Rapid and Accurate Varieties Identification of Different Crop Seeds Under Sample-Limited Condition Based on Hyperspectral Imaging and Deep Transfer Learning

Na Wu
Zhejiang University
Fei Liu
Zhejiang University
Yidan Bao
Zhejiang University
Mu Li
Jilin Academy of Agricultural Sciences
Wei Huang
Jilin Academy of Agricultural Sciences
Fanjia Meng
China Agricultural University
Chu Zhang
Zhejiang University
Yong He (yhe@zju.edu.cn)
Zhejiang University https://orcid.org/0000-0001-6752-1757

Research

Keywords: Hyperspectral imaging, Crop seeds, Deep learning, Transfer learning

Posted Date: November 20th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-108818/v1

Rapid and accurate varieties identification of different crop seeds under sample-limited condition based on hyperspectral imaging and deep transfer learning

Na Wu[a], Fei Liu[a], Yidan Bao[a], Mu Li[c], Wei Huang[c], Fanjia Meng[d], Chu Zhang[a], Yong He[a,b]*

[a] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

[b] State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310058, China

[c] Maize Research Institute, Jilin Academy of Agricultural Sciences, Gongzhuling 136100, China

[d] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

* Corresponding Author at: College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China.

E-mail: yhe@zju.edu.cn

Abstract

Background: Varieties identification of crop seeds is significant for breeders to screen out seeds with specific traits and for market regulators to detect seeds purity. Hyperspectral imaging technology provides a fast and non-destructive means for varieties identification. And deep learning algorithm is suitable for effective analysis of redundant spectral data. However, deep learning algorithms have serious big data dependency, while collecting high-quality large-scale samples was high-cost in many cases. This made it difficult to build an accurate identification model. This study aimed to explore a rapid and accurate method for varieties identification of different crop seeds under sample-limited condition based on hyperspectral imaging and deep transfer learning.

Results: Three deep neural networks with typical structures were designed based on a samples-rich Pea dataset. Obtained the highest accuracy of 99.57 %, VGG-MODEL was transferred to classify four target datasets (Rice, Oat, Wheat, Cotton) with limited samples. The accuracies of deep transferred model achieved 95 %, 99 %, 80.8 %, and 83.86 % on the four datasets, respectively. Using training sets with different sizes, deep transferred model could always obtain higher performance than other traditional methods. Visualization of training process and classification results confirmed the portability of common features of seed spectra and provided an interpreted method for rapid and accurate varieties identification of crop seeds.

Conclusions: This study combined hyperspectral imaging and deep transfer learning to identify varieties of different crop seeds, which was proved to be efficient under sample-limited condition. This facilitated crop variety screening process under the scenario of sample scarcity. It also provided a new idea for the detection of other qualities of crop seeds based on hyperspectral imaging under sample-limited condition.

Key words: Hyperspectral imaging, Crop seeds, Deep learning, Transfer learning

Background

High-quality seeds are conducive to continue excellent species and guarantee crop yield and quality. Due to the large differences of climate, soil, water resources in different regions, breeders have pointedly developed many crop varieties to adapt to the local planting environment. Growth rules, stress resistance and bio-chemical characteristics of different varieties crops vary greatly. For varieties that are still in breeding stage, screening a variety with specific traits often need to observe the phenotypic traits of the offspring plants, which is time-consuming and labor-intensive. As a seed carries all the genetic genes that develop into a plant, seed identification can be an alternative for screening variety with specific traits. For varieties that are already promoted widely, different varieties of seeds tend to be easily mixed with frequent circulation in the market. This makes the seed purity unable to be guaranteed. Conventionally, manual vision inspection method based on the external phenotypic traits of seeds, like color, texture and shape, is subjective and boring. The more accurate methods based on the internal biochemical properties of seeds, such as DNA molecular markers [1] and protein electrophoresis techniques [2], rely on chemical agents and complex operation. Accordingly, it is necessary to develop a fast and accurate method for varieties identification of crop seeds.

As hyperspectral imaging (HSI) can obtain spectral and spatial location information simultaneously during one scan, it has the capability of probing the internal and external phenotypic traits of samples rapidly. HSI has gained great and continuous attention in breed screening [3], plant phenotyping [4] and environment monitoring [5]. In seed-related tasks like determination of seed quality [6], diagnosis of seed diseases [7] and detection of seed component [8], HSI has been utilized as a fast and accurate alternative. Since hyperspectral image contains a large amount of redundant collinear information, diverse linear and nonlinear machine learning approaches like partial least squares discriminant analysis (PLS-DA), extreme learning machine(ELM), least square support vector machines (LSSVM), and partial least squares regression (PLSR) were introduced to couple the relationship between seed spectra and category label or component content [8-10].

In recent years, with the attention from academia and industry increasing, deep learning as the new state of art machine learning approach has also been applied in spectral analysis filed gradually [11-13]. Compared with traditional approaches, deep learning can extract various low-level and high-level features automatically through a multi-layered stack network structure [14]. This can reduce the requirement of prior knowledge

from specific tasks and human effort in feature engineering, which is very beneficial for the analysis of redundant and high-dimensional spectral data.

However, deep learning models have serious big data dependencies. A high-performance deep model requires enough samples to adequately learn the feature patterns hidden in the massive and redundant spectral data. Unfortunately, in some tasks like seeds screening with specific traits during breeding or quality detection of precious agricultural products, it is very difficult to establish a large-scale high-quality dataset because of the high cost of obtaining and labeling samples [4,15-16]. In addition, the precious data acquired at great expense is very easy to be outdated and difficult to be reused in new tasks. This greatly limits the rapid application of well-performing methods like deep learning in spectral analysis. Another problem is that deep models developed for different tasks are generally based on a common assumption, that is, training and testing data lie in same feature space and have same distribution [17]. Therefore, even for similar tasks, the tiny differences in the distribution of different datasets caused by the variation in sample state and spectral acquisition parameters will make the model not reusable.

The emergence of transfer learning brings hope for solving the above two problems. Allowing the training and testing data lie in different feature spaces, transfer learning mainly investigates methods how to transfer useful knowledge learnt from the relevant source domain into the target domain [18]. This not only relieve the demand for a large number of samples in target task, but also make reusing the useful knowledge like model structure and feature representation in the source domain possible. The target task can be expected completed using limited samples and computation overhead. Deep transfer learning is the product of the combination of deep learning and transfer learning. It aims to study how to make use of deep neural network to transfer knowledge and has been widely used in computer vision field [19-21].

However, deep transfer learning technique has not received much attention in the field of spectral analysis. Most studies perform task analysis at pixel level based on remote sensing images, such as, poverty mapping [22], image superresolution processing [23] and crop yield prediction [24]. For ground spectral image, [25] showed the effectiveness of deep transfer learning in predicting soil clay content in different soils. For seeds of different crops, although many differences were existed, there are also certain similarities, for example the seeds structure. Most seeds contain seed coat, embryo and endosperm. These structures contain some conventional chemical components like starch, fat and enzymes which are necessary for seed to develop

into seedling [26-27]. This may lead to the similarities among the spectral characteristics of different crop seeds. Therefore, when constructing a deep model for seed varieties classification of a certain crop based on HSI, the knowledge in the model has the possibility to be transferred into the classification tasks of other crop seeds. In this study, we aimed to investigate the feasibility of deep transfer learning technique for varieties identification of different crop seeds based on HSI.

The specific objectives were: (1) to develop a deep network model with excellent performance based on a sample-rich dataset; (2) to transfer common knowledge to the varieties classification of other crop seeds with sample-limited datasets through the deep network; (3) to evaluate the impact of training set size on the performance of transfer learning; (4) to visualize the training and transferring process of deep network and the classification results. We hope to provide a common framework for accurate varieties identification of crop seeds under sample-limited condition through this study.

Materials and methods

Sample collection and dataset description

This study investigated five kinds of crop seeds, including pea, rice, oat, wheat and cotton. The seeds were performed hyperspectral imaging immediately after collected. All images were obtained by a same line-scanning near-infrared hyperspectral imaging system covering a spectral range from 874.41 nm to 1733.91 nm with a resolution of 5 nm [28]. A hyperspectral image containing 256 spectral channels could be obtained through each scan by this system and then calibrated using the following formula to reduce the negative impact of dark current.

$$I_c = \frac{I_r - I_d}{I_w - I_d} \quad (1)$$

where I_r and I_c represented the raw hyperspectral image and the corrected image, I_w and I_d represented the white and dark reference image. Each seed in hyperspectral image was regarded as a region of interest (ROI). To get the mask of all the ROIs, the channel with strongest contrast between background and seeds was conducted a simple threshold segmentation and morphological operation. Then the spectral vectors of all pixels within each ROI were extracted and the bands in head and end range were removed to avoid noise introduced by instability of the system. The reserved spectra with a range of 975~1646 nm were further

processed by wavelet transform (WT). The spectrum vector representing a seed sample was finally obtained by averaging all the transformed pixel spectra in one ROI.

Five spectra datasets with similar but different distributions were established in this study. Their detailed collection parameters and description information were summarized in Table 1. It should be noted that different parameters were set for imaging different crop seeds clearly since different seeds have different external phenotypes such as size, height and color. The most abundant dataset, Pea dataset, contained a total of 10420 samples from four varieties named Baiyan (2697), Heiyan (2848), Changshouren (2849) and Zhewan 1 (2026) which were widely cultivated in southern China. Peas of the first two varieties generally need to be roasted before eating, while the latter two can be directly eaten due to the high water and sugar content. All the seeds were purchased from Lvfeng seed company in Hangzhou, Zhejiang, China in 2018. The sub dataset corresponding to each variety was randomly divided into a training set, a validation set and a testing set at a ratio of 3:1:1. Then the sub datasets with same class were merged and shuffled. Because of its large volume of data, Pea dataset was selected as source dataset.

