARS-SVM	69.49	48.17	44.21	42.24
	CARS-1D-CNN	76.79	57.65	57.85	54.29
2-class	**LDA**	**100.00**	**100.00**	**100.00**	**100.00**
	PCA-LDA	88.69	88.63	88.76	88.67
	PLS-DA	99.85	99.86	99.84	99.85
	SVM	98.07	98.03	98.10	98.06
	1D-CNN	99.55	99.56	99.54	99.55
	SPA-LDA	99.40	99.42	99.38	99.40
	SPA-PCA-LDA	90.92	90.87	90.99	90.90
	SPA-PLS-DA	99.40	99.40	99.40	99.40
	SPA-SVM	95.83	95.78	95.93	95.83
	SPA-1D-CNN	99.55	99.58	99.52	99.55
	CARS-LDA	**100.00**	**100.00**	**100.00**	**100.00**
	CARS-PCA-LDA	94.79	94.90	94.67	94.76
	CARS-PLS-DA	99.26	99.23	99.28	99.25
	CARS-SVM	97.47	97.44	97.49	97.46
	CARS-1D-CNN	99.11	99.07	99.16	99.10

PLS-DA, and 1D-CNN configured with 2 convolutional layers, a flatten layer, and 2 fully connected layers. Default parameters were used for other models, and the experiments were conducted in a Python 3.7.8 environment.

Table 5 illustrated that LDA demonstrated the best classification performance among the 8-class classification. It achieved overall accuracy, precision, recall and F1-score of 92.71%, 86.66%, 86.44%, and 86.50%, respectively. Models based on the feature wavelengths did not exhibit improved classification performance compared with original wavelength models, possibly due to the loss of some useful information in feature wavelengths. In the 2-class classification, LDA and CARS-LDA performed exceptionally well, attaining 100% in overall accuracy, precision, recall and F1-score. With the exception of PCA-LDA, all other models achieved overall accuracy above 90%. The results indicated that the LDA model outperformed others in all categories, accurately identifying *Aspergillus flavus* contamination in peanuts. However, none of the 5 machine learning algorithms accurately identified the contamination time of peanut kernels.

3.4.2. *Deep learning algorithms based on the GCN features*

Representative classical deep learning models often exhibit increased architectural complexity, incorporating advanced neural network structures. Therefore, a comprehensive evaluation of various representative classical deep learning algorithms was conducted for the identification of *Aspergillus flavus* contamination and differentiate contamination time in peanut kernels. Six algorithms were analyzed including DenseNet-121 (Huang et al., 2017), ResNet-50 (He et al., 2016), SqueezeNet (Forrest et al., 2016), VGG-16 (Simonyan & Zisserman, 2014) and Xception (Chollet, 2016), alongside the proposed DAASST.

Initially focusing on the 8-class classification task, DAASST 8 demonstrated outstanding performance across all evaluation metrics in Table 6. It achieved an overall accuracy of 99.40%, with precision, recall and F1-score all at 99.40%, while exhibiting a relatively short runtime of 55.31 s. Notably, DenseNet-121 also performed well, but slightly below DAASST in various metrics. In contrast, VGG-16 exhibited significantly inferior performance, indicating potential overfitting due to its deep network structure. Transitioning to the 2-class classification task, DAASST 2 once again showcased remarkable performance, achieving perfect scores of 100% in overall accuracy, precision, recall and F1-score, with a concise runtime of 50.46 s. SqueezeNet also demonstrated strong performance, particularly in overall accuracy and runtime, though slightly trailing in precision and recall. Conversely, ResNet-50 exhibited comparatively lower performance. VGG-16's performance was notably suboptimal, particularly in recall and F1-score.

Additionally, particular attention was devoted to the model's performance across 8-class classification tasks. Consequently, an in-depth analysis of the overall accuracy during the training process and validation process based on the 6 algorithms was conducted. It was observed

Table 6
Deep learning algorithm performance.

