Article

Variety Identification of Single Rice Seed Using Hyperspectral Imaging Combined with Convolutional Neural Network

Zhengjun Qiu [1,2], Jian Chen [1,2], Yiying Zhao [1,2], Susu Zhu [1,2], Yong He [1,2] and Chu Zhang [1,2,*]

[1]   College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China; zjqiu@zju.edu.cn (Z.Q.); cjian@zju.edu.cn (J.C.); zhaoyy@zju.edu.cn (Y.Z.); sszhu@zju.edu.cn (S.Z.); yhe@zju.edu.cn (Y.H.)
[2]   Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture, Hangzhou 310058, China
*   Correspondence: chuzh@zju.edu.cn

Abstract: The feasibility of using hyperspectral imaging with convolutional neural network (CNN) to identify rice seed varieties was studied. Hyperspectral images of 4 rice seed varieties at two different spectral ranges (380–1030 nm and 874–1734 nm) were acquired. The spectral data at the ranges of 441–948 nm (Spectral range 1) and 975–1646 nm (Spectral range 2) were extracted. K nearest neighbors (KNN), support vector machine (SVM) and CNN models were built using different number of training samples (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500 and 3000). KNN, SVM and CNN models in the Spectral range 2 performed slightly better than those in the Spectral range 1. The model performances improved with the increase in the number of training samples. The improvements were not significant when the number of training samples was large. CNN model performed better than the corresponding KNN and SVM models in most cases, which indicated the effectiveness of using CNN to analyze spectral data. The results of this study showed that CNN could be adopted in spectral data analysis with promising results. More varieties of rice need to be studied in future research to extend the use of CNNs in spectral data analysis.

Keywords: hyperspectral imaging; variety identification; rice seed; convolutional neural network

1. Introduction

Rice is one of the most common food crops planted in China and some other countries. Rice seeds are harvested as food to consumers and as seed for the following sowing season. Due to the vast planting areas across villages, towns, countries and continents, different varieties of rice seeds are developed to adapt to changes of growth environments (climates, soil, water, etc.) and improve nutrition and flavors. With the development of breeding techniques, more varieties of rice seeds are brought into the market, and the purity of rice seeds is critical for planters and consumers.

Different varieties of rice seed vary in physical and chemical characteristics, growth performances and stress tolerance. Rice seed varieties can be identified by inspecting the appearance characteristics such as size, color, shape and texture, or by determining the quality attributes such as protein, starch and aroma. Traditional methods for rice variety identification, like High Performance Liquid Chromatography (HPLC) and Gas Chromatography-Mass Spectrometer (GC-MS), are either time consuming or expensive, and generally are only applied on a small number of sampled seeds [1,2]. Thus, developing a rapid method to identify rice seed varieties is needed.

Hyperspectral imaging is a technique integrating visible/near-infrared spectroscopy and imaging to acquire spectral and spatial information of the samples simultaneously. A hyperspectral image

can be defined as a data cube of shape $w \times h \times b$, where w (width) and h (height) denote the spatial resolution and b denotes the number of spectral bands. From the perspective of imaging, there is an image of size $w \times h$ for each spectral band. From the perspective of spectroscopy, there is a spectrum of length b for each spatial pixel. Compared with non-imaging spectroscopy, it is possible to obtain spectra of multiple samples in one single scan using a hyperspectral imaging system. There are hyperspectral imaging systems for different spectral ranges, and identification of seed varieties has been conducted at different spectral ranges [3–6].

Hyperspectral imaging generates a large amount of data representing spectral and spatial information. Multivariate hyperspectral image analysis is required to deal with such amount of data based on different objectives, including spectral and image preprocessing, feature selection and extraction, regression and classification analysis [7–10]. Classification methods have been successfully used to identify seed varieties using hyperspectral imaging [11] and various modeling methods were used. Zhang et al. (2012) used a line scan hyperspectral imaging system to identify maize varieties using batches of maize kernels. Principal component analysis (PCA) and kernel PCA (KPCA) were applied to extract useful information from spectra, then least squares support vector machine (LS-SVM) and back-propagation neural network (BPNN) were used for classification. The classification accuracy was all over 90% [3]. Kong et al. (2013) used a near-infrared hyperspectral imaging system to classify rice seed varieties. The key point was to compare the performances of different discriminant models, including partial least square discriminant analysis (PLS-DA), soft independent modeling of class analogy (SIMCA), K-nearest neighbor (KNN), support vector machine (SVM) and random forest (RF). The classification accuracy of most discriminant models was over 90% [6]. Mahesh et al. (2013) used a near-infrared hyperspectral imaging system at the range of 960–1700 nm to classify 4 wheat classes. Particularly, seeds were collected with variations of crop year, growing location and water content. Seed classifications were conducted based on water content of 13%, 16% and 19%. The pairwise moisture specific models (linear discriminant analysis, quadratic discriminant analysis and BPNN) were built with good classification results. Moreover, the key wavelengths of classification of wheat classes were identified by a stepwise discriminant analysis [12]. Yang et al. (2015) used a hyperspectral imaging system at the spectral range of 400–1000 nm to identify 4 varieties of waxy corn with consideration of the placement of germ. Spectral information, texture features were extracted from each corn. PLS-DA and SVM models were built using full spectra and image features of selected important wavelengths. The results indicated that hyperspectral imaging can be applied to differentiate maize seed varieties and test seed purity effectively [13].