Table 1 Description of the datasets

Datasets		Par.	#Variety	#Total	#Training	#Validation	#Testing
Source	**Pea**	(15.5, 3, 12)	4	10420	6252	2084	2084
Target	**Rice**	(9, 3, 11)	3	750	150	300	300
	Oat	(15.2, 3, 11.5)	4	1000	200	400	400
	Wheat	(15, 3, 13)	5	1250	250	500	500
	Cotton	(14, 3, 11.5)	7	1750	350	700	700

Note: Par. represents parameters of hyperspectral imaging system including the distance between camera and seed plate (cm), the exposure time of the camera (ms), and the speed along the X-axis of seeds movement ($mm.s^{-1}$).

The other four sample-limited datasets were used as target datasets which involved several common staple crops. These datasets were designed to contain different numbers of seed varieties for investigating its impact on the transferring effect. Each variety in these datasets contains 250 samples and was further divided into three subsets at a ratio of 1:2:2 to reflect sample-limited condition.

The first target dataset only consisted 750 spectral samples of three varieties of rice seeds including Yongyou 9, Nuoyou 6211 and Zhongbaiyouhuazhan. These varieties are all hybrid rice with indica property

and belong to hybrid indica-japonica, hybrid indica-glutinous and hybrid indica rice, respectively. All seeds were collected by College of Agriculture and Biotechnology, Zhejiang University in 2019.

The second dataset was Oat dataset with the same number of varieties as the source dataset. It contained 1000 samples from four varieties named Bayan 6, Dingyan 2, Muwang and Jizhangyan 4 which were widely planted in the grasslands of northern China. The seeds harvested in 2017 were kindly provided by Academy of Agricultural and Animal Sciences, Inner Mongolia, China.

A total of 1250 samples from five varieties of wheat seeds, including Zhenmai 9, Annong 1124, Long-pingmian 6, Shannong 102 and Weilong 169, formed Wheat dataset. These five varieties were extensively cultivated in the winter wheat regions of southern China. The seed samples were friendly provided by Anhui longping high-tech seed industry co., LTD. in Hefei, Anhui, China in 2018.

The fourth dataset, Cotton dataset, was consisted by 1750 samples of seven varieties of cotton seeds. They were Jinxin 5, Jinxin 7, Shennongmian 1, Xinjiangzaomian 1, Xinluzaomian 29, Xinluzhong 52 and Xinluzhong 42. These varieties were mainly grown in Xinjiang Uyghur Autonomous Region, the largest cotton-producing region in China. And the cotton seeds were collected by Shihezi University in 2016.

In this study, multiple deep neural networks with different structures were firstly developed using the source dataset. Then the optimal deep model was selected as the model to be transferred through comparing the classification accuracies. Transfer learning technique was investigated to transfer useful knowledge from the optimal deep model to the analysis of four target datasets. The training set of each target dataset were further transformed into 10 datasets to analyze the impact of sample size on transfer learning by randomly selected 10 %~100 % samples from original training set. Four commonly-used multivariate analysis methods, including two linear methods: linear discriminant analysis (LDA), PLS-DA, and two nonlinear methods: multi-layer perceptron (MLP) and support vector machines (SVM) were introduced as the benchmarks.

Deep classification models development

In computer vision field, the huge image library, ImageNet, has spawned many excellent deep learning models like VggNet, InceptionNet and ResNet [29]. The specialness of VggNet is using small convolution kernels. The designers believed that using multiple convolution layers equipped with a 3×3 kernel to replace a convolution layer with a 5×5 kernel could not only reduce the network parameters, but also increase non-linear mapping, thereby increasing the representation capability [30]. ResNet is also an outstanding network

with many variations. What makes it special is the introduction of residual learning idea. The residual module directly bypasses the input of a certain layer to the output, which makes ResNet only need to learn the residual between the input and output [31]. This manner solves the problem of performance degradation when the network depth increases. InceptionNet was born in the ILSVRC2014 competition. The biggest innovation of this network is introducing a module called Inception to replace the common structure of convolution layer cascading pooling layer [32]. This Inception module contains four branches with different receptive fields to perceive the input patterns. By utilizing this module, InceptionNet can increase its width and learn more local features of different scales.

Inspired by these network structures, three one-dimensional deep neural networks were developed for the source dataset in this study, as shown in Fig. 1.

The first one was VGG-MODEL. Two V blocks containing two convolution layers equipped with a 1×3 kernel were designed to extract the feature patterns hidden in the spectral vectors. A Batch Normalization (BN) and an activation function exponential linear unit (ELU) were inserted after each convolution for the purpose of reducing the overfitting risk and speeding up the convergence process. The number of convolution filters was set to 16 for the first V block, and 32 for the second V block. A max pooling layer was placed behind each V block to reduce feature dimension. The last max pooling layer was followed by a Flatten layer to convert its output features into a one-dimensional vector form. Layer Fc1 and Fc2, consisting of 64 and 4 neurons respectively, were used to perform classification task like traditional neural networks. BN and ELU were also used behind Fc1. VGG-MODEL finally output the probability of the input spectral vector belonging to each category through a softmax function.

The second one was RES-MODEL. The first part of this network was similar to half of V block, which contained a convolutional layer followed by BN, ELU and a max pooling layer. The difference was that the convolutional layer used 32 kernels with a size of 1×7. The second part consisted of four cascaded residual modules, R block. This module was similar to the V block but added a transmission channel from input to output. The number of convolutional filters in the first R block was 32, and was doubled when the blocks going deeper. An average pooling layer was placed after the last R block to average the features in the spectral dimension. This layer could decrease the parameters in fully connected layers, thereby reducing over-fitting

risk. The last part of RES-MODELDE was similar to that of VGG-MODEL, but just equipped with one fully connected layer Fc1 with 4 neurons.

The third one was INCEPTION-MODEL. Having the same structure as that of RES-MODEL, the first part of this network utilized 16 convolution filters with a size of 1×3. It was followed by four I blocks, each of which cascaded a max pooling layer except the last one. The number of filters in the first I block was 16, and was also doubled when the blocks going deeper. As shown in Fig. 2, the I block transmitted its input to four parallel branches. Three of them were convolution layers with 1×1, 1×3, 1×5 kernels, respectively. They were employed to extract local spectral features at different scales. A 1×1 convolution was placed before 1×3 and 1×5 convolution to reduce the number of input channels. The last branch performed the max pooling operation. The introduction of I block allowed INCEPTION-MODEL to increase its width and obtain richer representation information without slowing down running speed. The end of the INCEPTION-MODEL was similar to that of RES-MODELDE.

To fairly compare the performance of these three deep models, all networks employed cross-entropy as objective function and used stochastic gradient descent (SGD) optimization algorithm. The learning rate and momentum of these networks were all set to 0.001 and 0.9, respectively. After debugging for many times, the number of samples input into the network at one time, *batch_size*, was set to 128, and the number of training iterations, *epoch*, was set to 400. All networks were trained using the training set of the source dataset, and the best models with the highest accuracy on the validation set were saved to evaluate the effectiveness of the corresponding network on the testing set. The detailed parameters of these three networks were shown in Additional file 1: Table S1.

Transfer learning strategy

As an emerging tool in machine learning, transfer learning was proposed to remit the requirement of models for sufficient training data by transferring available knowledge from relevant source domain to target domain [18]. The common process of transfer learning was shown in Fig. 3a. We defined a domain $\mathcal{D} = \{\mathcal{X}, \mathrm{P}(\mathcal{X})\}$ where $\mathcal{X}$ represented a feature space and $\mathrm{P}(\mathcal{X})$ represented its probability distribution, and a task $\mathcal{T} = \{\mathcal{Y}, f(.)\}$ where $\mathcal{Y}$ represented a label space and f represented a transformed function. When the task $\mathcal{T}$ was performed in the domain $\mathcal{D}$, f modeled $\mathrm{P}(y|x)$, where $y \in \mathcal{Y}, x \in \mathcal{X}$. In transfer learning field, there

are two domains: source domain $\mathcal{D}_S$ with task $\mathcal{T}_S$ and target domain $\mathcal{D}_T$ with task $\mathcal{T}_T$. The main goal of transfer learning is to improve the performance of transformed function in target domain $f_T(.)$ using the knowledge learned in $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{D}_S$ (or $\mathcal{T}_S$) and $\mathcal{D}_T$ (or $\mathcal{T}_T$) are different but relevant, that is, $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$.

For deep transfer learning, $f(.)$ is various deep models designed for specific tasks. These deep models that take a large amount of data and time to train contain rich knowledge. Some knowledge is closely related to the specific task, while other is some common knowledge which can be shared between different tasks or objects. Deep transfer learning aims to transfer the common knowledge into current target task to avoid collecting a large number of samples to learn this knowledge repeatedly, and thus can achieve rapid modeling. The model's structure and the network's weight are two important kinds of knowledge contained in the deep model. In this study, the optimal deep model structure based on the source dataset was reused to simplify and shorten the modeling process as much as possible. Since the number of seed varieties of different crops was different, the number of neurons in the output layer of the model was modified correspondingly. As the initial weights have a great influence on the convergence speed and the final performance of the network, this study transferred the weights of the optimal deep model based on the source dataset to the models based on the target datasets according to the network structure. It should be noted that the weights of the last fully connected layer in deep models based on the Rice, Wheat and Cotton dataset needed to be randomly initialized because the number of output neurons in these models were different from that in models based on source Pea dataset. During the transferring process, the weights of the layers before the flatten layer were frozen, and the target dataset were used to fine tune the subsequent full connected layers (Fig. 3b). The first reason was that the target dataset was too small to retrain the entire network. The second reason was that the convolutional layers before the flatten layer had extracted important feature patterns of the seed spectra, which could be reused in the target tasks. According to the size of the target datasets, the *batch_size* of the transferred network was set to 3, and the learning rate was set to 0.0001. The other configurations were the same as those of the deep network based on the source dataset.