Category	Algorithm	Overall accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Run time (s)
8-class	DenseNet-121	99.17±0.20	99.26±0.00	99.14±0.27	99.20±0.13	169.47±7.65
	ResNet-50	98.45±1.80	98.51±1.67	98.45±1.80	98.48±1.73	168.54±4.49
	SqueezeNet	96.90±2.90	98.29±1.57	96.90±2.90	97.57±1.77	27.81±1.93
	VGG-16	53.12±0.00	10.62±23.76	10.62±23.76	10.62±23.76	144.05±5.40
	Xception	93.51±13.18	93.51±13.18	93.51±13.18	93.51±13.18	323.28±2.15
	DAASST 8	**99.40±0.00**	**99.40±0.00**	**99.40±0.00**	**99.40±0.00**	**55.31±0.30**
2-class	DenseNet-121	99.91±0.08	100.00±0.00	99.81±0.17	99.90±0.09	175.81±5.71
	ResNet-50	93.84±13.61	92.16±17.54	99.56±0.66	94.97±11.08	164.39±4.54
	SqueezeNet	99.97±0.07	99.94±0.14	100.00±0.00	99.97±0.07	25.41±1.96
	VGG-16	46.88±0.00	46.88±0.00	100.00±0.00	63.83±0.00	142.84±3.97
	Xception	99.97±0.07	100.00±0.00	99.94±0.14	99.97±0.07	324.97±2.83
	DAASST 2	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**50.46±2.32**

that DAASST 8, DenseNet-121, ResNet-50, and Xception displayed commendable convergence speeds during the training process (Fig. 5A). DAASST 8 achieved nearly flawless accuracy at the 10th epoch, while DenseNet-121, ResNet-50 and Xception demonstrated relatively stable performance throughout the entire training process, with overall accuracy fluctuating between 0.98% and 0.99%. In contrast, SqueezeNet exhibited suboptimal performance in the initial stages but gradually converged with ongoing training, while VGG-16 failed to converge. During the validation process, DAASST 8 exhibited unparalleled convergence speed, reaching an overall accuracy of 99.81% as early as the 3rd epoch (Fig. 5B). This underscored the robustness of DAASST 8 in handling multi-class classification tasks. Although DenseNet-121 and SqueezeNet showed improvements, their performance fell short of that achieved by DAASST 8.

In summary, DAASST consistently outperformed other models in both 8-class and 2-class classification tasks. Its unique dual-aspect attention fusion layer and transformer architecture endowed it with enhanced expressive and generalization capabilities for classification tasks. Compared to classical deep learning models, DAASST not only achieved significant improvements in accuracy but also demonstrated high efficiency in runtime. These findings underscored the broad potential applications of DAASST in practical domains and provide valuable insights for the advancement of deep learning models in complex classification tasks.

4. Conclusion

The findings underscored the potential of DAASST in practical applications, offering a robust solution for identifying *Aspergillus flavus* contamination in peanut kernels and characterizing contamination time. The unique contributions of this study lie in the integration of advanced feature extraction techniques, attention mechanisms, and transformer architectures, providing a comprehensive and efficient approach to hyperspectral image classification. Although the results are encouraging, it's crucial to acknowledge that the increased complexity of deep learning models and hyperspectral images presents challenges for deploying online detection systems. Further research could explore the applicability of the proposed methodology to other domains and extend the understanding of hyperspectral imaging in online detection deployment.

Ethical approval

This article has no any study with human participants or animals by any of the authors.

CRediT authorship contribution statement

Haifang Wang: Methodology, Investigation, Conceptualization. **Shiling Li:** Writing – original draft, Methodology, Investigation, Conceptualization. **Ibrahim A. Darwish:** Writing – review & editing, Writing – original draft, Funding acquisition. **Jiashuai Sun:** Writing – original draft, Methodology, Investigation. **Lingjun Geng:** Writing – original draft, Investigation, Conceptualization. **Xijun Shao:** Writing – review & editing, Writing – original draft, Methodology. **Haowei Dong:** Writing – review & editing, Writing – original draft, Investigation. **Xia Sun:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Lianming Xia:** Writing – original draft, Conceptualization. **Zhen Guo:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Qi Zhang:**