Deep learning, a very hot topic in artificial intelligence (AI), has gained attentions from different fields. Convolutional neural networks (CNNs) are one of the most popular deep learning models. A typical CNN includes convolutional layers for feature extraction, pooling layers for feature compression and fully-connected layers for classification. Multiple convolutional layers and pooling layers can be chained to learn features of increasing levels of abstraction. CNNs have achieved state-of-the-art performance in various tasks, especially in computer vision. In the field of hyperspectral image analysis, CNNs have been employed for classification of two-dimensional image and three-dimensional hyperspectral remote sensing data.

Since CNNs are generally used for two-dimensional and three-dimensional data classification, trails using CNNs to handle one-dimensional spectral data are challenging. Researchers have tried to develop CNNs for spectral data analysis. Acquarelli et al. (2017) proposed a CNN architecture for vibrational spectroscopic data analysis [14]. Liu et al. (2017) proposed a CNN architecture for Raman spectra recognition [15]. One drawback of CNN is the number of training samples needed. For non-image spectroscopy, collecting large number of spectra is time-consuming. A hyperspectral image of multiple objects can be obtained in one scan. After image segmentation, spectra of each object can be extracted. This enables fast collection of samples for CNN training. However, the sample collection procedure is still not free. A study on the influence of training samples on CNNs would help in the trade-off between model performance and sample collection cost.

The objective of this study was to identify rice seed varieties using hyperspectral imaging combined with convolutional neural network. The specific objectives were to: (1) develop deep learning models using full spectra to identify seed variety at two spectral ranges; (2) to compare the results of deep learning methods with SVM and KNN; (3) to evaluate the influence of the number of training samples on KNN, SVM and CNN.

2. Materials and Methods

2.1. Sample Preparation

Four varieties of rice seeds (including Japonica rice: Xiushui 134, Zhejing 99; Indica rice: Zhongjiazao 17 and Zhongzao 39) harvested in 2016 were collected from a local seed company in Hangzhou, Zhejiang Province, China. The rice seeds were dried by natural air-drying. The seeds were cleaned and packed in paper bags and brought into the laboratory. In all, 5664, 5394, 4933, and 4916 rice seeds of Xiushui 134, Zhejing 99, Zhongjiazao 17 and Zhongzao 39 were obtained, respectively. Different category values were assigned to the four varieties of rice seeds, Xiushui 134, Zhejing 99, Zhongjiazao 17 and Zhongzao 39 were assigned as 1, 2, 3 and 4 respectively. Figure 1 shows samples of the four varieties of rice seeds.
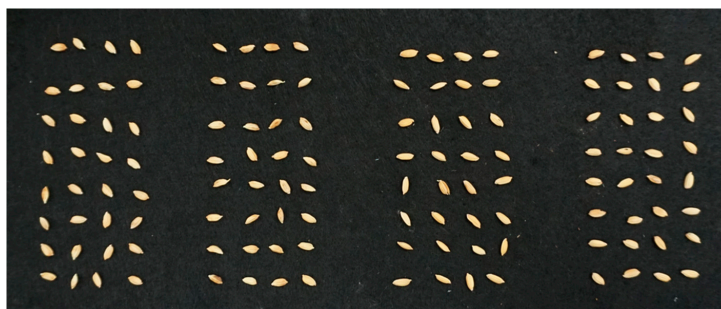

Figure 1. Rice seeds of 4 varieties. From left to right: Xiushui 134, Zhejing 99, Zhongjiazao 17 and Zhongzao 39.

2.2. Hyperspectral Imaging Acquisition

Hyperspectral image acquisition was conducted on a visible near-infrared hyperspectral imaging system covering the spectral range of 380–1030 nm and a near-infrared hyperspectral imaging system covering the spectral range of 874–1734 nm.