Comparison methods

In this study, the deep neural networks based on the source dataset and four target datasets were compared with conventional linear and nonlinear multivariate analysis methods to confirm their validities in spectra analysis from both data-rich and data-poor sides.

LDA aims to find an optimal projected direction for raw variables. In the projected feature space, samples between classes hold maximal dispersion degree, while samples within classes hold minimal dispersion degree [33]. This projection manner facilitates transforming the samples into a linear separable state. The number of variables in the projected space, *n_lda*, is the only parameter which needs to be adjusted. We set *n_lda* to 1~20, and selected the optimal *n_lda* according to the classification performance of LDA.

The core principle of PLS-DA is also to conduct a linear transformation. Unlike LDA, the transformed latent variables (LVs) can not only carry the main information hidden in the raw variables, but also maximize the correlation between the independent and the dependent variables [34]. In spectral analysis, the number of LVs, *n_pls-da*, that minimize the sum of prediction residual error was usually selected. The range of *n_pls-da* was also set to 1~20 in this study.

SVM can enable raw linear unseparable variables to transform into a linear separable space through a nonlinear kernel function [33]. Because of the ability to cluster samples with same categories closely and make them tend to be linear separable, radial basis function (RBF) kernel was often used with SVM in many spectral analysis tasks. In this study, SVM equipped with RBF kernel was introduced as a nonlinear benchmark. Two parameters, penalty coefficient c and kernel parameter g were set to {10, 100, 1000, 10000} and {0.1, 0.01, 0.001, 0.0001}, respectively.

MLP is a fully connected artificial neural network with one or more hidden layers [35]. To obtain optimal performance, a total of 32 structures were attempted to process the source dataset, which contained 1~4 hidden layers, and each was equipped with 8 configurations for nodes in hidden layers. The number of nodes in hidden layers of the structure with 4 hidden layers was set to {(200-100-50-25), (180-90-45-23), (160-80-40-20), (140-70-35-18), (120-60-30-15), (100-50-25-13), (80-40-20-10), (60-30-15-8)}, and was simplified as the number of hidden layer decreases. For the target datasets, 24 same structures with 1~3 hidden layers were tried to get the optimal classifier.

Parameters of all models in this study were adjusted toward optimal states using corresponding validation set. All models were coded using python language in Spyder 3.2.6 environment (Anaconda, Austin, TX,

USA). The famous machine learning library, Sklearn (https://scikit-learn.org/stable/), was introduced to implement the conventional models, and the popular deep learning framework, Keras, was employed to program deep models. A Win10 64-bit operating system with Inter (R) Core(TM) i5-8500 CPU and 8GB RAM constituted the basic platform.

Model Visualization

Model visualization is significant for intuitively understanding the decision-making mechanism and clearly showing the computational result. In this study, visualization techniques were investigated from the perspective of training process of the deep model and classification results of the crop seeds. The accuracy and loss of the optimal deep model were exhibited to examine the convergence state. In addition, the raw seed spectra in different datasets and the feature representation of different layers in optimal deep model were extracted and their distributions were then expressed by t-distribution stochastic neighbor embedding (t-SNE). As an effective method for high-dimensional data visualization, t-SNE converts the similarity between sample points in high-dimensional space into Gaussian joint probability form, and constructs a similar probability distribution in low-dimensional space [36]. The ability of maintaining local structure of data is conducive for observing data patterns in low-dimensional space. Moreover, the advantage of hyperspectral imaging to obtain both spatial and spectral information was fully exploited. The sample's label predicted by the deep model was projected into the corresponding spatial position and represented by different colors to establish classification maps of crop seeds.

Results and discussion

Spectroscopic analysis

The average spectra with standard deviation of different varieties of seeds in five datasets was shown in Fig. 4. It could be obviously observed that these spectral curves possessed similar fluctuation patterns and locations of peaks and valleys. The absorption bands at approximately 1119.45 and 1206.92 nm were caused by the second overtone of carbohydrates (C–H stretch) [37]. The peak near 1307.97 nm (in the range of 1254 ~1348 nm) was reported to be associated with the combinations of the first overtone of amide B (N–H stretch) and the fundamental vibrations of amide II and III (C–N stretch and N–H in-plane bend) [38]. The band at 1469.95 nm (in the region of 1410~1502 nm) could be attributed to the first overtone of Amide A (N–H

stretch), which might be the key band for protein detection [38-39]. The similar chemical components contained in the crop seeds leaded to the similarities between the spectral curves. This also meant that certain common features were hidden in the spectral information of different crop seeds, which laid a foundation for effective transfer learning.

However, for seeds of different varieties of the same crop, there were also some heterogeneities between their spectral curves due to the content difference of chemical components. For example, the spectral curves of four varieties of pea seeds in the source dataset were naturally divided into two groups: Baiyan and Heiyan formed one group while Changshouren and Zhewan 1 formed the other one. This was consistent with the classification results according to edible manner which results from the difference of sugar and water content. In addition, for the Rice target dataset, it was because of the introduction of japonica characteristic that the reflectance of variety Yongyou 9 was quite different from the other two varieties. [40] also confirmed spectra differences were existed between different varieties of rice seeds. [41] found the optical characteristics of different varieties of hybrid okra and luffa seeds were very different and the metabolic analysis results showed that the components content of different seeds varied greatly. The heterogeneity of the spectral features between different varieties could be utilized as the basis of using hyperspectral imaging to identify different varieties of crop seeds.

Identification results analysis on source dataset

The accuracies and the optimal parameters of all models on training set and testing set of the source dataset were summarized in Table 2. The over-fitting phenomenon for all models was not serious due to the large-scale training set which might contain the spectral patterns in the testing set.

The accuracies of three deep models on the testing set were all above 99 % which was higher than most conventional methods. Owing to the convolution operation, the deep models were able to extract much discriminative information hidden in the raw redundant spectral data. Their performance superiorities were predictable. VGG-MODEL with accuracy of 99.57 % on the testing set was slightly conspicuous than the other two models. It was reported in computer vision field that the difficulty of improving performance increases with the model performance being better. For example, in 2014 ILSVRC competition, a 22-layers Inception-Net won the championship with a top-5 error rate of 6.7 % which was only 0.6 % lower than the runner-up,

VGGNet with 19-layers structure [32]. In addition, the high version of Inception, Inception-v4, achieved a top-5 error rate of 3.08 % which was only 0.42 % lower than the previous version, Inception V3 [42].

Table 2 Identification accuracies and optimal parameters of all models on source dataset

Methods	Par.	Training (%)	Testing (%)
VGG-MODEL	(16, 32, 128, 201)	99.98	99.57
RES-MODEL	(32, 32, 64, 64, 128, 194)	99.76	99.14
INCEPTION-MODEL	(16, 32, 64, 128, 256, 349)	100	99.09
LDA	(1)	99.39	98.90
PLS-DA	(20)	87.14	86.90
SVM	(104, 10-3)	99.70	99.28
MLP	(200, 100, 50, 25)	93.81	93.52

Note: Par. represents (number of major convolution filters, best *epoch*) for deep models, (*n_lda*) for LDA, (*n_pls-da*) for PLS-DA, (*c*, *g*) for SVM, (number of nodes in hidden layers) for MLP.

Since the structures of the three deep models were continuously adjusted to the optimal states according to the source dataset, they possessed different depths. In this study, INCEPTION-MODEL and RES-MODEL had a deeper structure than VGG-MODEL. In general, the deeper the model is, the richer the extracted features are. But this was based on a very large dataset like ImageNet and it should be guaranteed that the gradient would not disappear during model training. [43] developed a network with inception structure named Deepspectra which showed better performance than a comparison network, Model3, similar structure to VGG-MODEL. However, the authors also pointed out that the superiority of deep model was not in processing small dataset. Model 3 in [43] couldn't learn enough effective patterns from very little information. They also indicated that the performance of Model 3 improved significantly when the size of sample set increased slightly. The source dataset in this study was much larger than all the datasets in [43] and was enough for VGG-MODEL training. For structure like ResNet, [44] compared the classification accuracy of a developed ResNet with a general deep convolutional neural network on cotton datasets, and also found that ResNet was not as effective as the latter one.

The structure of the optimal model for a specific dataset was the result of constant tradeoff and adjustment, which was closely related to the size and distribution of the sample set. A complex deep network couldn't always obtain higher performance than a simple one. In this study, for the source dataset, VGG-MODEL with the simplest structure and the shallowest depth got a small victory when faced the relatively complex INCEPTION-MODEL and RES-MODEL. For conventional models, SVM performed best and its accuracy on the testing set was second only to VGG-MODEL. But different from our common sense that nonlinear models generally performed better than linear models, nonlinear model MLP achieved lower accuracy than linear model LDA in this study [41, 45-46]. Thus, many models need to be tried to determine which is the optimal one if we use traditional multivariate analysis methods [45-48]. Conversely, deep models will generally achieve satisfactory results if the training data is sufficient and the structure is properly designed. In the field of spectral analysis, deep learning is a very competitive and potential tool.

Identification results analysis on target datasets

Although deep learning might not work well on small dataset, its advantages would carry forward again after combining with transfer learning. To verify the effect of deep transfer learning, the slightly better-performing VGG-MODEL was used as the source model to be transferred in this study. 10 training sets with different sizes were built based on the original training sets to investigate the influence of training set size on the transferring effect. The identification results of the deep transferred model and the comparison methods were showed in Fig. 5.

It could be seen that the deep transferred model was the only model that always performed well on the four datasets. For the 100 % training set which was still very small compared with the training set of the source dataset, the deep model achieved accuracies of 95 %, 99 %, 80.8 %, and 83.86 % on the testing set of Rice, Oat, Wheat and Cotton dataset, respectively. Such good results could not be obtained if these small datasets were used to build deep models directly. It was because of combination with transfer learning that deep learning model could also obtain satisfactory results on small datasets. Transfer learning enabled deep learning to take advantage of itself, and also avoided the requirement for a large number of samples [19]. As similar patterns were existed among the spectra vectors of different crop seeds, varieties identification of different crop seeds belonged to different but similar tasks in the same domain. Thus, transfer learning was very suitable for the spectral analysis in this study.