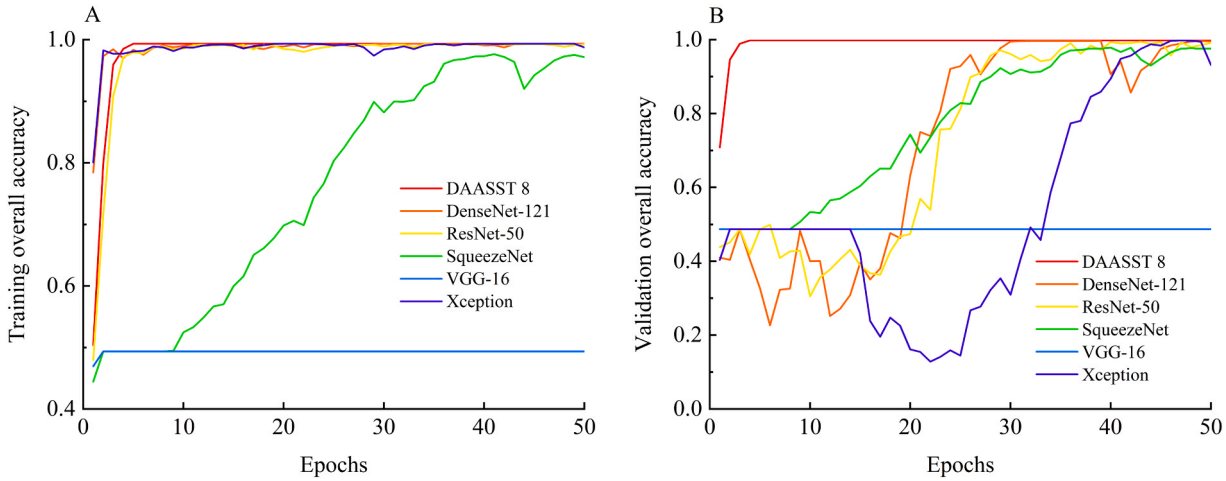


Fig. 5. Overall accuracy curves of different deep learning algorithms, (A) training overall accuracy curves of different deep learning algorithms, (B) validation overall accuracy curves of different deep learning algorithms.

Writing – original draft, Methodology. **Yemin Guo:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Jing Zhang:** Writing – original draft, Methodology, Investigation, Conceptualization.

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.

Data Availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 32372438, 31772068, 31872909), Funding Project for the Central Government to Guide the Development of Local Science and Technology (YDZX2022163), Shandong Province Major Applied Technology Innovation Project (SD2019NJ007), Technological innovation guidance project of Department of Science& Technology of Gansu Province (22CX8NA023) and Weifang Science and Technology Development Project (2021ZJ1103). The authors extend their appreciation to the Researchers Supporting Project number (RSPD2024R944), King Saud University, Riyadh, Saudi Arabia, for funding this work.

References

Azad, R., Kazerouni, A., Heidari, M., Aghdam, E.K., Molaei, A., Jia, Y., Jose, A., Roy, R., Merhof, D., 2024. Advances in medical image analysis with vision transformers: a comprehensive review. Med. Image Anal. 91, 103000 https://doi.org/10.1016/j.media.2023.103000.

Berardo, N., Pisacane, V., Battilani, P., Scandolara, A., Pietri, A., Marocco, A., 2005. Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy. J. Agric. Food Chem. 53 (21), 8128–8134. https://doi.org/10.1021/jf0512297.

Chu, X., Wang, W., Yoon, S.C., Ni, X., Heitschmidt, G.W., 2017. Detection of aflatoxin B_1 (AFB_1) in individual maize kernels using short wave infrared (SWIR) hyperspectral imaging. Biosyst. Eng. 157, 13–23. https://doi.org/10.1016/j.biosystemseng.2017.02.005.

Femenias, A., Gatius, F., Ramos, A.J., Sanchis, V., Marín, S., 2020. Use of hyperspectral imaging as a tool for *Fusarium* and deoxynivalenol risk management in cereals: a review. Food Control 108, 106819. https://doi.org/10.1016/j.foodcont.2019.106819.