As for visible near-infrared hyperspectral imaging system, an imaging spectrograph (ImSpector V10E; Spectral Imaging Ltd., Oulu, Finland) with the spectral resolution of 2.8 nm, a high-performance CCD camera (Hamamatsu, Hamamatsu City, Japan) with 672 × 512 (spatial × spectral) pixels, a camera lens (OLES23; Specim, Spectral Imaging Ltd., Oulu, Finland) were assembled.

As for near-infrared hyperspectral imaging system, an imaging spectrograph (ImSpector N17E; Spectral Imaging Ltd., Oulu, Finland), a high-performance camera (Xeva 992; Xenics Infrared Solutions, Leuven, Belgium) with 326 × 256 (spatial × spectral) pixels, a camera lens (OLES22; Specim, Spectral Imaging Ltd., Oulu, Finland) were assembled.

The two systems were integrated in one platform. In the platform, two 150 W tungsten halogen lamps (3900 Lightsource, Illumination Technologies Inc., Elbridge, NY, USA) were used for illumination of the two systems; a conveyer belt driven by a stepper motor (Isuzu Optics Corp., Taiwan, China) was used for sample motion for line scan. The two systems were controlled by two data acquisition and preprocessing software (Spectral Image-V10E and Spectral-Image-Xenics 17E, Isuzu Optics Corp., Taiwan, China).

To acquire clear and nondeformable hyperspectral images, the moving speed of the conveyer belt, the exposure time of the camera and the height between the lens of the camera and the samples for the visible/near-infrared hyperspectral imaging system were adjusted to 2.3 mm/s, 0.04 s and 22.2 cm,

respectively. The moving speed of the conveyer belt, the exposure time of the camera and the height between the lens of the camera and the samples for the near-infrared hyperspectral imaging system were adjusted to 18.8 mm/s, 3 ms and 20 cm, respectively.

2.3. Hyperspectral Image Correction

The acquired hyperspectral images should be corrected to reduce the influence of dark current [16]. Raw hyperspectral images can be corrected by the following equation:

$$I_c = \frac{I_{raw} - I_{dark}}{I_{white} - I_{dark}} \tag{1}$$

where I_c is the corrected image, I_{raw} is the raw image, I_{dark} is the dark reference image acquired by covering the camera lens completely with its opaque cap. I_{white} is the white reference image acquired by using a white Teflon tile with nearly 100% reflectance.

2.4. Spectra Extraction and Preprocessing

Spectral information was extracted from the hyperspectral images after image correction. Due to limitation of the hyperspectral imaging systems, the reflectance values of the first several bands and the last several bands are noisy and cannot be trusted. Thus, only trusted range of spectrum was used for further analysis. The spectral range of 441–949 nm was selected for visible/near-infrared hyperspectral images and the spectral range of 975–1646 nm was selected for near-infrared hyperspectral images. Wavelet transform (WT) is an efficient method for spectra preprocessing [17]. In this study, WT using Daubechies 8 as basis function and a decomposition level 3 was applied as a preprocessing step to obtain smoothed spectra. After preprocessing of the spectral dimension, image segmentation was applied on the spatial dimensions. Generally, reflectance values of the black background were close to zero, so pixels of rice seeds can be easily segmented by thresholding. Then, each group of connected foreground pixels is regarded as a rice seed. At last, for each rice seed, all pixel-wise spectra were averaged to one spectrum for further analysis.

2.5. Multivariate Data Analysis

2.5.1. Support Vector Machine

Support vector machine (SVM) is a supervised pattern recognition method [18]. Due to the effectiveness of handling linear and nonlinear data efficiently, SVM has been widely used in spectral data analysis. SVM maps the original data into higher dimensional spaces and constructs a hyperplane or a set of hyperplanes to maximize the distance between the nearest samples of different classes. The selection of an appropriate kernel function is essential in SVM and affects the performance of SVM. Radial basis function (RBF) is a widely used kernel function with the ability to deal with nonlinear data efficiently. To conduct SVM using RBF kernel, penalty coefficient C and the kernel parameter g should be determined to obtain optimal performances, and a grid-search procedure is generally used to determine C and g. In this study, a grid-search procedure was applied to optimize C and g, and the searching range of C and g were from 2^{-8} to 2^8.

2.5.2. K-Nearest Neighbor

K-nearest neighbor (KNN) is a widely used pattern recognition method [19]. It calculates the distances between an unknown sample and the samples in the predefined training set. The number of nearest samples (k) to the unknown sample is manually defined. The category of the unknown sample is determined by the categories of its k nearest samples in the training set. The determination of k is crucial for KNN. In this study, k was optimized by comparing KNN models using k from 3 to 20 with a step of 1.