Although other models could also achieve good performance on some datasets, they could not always perform well on all. For example, for the 100 % training set, LDA model achieved an accuracy of 87.71 % on the Cotton dataset which was even higher than that achieved by the deep transferred model. However, it just obtained accuracies of 93.33 %, 94.5 %, and 71.2 % on the Rice, Oat and Wheat datasets. SVM performed relatively stable on the four datasets like performed in [40-41, 46], and achieved accuracies of 90 %, 96.75 %, 76.6 % and 75.86 % when using the 100 % training sets. As expected, as a shallow neural network, MLP performed much worse than the deep neural network, especially on the Wheat dataset. It only got an accuracy of 32.4 % on the testing set which was just a little better than random guessing. Shallow neural network could not extract valuable discriminative information from redundant spectral data, which led to unsatisfactory results [49]. PLS-DA commonly used in spectral analysis was also very unstable. Although it could obtain an accuracy of 92.25 % on the Oat dataset, it performed rather badly on the Wheat and Cotton datasets which contained more varieties. This was consistent with the results in [41]. With the increasing of the number of seed varieties, the possibility of samples being linearly separable became smaller, and the difficulty of distinguishing different varieties became greater. In a word, for traditional multivariate analysis methods, different datasets might correspond to different optimal models. Conversely, deep transferred models could generally achieve satisfactory results.

In addition, it was worth mentioning that deep transferred model could also achieve good results when fine-tuned using very small datasets. For example, using 10 % training set which only contained 5 samples for each variety, it could achieve accuracies of 86.67 %, 88.74 % and 70.14 % on the Rice, Oat and Cotton dataset. And the accuracy rose rapidly with the increasing of training set size. Even on the Wheat dataset where all models failed, deep transferred model outperformed all the conventional methods. Deep transfer learning brings hope for scenarios with very limited samples. On the contrary, the accuracies of most conventional methods were very low when trained using such a small dataset, and the increased speed were relatively slowly. The result that LDA got a high classification accuracy of 93 % on the Oat dataset was unexpected. This could be because this little training set just fitted the classification rule of LDA, since its accuracy dropped to 80 % soon for 30 % training set, and then slowly increased.

Moreover, it could be observed that almost models showed high accuracies on the Rice and Oat dataset but performed badly on the Wheat and Cotton dataset. The samples distribution of a dataset with few categories was generally simple. Contrarily, the distribution of a dataset with more categories was relatively complicated, which was not conducive to discrimination. Thus, dataset was an important factor affecting the performance of models [43, 50]. In addition, it could be seen that deep transferred model got the best performance in Oat dataset. Using just 20 % training set, it obtained an accuracy of 97.25 % on the testing set. Since the Oat dataset had the same number of varieties as the source dataset, all the weight parameters in the source deep model including the weights in fully connected layers could be transferred. This allowed maximum transferring of features in the source model.

Model Visualization

In order to inspect the convergence status, the accuracy and the loss curves of the training and validation set of five datasets (for the target datasets, the 100 % training set was used) were shown in Fig. 6. For the source Pea dataset, it could be seen that these two measurements on the training set and the validation set were basically the same and soon entered stable state with little fluctuation. According to the program design, the VGG-MODEL at epoch = 201 where the accuracy on the validation set reached maximized was saved and used for subsequent analysis. For target datasets, the accuracy and the loss curves changed into stable states with large fluctuations. Different from the source dataset, these two measurements on the validation set of the target datasets performed better than that on the training set. We guessed that it might be because of the small sample size of the training set. Similar to the source dataset, the deep models at epoch = 327, 93, 262, and 184 were saved for the Rice, Oat, Wheat and Cotton datasets.

Visualizing the feature distribution at each layer of deep network was another important channel to understand the training process of deep model [51-53]. In this study, t-SNE technique was used to visualize the original high-dimensional spectra and the features output by the Flatten, Fc1 and Fc2 layer in deep model in a two-dimensional space, as shown in Fig. 7 and Fig. S1~S4 in Additional file 2~5. For the Pea dataset, the original spectral samples were aggregated into two categories, which was consistent with the average spectral analysis. After passed the Flatten layer, the samples with easily confused categories like Baiyan and Heiyan, or Changshouren and Zhewan 1, were gradually became distinguishable. As the layers of the network deepened, the samples within a category were clustered closely, while those between different categories

429 became discrete. The samples were clearly gathered into four categories after output by the Fc2 layer. It could
430 be seen that the deep model gradually transformed the samples from a cluttered state to a distinguishable
431 state. This was why the deep model could obtain better performance than the traditional methods which used
432 the original spectral vectors directly.

433 For the four target datasets, the raw spectra in Rice and Oat dataset, especially in Rice dataset, were
434 slightly more regular than those in Wheat and Cotton dataset. The variety Yongyou9 was strongly distin-
435 guishable from the other two varieties. This was also consistent with previous spectral analysis. Thus, most
436 traditional methods performed better on the Rice and Oat dataset than on the other two datasets. Since all the
437 weights before the Flatten layer were transferred from the source deep model directly, the features output by
438 the Flatten layer in the deep transferred model contained the spectral patterns learnt from the source dataset.
439 From Fig. S1~S4 in Additional file 2~5, it could be seen that for the Rice, Oat and Cotton dataset, the features
440 output by the Flatten layer presented a more aggregated distribution pattern than the raw spectral samples.
441 This intuitively verified the effectiveness of transfer learning. Through transferring, the spectral features
442 learnt from the source dataset was reused, which facilitates the identification of small target datasets. The
443 Wheat dataset might be too cluttered so that the output of the Flatten layer didn't show distinguishability.
444 The target datasets began to work from Fc1 layer. The samples gradually showed strong separability with the
445 layers deepened. After output by the final Fc2 layer, the Rice and the Oat samples had been divided into 3
446 and 4 categories, respectively. Thus, the deep transferred model achieved two high accuracies of 95 % and
447 99 %. However, the Wheat and the Cotton samples had still some overlapping phenomenon, which led to
448 relatively low accuracies of deep transferred model.

449 The classification visualization of crop seeds was helpful for breeders to select varieties that meet re-
450 quirements and for market supervision authorities to check seed purity. In this study, the advantage of hyper-
451 spectral imaging to obtain spectral and spatial information simultaneously was fully exploited. The categories
452 of pea seeds predicted by the optimal model, VGG-MODEL, were visualized in a map, as shown in the Fig.
453 8. In the original hyperspectral images, variety Baiyan and Heiyan showed similar smooth texture features,
454 while variety Changshouren and Zhewan 1 showed rough texture due to water loss during drying process.
455 According to human vision, these four varieties were naturally divided into two categories, which was con-
456 sistent with the visualization analysis of the samples' distribution. Among the predicted 180 seeds, only 2

seeds with variety Heiyan were misclassified into similar Baiyan category. This accuracy was sufficient for variety selection during breeding process or purity detection in actual production.

Conclusion

In this study, we attempted to use hyperspectral imaging and deep transfer learning to achieve accurate and rapid varieties identification of crop seeds under sample-limited condition. The VGG-MODEL based on the sample-rich dataset stood out from three deep neural networks with typical structures and was utilized as the deep source model to be transferred. The transfer results on four small target datasets showed that deep transferred model could make full use of the common spectral features of crop seeds extracted by the source deep model and could still achieve better performance than traditional multivariate analysis methods under sample-limited condition, especially under the condition of very few samples. Giving a glimpse into the process of deep transfer learning, the visualization of the feature distribution at each layer of deep network further confirmed the portability of common spectral features, and revealed the reason why the deep network achieved high accuracy. The visualization of classification results provided an intuitive and convenient manner for varieties identification of crop seeds. In conclusion, hyperspectral imaging combined with deep transfer learning was a great potential tool to be applied in the identification of seed varieties with limited samples, which will greatly accelerate seed screening process in the fields with scarce samples, for example in breeding field. This study also provided a new idea for the detection of other qualities of crop seeds based on hyperspectral imaging under sample-limited condition.

Additional files

Additional file 1: Table S1. The detailed parameters of three deep neural networks.

Additional file 2: Fig. S1. Feature visualization of VGG-MODEL on Rice dataset using t-SNE.

Additional file 3: Fig. S2. Feature visualization of VGG-MODEL on Oat dataset using t-SNE.

Additional file 4: Fig. S3. Feature visualization of VGG-MODEL on Wheat dataset using t-SNE.

Additional file 5: Fig. S4. Feature visualization of VGG-MODEL on Cotton dataset using t-SNE.

Abbreviations

BN: Batch Normalization; ELM: extreme learning machine; ELU: exponential linear unit; HSI: hyper-spectral imaging; LDA: linear discriminant analysis; LSSVM: least square sup-port vector machines; LV: latent variables; MLP: multi-layer perceptron; PLS-DA: partial least squares discriminant analysis; PLSR: partial least squares regression; RBF: radial basis function; SVM: support vector machines; t-SNE: t-distribution stochastic neighbor embedding.

Declarations

Authors' contributions

All authors have made significant contributions to this study. NW, FL and YH conceived and designed the experiments. NW, ML and CZ performed the experiment and conducted data analysis. NW, FL interpreted data analysis. FL, YB, WH, FM and YH provided suggestions on the experiment design. YB and YH provided financial support. NW drafted the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by National Key R&D Program of China (2018YFD0101002).

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable.

Acknowledgements

Not applicable.

References

1. Ye S, Wang Y, Huang D, Li J, Gong Y, Xu L, Liu L. Genetic purity testing of F1 hybrid seed with molecular markers in cabbage (*Brassica oleracea var. capitata*). Sci Hortic. 2013;155:92-96.