Femenias, A., Gatius, F., Ramos, A.J., Teixido-Orries, I., Marín, S., 2022. Hyperspectral imaging for the classification of individual cereal kernels according to fungal and mycotoxins contamination: a review. Food Res. Int. 155, 111102 https://doi.org/10.1016/j.foodres.2022.111102.

Feng, C.-H., Makino, Y., Oshita, S., García Martín, J.F., 2018. Hyperspectral imaging and multispectral imaging as the novel techniques for detecting defects in raw and processed meat products: Current state-of-the-art research advances. Food Control 84, 165–176. https://doi.org/10.1016/j.foodcont.2017.07.013.

Gao, J., Zhao, L., Li, J., Deng, L., Ni, J., Han, Z., 2021. Aflatoxin rapid detection based on hyperspectral with 1D-convolution neural network in the pixel level. Food Chem. 360, 129968 https://doi.org/10.1016/j.foodchem.2021.129968.

Guo, Z., Zhang, J., Dong, H., Sun, J., Huang, J., Li, S., Ma, C., Guo, Y., Sun, X., 2023. Spatio-temporal distribution patterns and quantitative detection of aflatoxin B_1 and total aflatoxin in peanut kernels explored by short-wave infrared hyperspectral imaging. Food Chem. 424, 136441 https://doi.org/10.1016/j.foodchem.2023.136441.

Guo, Z., Zhang, J., Ma, C., Yin, X., Guo, Y., Sun, X., Jin, C., 2023. Application of visible-near-infrared hyperspectral imaging technology coupled with wavelength selection algorithm for rapid determination of moisture content of soybean seeds. J. Food Compos. Anal. 116, 105048 https://doi.org/10.1016/j.jfca.2022.105048.

Guo, Z., Zhang, J., Sun, J., Dong, H., Huang, J., Geng, L., Li, S., Jing, X., Guo, Y., Sun, X., 2024. A multivariate algorithm for identifying contaminated peanut using visible and near-infrared hyperspectral imaging. Talanta 267, 125187. https://doi.org/10.1016/j.talanta.2023.125187.

Han, Z., Gao, J., 2019. Pixel-level aflatoxin detecting based on deep learning and hyperspectral imaging. Comput. Electron. Agric. 164, 104888 https://doi.org/10.1016/j.compag.2019.104888.

He, X., Yan, C., Jiang, X., Shen, F., You, J., Fang, Y., 2021. Classification of aflatoxin B_1 naturally contaminated peanut using visible and near-infrared hyperspectral imaging by integrating spectral and texture features. Infrared Phys. Technol. 114, 103652 https://doi.org/10.1016/j.infrared.2021.103652.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

Huang, C., Li, X., Wen, Y., 2021. AN OTSU image segmentation based on fruitfly optimization algorithm. Alex. Eng. J. 60 (1), 183–188. https://doi.org/10.1016/j.aej.2020.06.054.

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. Proc. - 0th IEEE Conf. Comput. Vis. Pattern Recognit., CVPR 2017, 4700–4708. https://doi.org/10.1109/CVPR.2017.243.

Chollet, F., 2016. Xception: Deep learning with depthwise separable convolutions. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1800-1807.

Islam, S., Elmekki, H., Elsebai, A., Bentahar, J., Drawel, N., Rjoub, G., Pedrycz, W., 2024. A comprehensive survey on applications of transformers for deep learning tasks. Expert Syst. Appl. 241, 122666 https://doi.org/10.1016/j.eswa.2023.122666.

Kaya-Celiker, H., Mallikarjunan, P.K., Kaaya, A., 2015. Mid-infrared spectroscopy for discrimination and classification of *Aspergillus spp*. contamination in peanuts. Food Control 52, 103–111. https://doi.org/10.1016/j.foodcont.2014.12.013.