2. Shuaib M, Zeb A, Ali Z, Ali W, Ahmad T, Khan I. Characterization of wheat varieties by seed storage-protein electrophoresis. Afr J Biotechnol. 2007;6:497-500.

3. Feng X, Peng C, Chen Y, Liu X, Feng X, He Y. Discrimination of CRISPR/Cas9-induced mutants of rice seeds using near-infrared hyperspectral imaging. Sci Rep. 2017;7:15934.

4. Sun D, Cen H, Weng H, Wan L, Abdalla A, El−Manawy AI, Zhu Y, Zhao N, Fu H, Tang J, Li X, Zheng H, Shu Q, Liu F, He Y. Using hyperspectral analysis as a potential high throughput phenotyping tool in GWAS for protein content of rice quality. Plant Methods. 2019;15:54.

5. Stuart MB, McGonigle AJS, Willmott JR. Hyperspectral imaging in environmental monitoring: a review of recent developments and technological advances in compact field deployable systems. Sensors. 2019;19(14):3071.

6. Shrestha S, Knapič M, Žibrat U, Deleuran LC, Gislum R. Single seed near-infrared hyperspectral imaging in determining tomato (*Solanum lycopersicum* L.) seed quality in association with multivariate data analysis. Sens Actuators B Chem. 2016;237:1027-1034.

7. Wu N, Jiang H, Bao Y, Zhang C, Zhang J, Song W, Zhao Y, Mi C, He Y, Liu F. Practicability investigation of using near-infrared hyperspectral imaging to detect rice kernels infected with rice false smut in different conditions. Sens Actuators B Chem. 2020;308:127696.

8. Caporaso N, Whitworth MB, Fisk ID. Protein content prediction in single wheat kernels using hyperspectral imaging. Food Chem. 2018;240:32-42.

9. Weng H, Lv J, Cen H, He M, Zeng Y, Hua S, Li H, Meng Y, Fang H, He Y. Hyperspectral reflectance imaging combined with carbohydrate metabolism analysis for diagnosis of citrus Huanglongbing in different seasons and cultivars. Sens Actuators B Chem. 2018;275:50-60.

10. Kong W, Zhang C, Huang W, Liu F, He Y. Application of hyperspectral imaging to detect *Sclerotinia sclerotiorum* on oilseed rape stems. Sensors. 2018;18:123.

11. Wei W, Zhang J, Zhang L, Tian C, Zhang Y. Deep cube-pair network for hyperspectral imagery classification. Remote Sens. 2018;10:783.

12. Yu X, Lu H, Wu D. Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflectance imaging. Postharvest Biol Technol. 2018;141:39-49.

13. Jin X, Jie L, Wang S, Qi HJ, Li SW. Classifying wheat hyperspectral pixels of healthy heads and *Fusarium* head blight disease using a deep neural network in the wild field. Remote Sens. 2018;10:395.

14. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436-444.

15. Lee H, Kim MS, Lim H, Park E, Lee W-H, Cho B-K. Detection of cucumber green mottle mosaic virus-infected watermelon seeds using a near-infrared (NIR) hyperspectral imaging system: application to seeds of the “Sambok Honey” cultivar. Biosyst Eng. 2016;148:138-147.

16. Xu J, Riccioli C, Sun D. Comparison of hyperspectral imaging and computer vision for automatic differentiation of organically and conventionally farmed salmon. J Food Eng. 2017;196:170-182.

17. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data 2016;3:9.

18. Pan SJ, Yang Q. A survey of transfer learning. IEEE Trans. Knowl. Data Eng. 2010;22(10):1345-1359.

19. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In: Proceedings of the international conference on artificial neural networks; 2018, p. 270-279.

20. Mohanty SP, Hughes DP, Salathé M. Using deep learning for image-based plant disease detection. Front. Plant Sci. 2016;7:1419.

21. Ghazi MM, Yanikoglu B, Aptoula E. Plant identification using deep neural networks via optimization of transfer learning parameters. Neurocomputing. 2017;235:228-235.

22. Xie M, Jean N, Burke M, Lobell D, Ermon S. Transfer learning from deep features for remote sensing and poverty mapping. In: Proceedings of the 30th AAAI conference on artificial intelligence; 2016, p. 3929-3935.

23. Yuan Y, Zheng X, Lu X. Hyperspectral image superresolution by transfer learning. IEEE J Sel Top Appl Earth Observ. 2017;10:1963-1974.

24. Wang AX, Tran C, Desai N, Lobell D, Ermon S. Deep transfer learning for crop yield prediction with remote sensing data. In: Proceedings of the 1st ACM SIGCAS conference on computing and sustainable societies; 2018, p. 1-5.

25. Liu L, Ji M, Buchroithner M. Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery. Sensors. 2018;18:3169.

26. Zhao M, Zhang H, Yan H, Qiu L, Baskin CC. Mobilization and role of starch, protein, and fat reserves during seed germination of six wild grassland species. Front Plant Sci. 2018;9:234.

27. Beníteza V, Canteraa S, Aguileraa Y, Molláa E, Estebana RM, Díazb MF, Martín-Cabrejasa MA. Impact of germination on starch, dietary fiber and physicochemical properties in non-conventional legumes. Food Res Int. 2013;50(1):64-69.

28. Wu N, Zhang C, Bai X, Du X, He Y. Discrimination of chrysanthemum varieties using hyperspectral imaging combined with a deep convolutional neural network. Molecules. 2018;23:2381.

29. Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th international conference on neural information processing systems; 2012, p. 1097-1105.

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

31. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016, p. 770-778.

32. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015, p. 1-9.

33. Gerhardt N, Schwolow S, Rohn S, Pérez-Cacho PR, Galán-Soldevilla H, Arce L, Weller P. Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: comparison of different processing approaches by LDA, kNN, and SVM. Food Chem. 2019;286:307-308.

34. Kandpal LM, Lohumi S, Kim MS, Kang JS, Cho BK. Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds. Sens Actuators B Chem. 2016;229:534-544.

35. Taud H, Mas JF. Multilayer perceptron (MLP). Geomatic Approaches for Modeling Land Change Scenarios; 2017, p. 451-455.

36. van der Maaten L, Hinton G. Viualizing data using t-SNE. J Mach Learn Res. 2008;9:2579-2605.

37. Wu N, Zhang Y, Na R, Mi C, Zhu S, He Y, Zhang C. Variety identification of oat seeds using hyperspectral imaging: investigating the representation ability of deep convolutional neural network. RSC Adv. 2019;9:12635-12644.

38. Daszykowski M, Wrobel MS, Czarnik-Matusewicz H, Walczak B. Near-infrared reflectance spectroscopy and multivariate calibration techniques applied to modelling coffee beverages using near infrared spectroscopy the crude protein, fibre and fat content in rapeseed meal. Analyst. 2008;133:1523-1531.

39. Ribeiro JS, Ferreira MMC, Salva TJG, Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. Talanta. 2011;83:1352-1358.

40. Qiu Z, Chen J, Zhao Y, Zhu S, He Y, Zhang C. Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. Appl Sci. 2018;8:212.

41. Nie P, Zhang J, Feng X, Yu C, He Y. Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning. Sens Actuators B Chem. 2019;296:126630.

42. Szegedy C, Loffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the impact of residual connections on learning; 2016. arXiv preprint arXiv:1602.07261.

603 43. Zhang X, Lin T, Xu J, Luo X, Ying Y. DeepSpectra: An end-to-end deep learning approach for quantita-
604 tive spectral analysis. Anal Chim Acta. 2019;1058:48-57.

605 44. Zhu S, Zhou L, Zhang C, Bao Y, Wu B, Chu H, Yu Y, He Y, Feng L. Identification of soybean varieties
606 using hyperspectral imaging coupled with convolutional neural network. Sensors. 2019;19(19):4065.

607 45. Liu D, Li Q, Li W, Yang B, Guo W. Discriminating forchlorfenuron-treated kiwifruits using a portable
608 spectrometer and Vis/NIR diffuse transmittance spectroscopy technology. Anal Methods. 2017;9:4207-4214.

609 46. Bao Y, Mi C, Wu N, Liu F, He Y. Rapid classification of wheat grain varieties using hyperspectral
610 imaging and chemometrics. Appl Sci. 2019;9:4119.

611 47. Zheng W, Fu X, Ying Y. Spectroscopy-based food classification with extreme learning machine. Chemo-
612 metrics Intell Lab Syst. 2014;139:42-47.

613 48. Teyea E, Anyidoho E, Agbemafle R, Sam-Amoah LK, Elliott C. Cocoa bean and cocoa bean products
614 quality evaluation by NIR spectroscopy and chemometrics: a review. Infrared Phys Technol.
615 2020;104:103127.

616 49. Chen Y, Lin Z, Zhao X, Wang G, Gu Y. Deep learning-based classification of hyperspectral data. IEEE
617 J Sel Top Appl Earth Obs Remote Sens. 2014;7:2094-2107.

618 50. Özdemir A, Yavuz U, Dael FA. Performance evaluation of different classification techniques using dif-
619 ferent datasets. Int J Elec & Comp Eng. 2019;9:3584-3590.

620 51. Zintgraf LM, Cohen TS, Adel T, Welling M. Visualizing deep neural networks decisions: prediction
621 difference analysis; 2017. arXiv preprint arXiv:1702.04595.

622 52. Zhang QS, Zhu SC. Visual interpretability for deep learning: a survey. Front Inform Technol Electron
623 Eng. 2018;19(1):27-39.

624 53. Lin TY, Maji S. Visualizing and understanding deep texture representations. In: Proceedings of the IEEE
625 conference on computer vision and pattern recognition; 2016, p. 2791-2799.

626

627

628 **Figure captions**

629 **Fig. 1**. The structures of three developed deep neural networks

630 **Fig. 2**. The inner structures of three typical blocks

631 **Fig. 3**. Transfer learning strategy. **a** The common process of transfer learning. **b** The deep transfer learning

632 strategy in this study

633 **Fig. 4**. The average spectra with standard deviation of five crop seeds

634 **Fig. 5**. Classification accuracies of all models on the four target datasets

635 **Fig. 6**. The accuracies and losses of VGG-MODEL on the training and validation set

636 **Fig. 7.** Feature visualization of VGG-MODEL on Pea dataset using t-SNE

637 **Fig. 8**. Classification maps of pea seeds

638

Figures

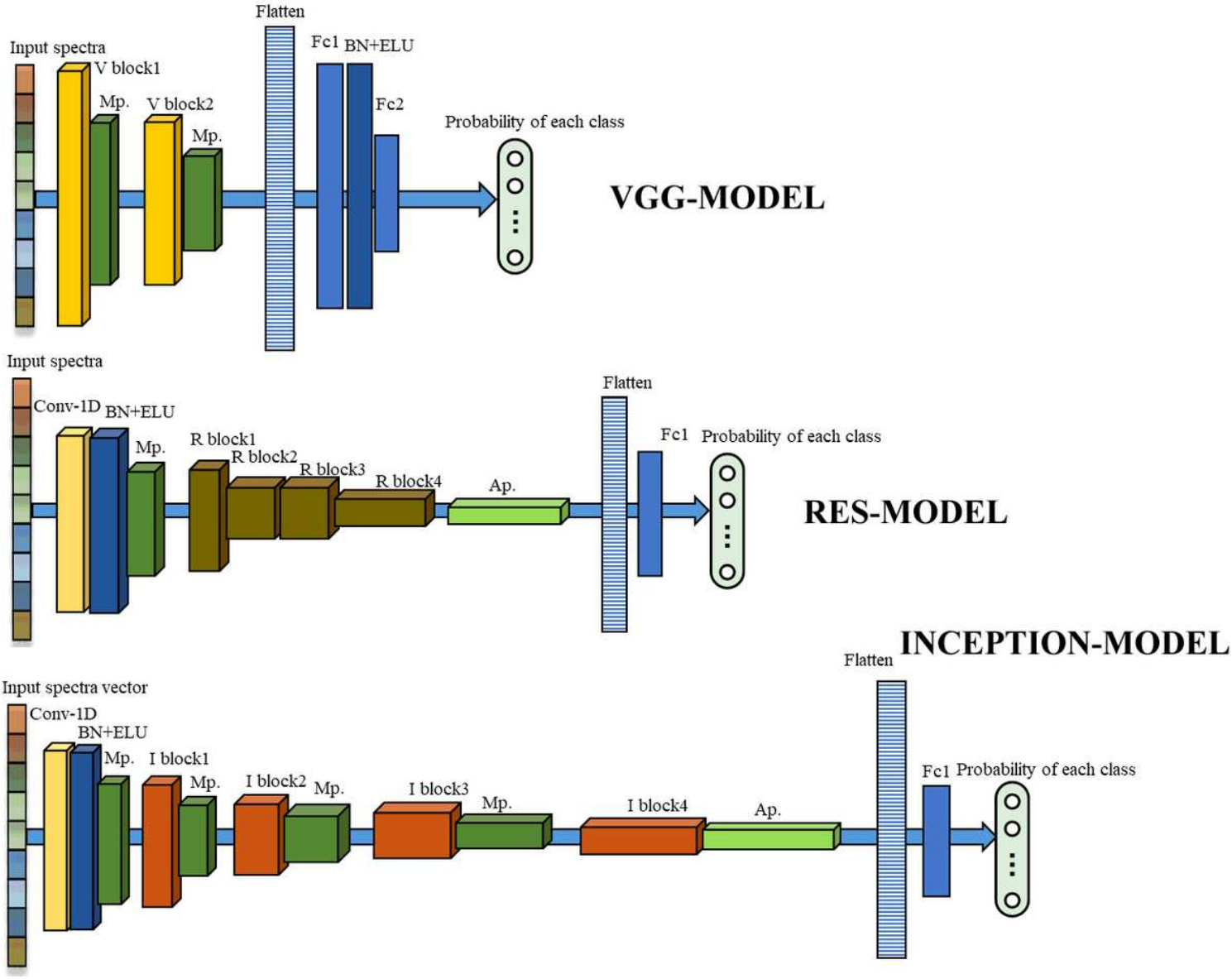

Figure 1

The structures of three developed deep neural networks

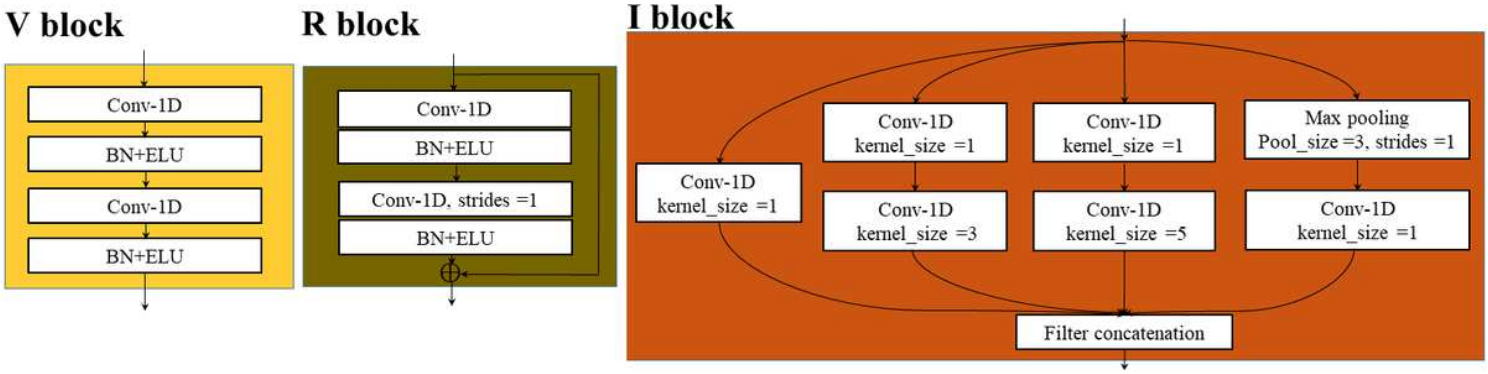

Figure 2

The inner structures of three typical blocks

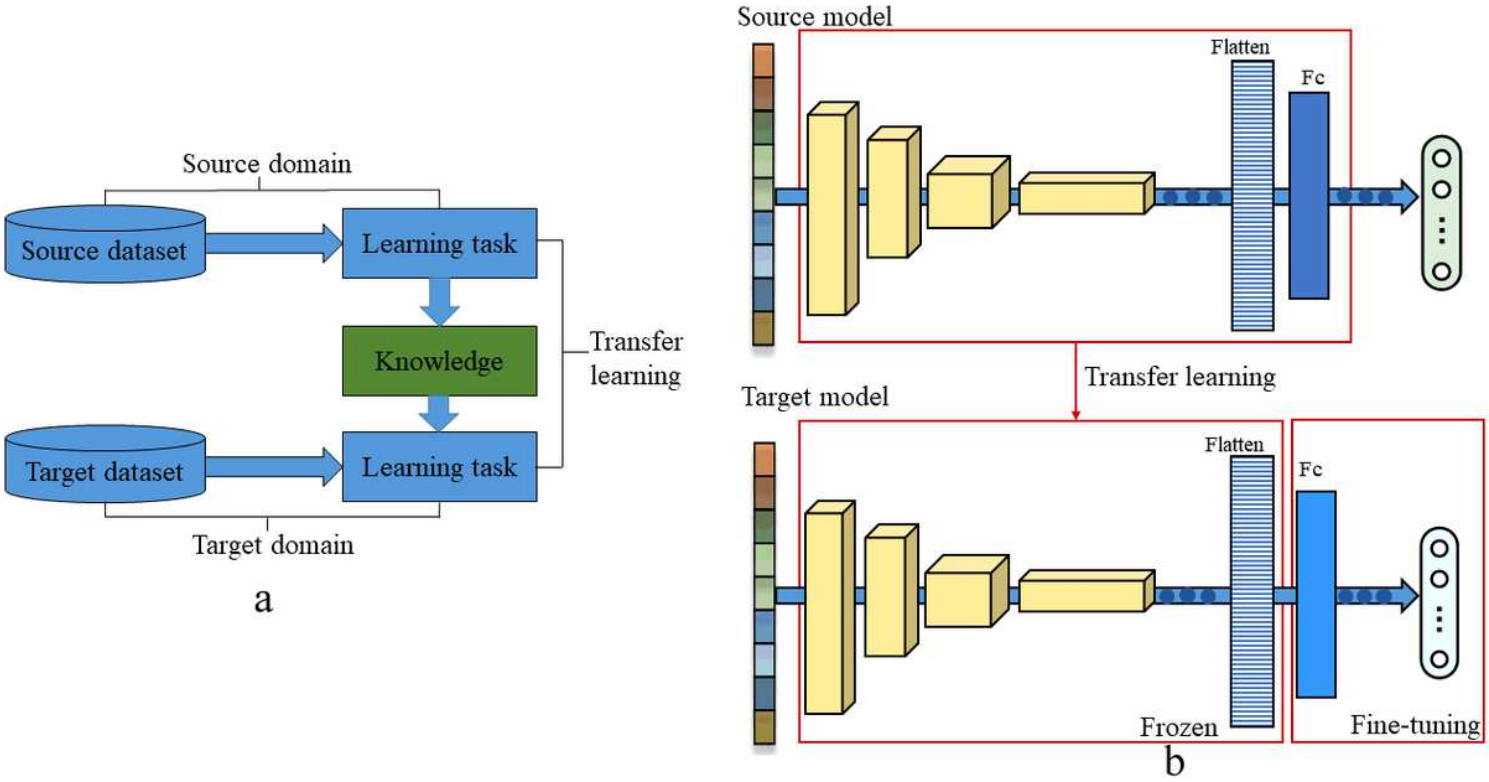

Figure 3

Transfer learning strategy. a The common process of transfer learning. b The deep transfer learning strate-gy in this study