Kimuli, D., Wang, W., Wang, W., Jiang, H., Zhao, X., Chu, X., 2018. Application of SWIR hyperspectral imaging and chemometrics for identification of aflatoxin B_1 contaminated maize kernels. Infrared Phys. Technol. 89, 351–362. https://doi.org/10.1016/j.infrared.2018.01.026.

Liu, Z., Jiang, J., Qiao, X., Qi, X., Pan, Y., Pan, X., 2020. Using convolution neural network and hyperspectral image to identify moldy peanut kernels. LWT 132, 109815. https://doi.org/10.1016/j.lwt.2020.109815.

Long, Y., Huang, W., Wang, Q., Fan, S., Tian, X., 2022. Integration of textural and spectral features of Raman hyperspectral imaging for quantitative determination of a single maize kernel mildew coupled with chemometrics. Food Chem. 372, 131246 https://doi.org/10.1016/j.foodchem.2021.131246.

Qiao, X., Jiang, J., Qi, X., Guo, H., Yuan, D., 2017. Utilization of spectral-spatial characteristics in shortwave infrared hyperspectral images to classify and identify fungi-contaminated peanuts. Food Chem. 220, 393–399. https://doi.org/10.1016/j.foodchem.2016.09.119.

Roy, S.K., Krishna, G., Dubey, S.R., Chaudhuri, B.B., 2020. HybridSN: exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. IEEE Geosci. Remote Sens. Lett. 17 (2), 277–281. https://doi.org/10.1109/LGRS.2019.2918719.

Forrest N. Iandola, Matthew W. Moskewicz, Khalid Ashraf, Song Han, William J. Dally, Kurt Keutzer, 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. CoRR abs/1602.07360.

K. Simonyan, A. Zisserman. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Stuart, B., 2004. Infrared Spectroscopy: Fundamentals and Applications. John Wiley & Sons, pp86.

Sun, J., Wang, G., Zhang, H., Xia, L., Zhao, W., Guo, Y., Sun, X., 2020. Detection of fat content in peanut kernels based on chemometrics and hyperspectral imaging technology. Infrared Phys. Technol. 105, 103226 https://doi.org/10.1016/j.infrared.2020.103226.

Sun, L., Zhao, G., Zheng, Y., Wu, Z., 2022. Spectral-spatial feature tokenization transformer for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2022.3144158.

Tao, F., Yao, H., Hruska, Z., Kincaid, R., Rajasekaran, K., Bhatnagar, D., 2020. A novel hyperspectral-based approach for identification of maize kernels infected with diverse *Aspergillus flavus* fungi. Biosyst. Eng. 200, 415–430. https://doi.org/10.1016/j.biosystemseng.2020.10.017.

Wu, D., Sun, D.W., 2013. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: a review-part II: applications. Innov. Food Sci. Emerg. Technol. 19, 1–14. https://doi.org/10.1016/j.ifset.2013.04.016.

Xu, Z., Hu, H., Wang, T., Zhao, Y., Zhou, C., Xu, H., Mao, X., 2023. Identification of growth years of Kudzu root by hyperspectral imaging combined with spectral-spatial feature tokenization transformer. Comput. Electron. Agric. 214, 108332 https://doi.org/10.1016/j.compag.2023.108332.

Yuan, D., Jiang, J., Qi, X., Xie, Z., Zhang, G., 2020. Selecting key wavelengths of hyperspectral imagine for nondestructive classification of moldy peanuts using ensemble classifier. Infrared Phys. Technol. 111, 13058. https://doi.org/10.1016/j.infrared.2020.103518.

Zhang, S., Yin, Y., Liu, C., Li, J., Sun, X., Wu, J., 2023. Discrimination of wheat flour grade based on PSO-SVM of hyperspectral technique. Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 302, 123050 https://doi.org/10.1016/j.saa.2023.123050.

Zhao, X., Wang, W., Chu, X., Li, C., Kimuli, D., 2017. Early detection of *Aspergillus parasiticus* infection in maize kernels using near-infrared hyperspectral imaging and multivariate data analysis. Appl. Sci. 7 (1), 90. https://doi.org/10.3390/app7010090.