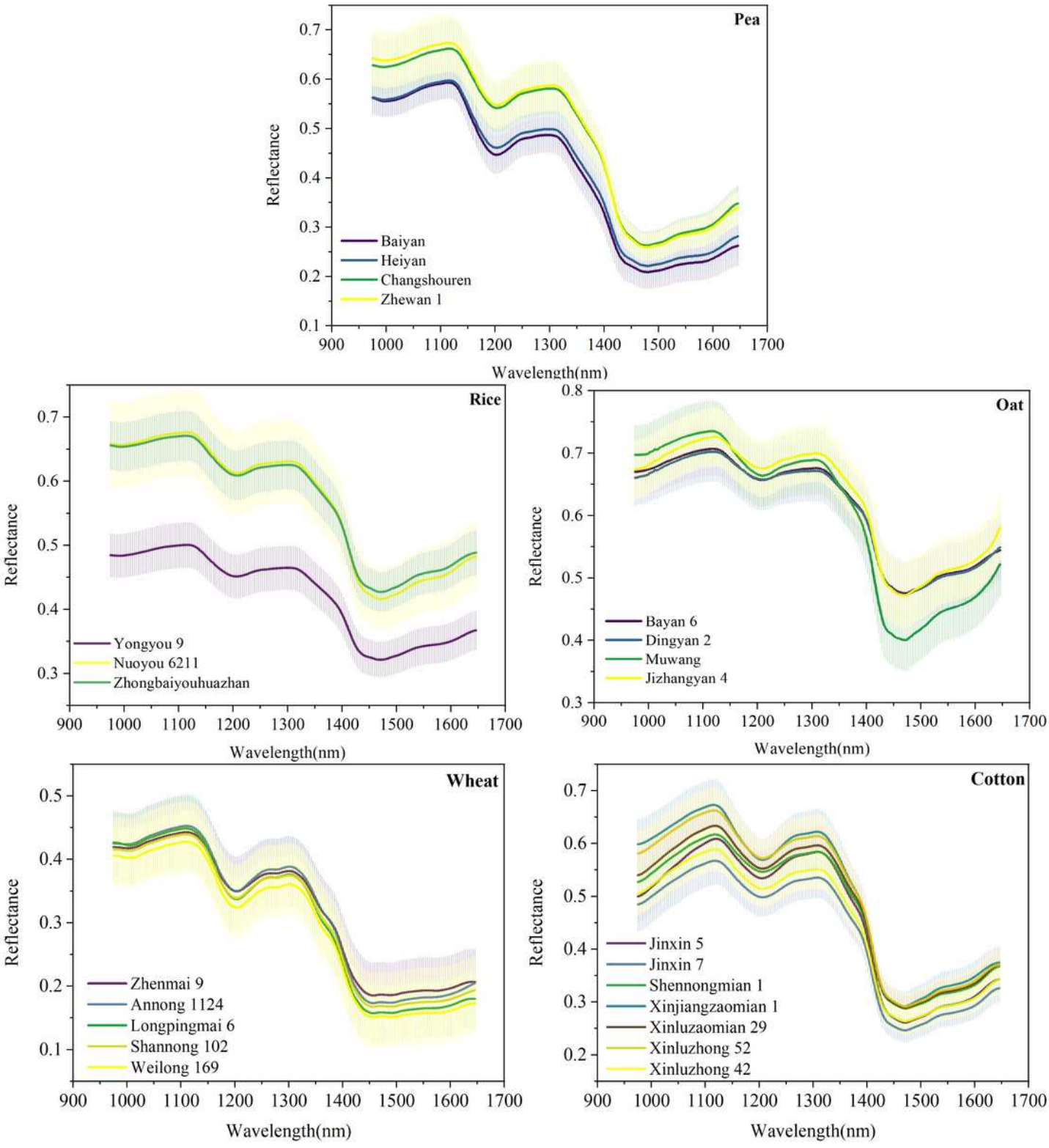

Figure 4

The average spectra with standard deviation of five crop seeds

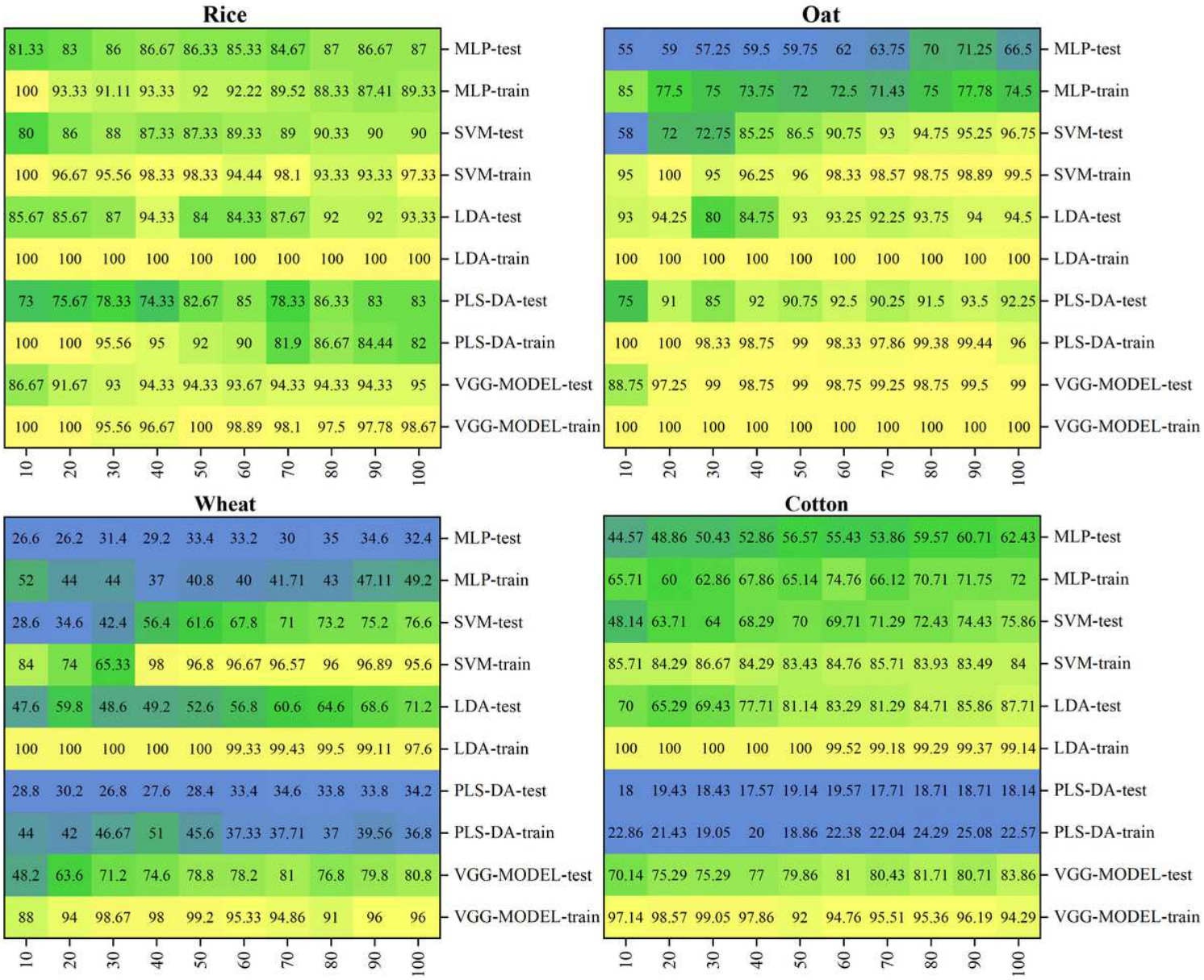

Figure 5

Classification accuracies of all models on the four target datasets

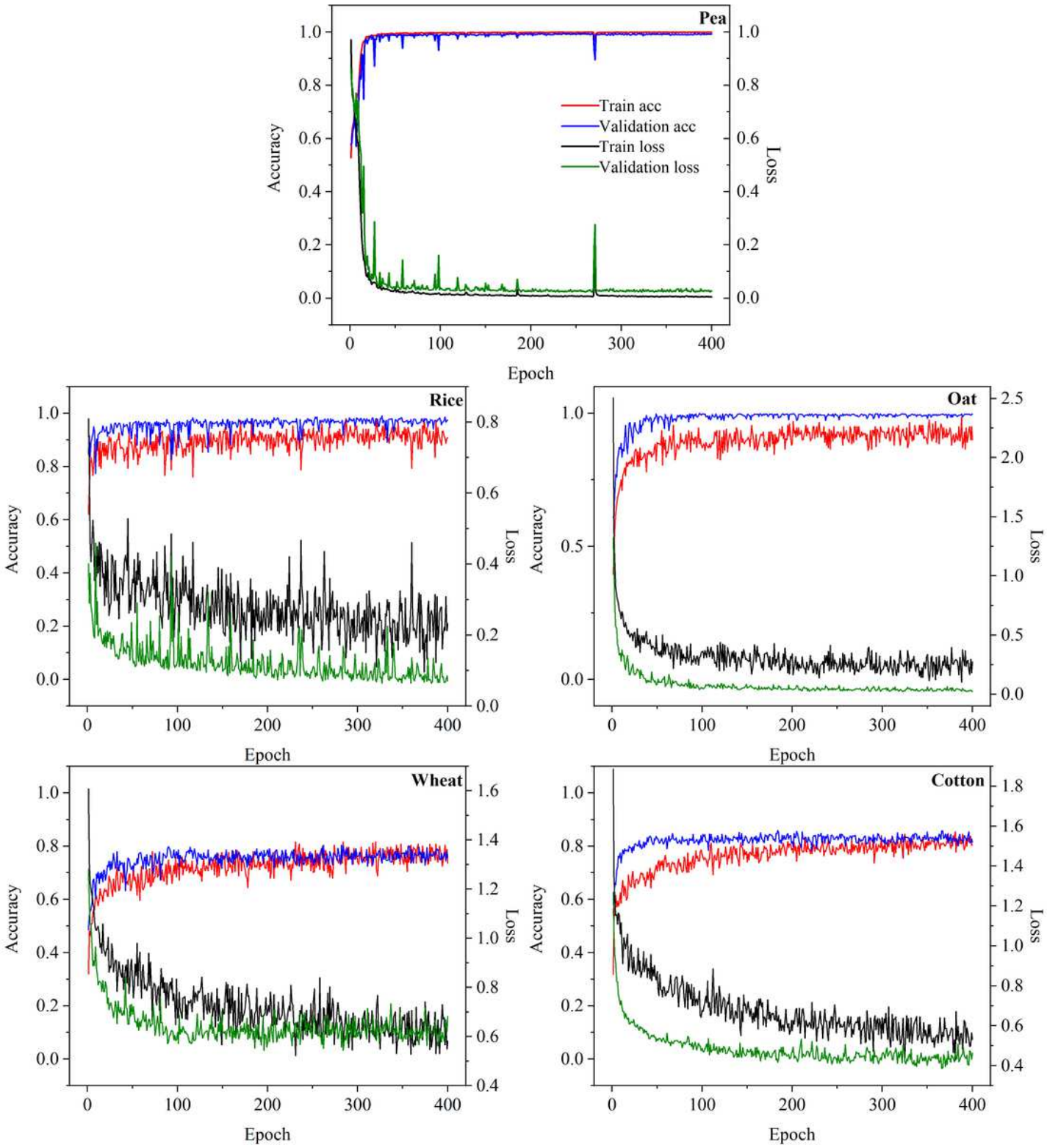

Figure 6

The accuracies and losses of VGG-MODEL on the training and validation set

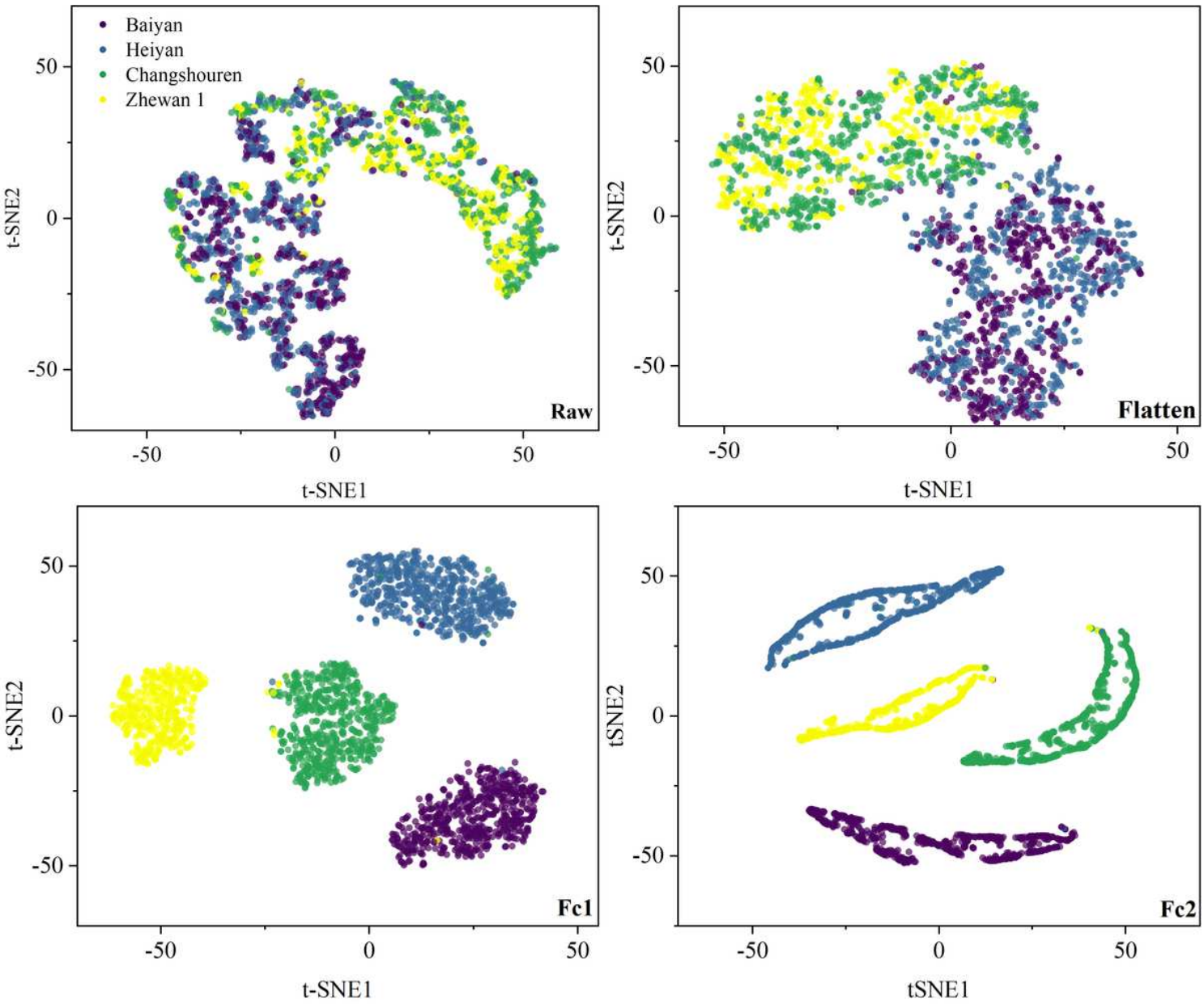

Figure 7

Feature visualization of VGG-MODEL on Pea dataset using t-SNE

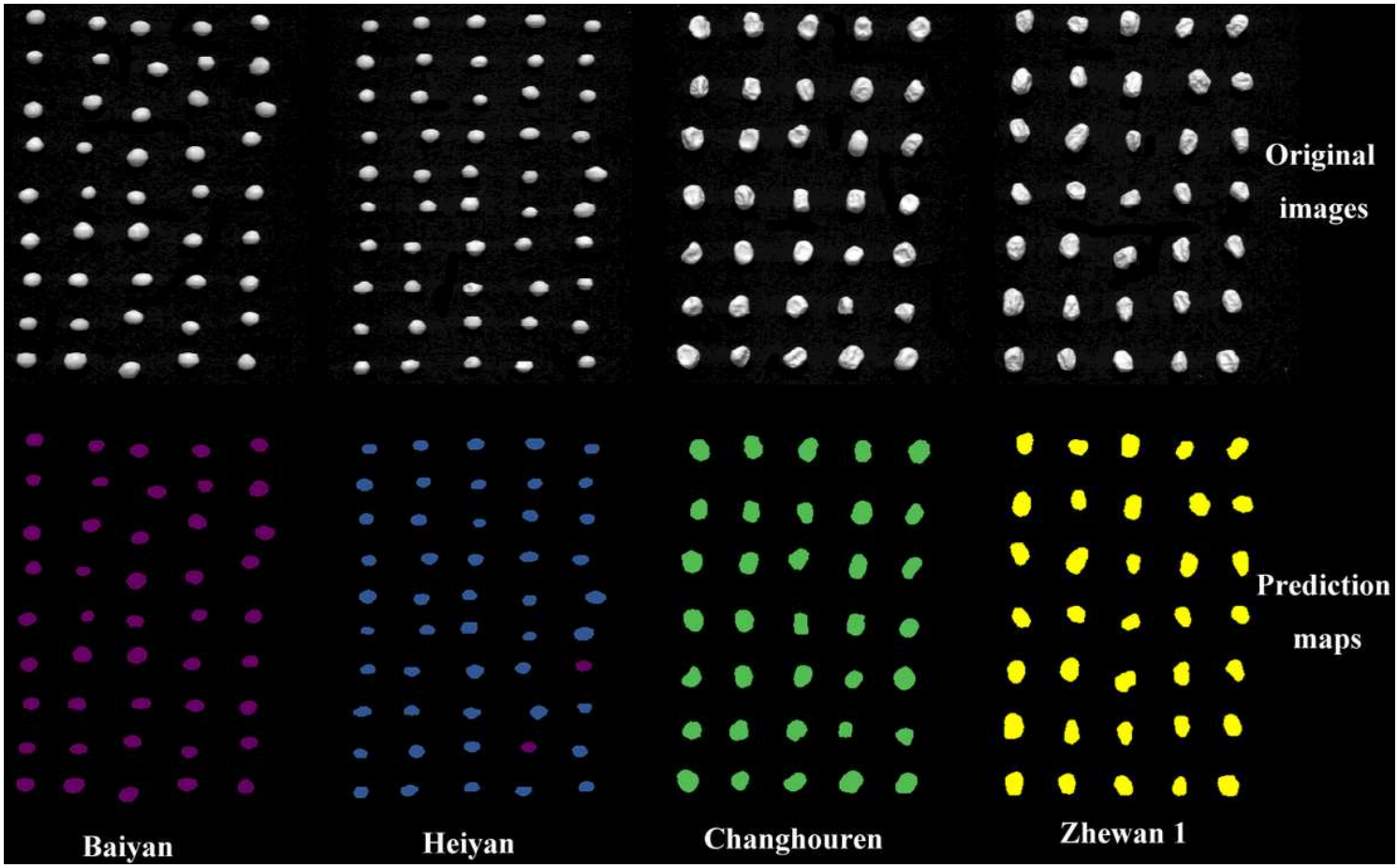

Figure 8

Classification maps of pea seeds

Supplementary Files

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

- Additionalfile1TableS1.docx
- Additionalfile2Fig.S1.tif
- Additionalfile3Fig.S2.tif
- Additionalfile4Fig.S3.tif
- Additionalfile5Fig.S4.tif