2.5.3. Convolutional Neural Network

Convolutional Neural Network Architecture

The architecture of our convolutional neural network (CNN) is shown in Figure 2. We adapted the design of VGGNet [20] for one-dimensional spectra inputs. Patterns in spectral curves share some common characteristics with image patterns. For example, peaks and minimums in spectral curve is similar to edges in images. VGGNet is chosen for its high performance in image classification tasks and its modular design make it easy to be modified and extended.
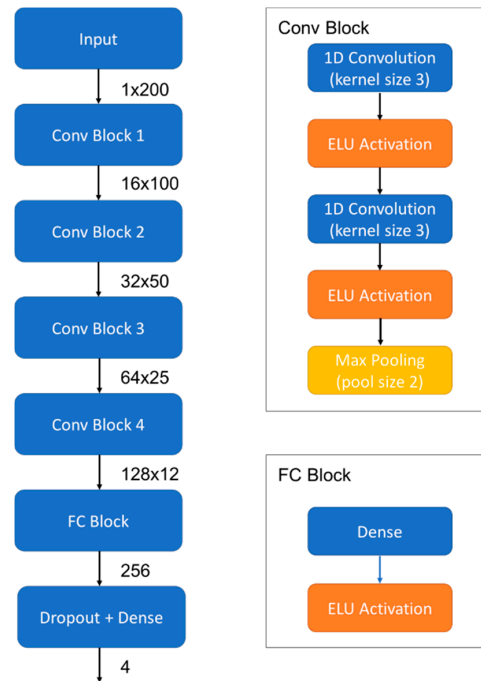

Figure 2. The architecture of CNN.

In Figure 2, an example input of 200 spectral bands was used to show the output size of each block. There are five main blocks in the architecture. The first four blocks are convolutional blocks (Conv Blocks), each of which consists of two consecutive convolutional layers followed by a max pooling layer. As blocks going deeper, the number of convolutional filters is doubled (starting from 16 and ending with 128). All convolutional layers use a kernel size of 3, stride of 1 and padding of 1. A convolutional layer has local connections to its input and can be trained to learn local patterns. By chaining convolutional layers together, deeper layers could have connection to larger part of the raw input. Thus, different layers "see" the raw input and learn features at different levels. The last block is a block of a fully connected layer (FC Block). The fully connected layer is used to learn combinations of features extracted by the convolutional layers. At last, a dropout layer and a dense layer were added as the output layer [21].

The original VGGNet architecture used the rectified linear unit (ReLU) as activation function. The exponential linear unit (ELU) was reported to speed up learning and outperform ReLU [22]. In our experiment, ELU activation showed better performance than ReLU activation with batch normalization [23]. Thus, ELU is utilized in the architecture. ELU function is defined as

$$f(x) = \begin{cases} x & if \ x > 0 \\ \alpha(\exp(x) - 1) & if \ x \leq 0 \end{cases} \tag{2}$$

The output of the CNN is followed by a softmax function to produce values in the range [0, 1] as classification confidence scores. A classification loss is then calculated by comparing the confidence scores and true labels of samples. The softmax function and the loss function are defined as follows,

$$p_{ij} = \frac{e^{z_{ij}}}{\Sigma_{k=1}^{K} e^{z_{ik}}} \ for \ j = 1, \ldots, K \tag{3}$$

$$Loss = -\sum_{i} \sum_{j} label_{ij} \log(p_{ij}) \tag{4}$$

where z denotes the output of CNN, i denotes a sample, j denotes a class and K denotes the total number of classes.

CNN Training

Before training, a normalization step was performed by subtracting the mean and dividing by the standard deviation. The mean and the standard deviation of training data were saved for preprocessing of test data. The weights of the CNN were initialized using the strategy introduced by [24]. The training was carried out by optimizing the softmax cross entropy loss using the Adam algorithm [25]. During training, the learning rate (η) was gradually reduced according to the following equation:

$$\eta = \frac{\eta_0}{1 + kt} \tag{5}$$

where η_0 denotes the initial learning rate, t denotes the number of epochs, and k controls the speed of learning rate reduction.

A grid-search was applied to find the best combination of hyperparameters. Finally, the batch size was set to 256, the α in ELU was set to 1.0 and the dropout ratio was set to 0.5. The CNN was trained for 200 epochs with $\eta_0 = 0.0005$ and $k = 0.045$.

3. Results

3.1. Spectral Profiles

After spectra extraction and spectral preprocessing, rice seeds were randomly divided into a training set with 3000 samples of each variety and a hold-out test set with 2664, 2394, 1933, and 1916 samples of Xiushui 134, Zhejing 99, Zhongjiazao 17 and Zhongzao 39, respectively. To train models with different sizes of training set, subsets of 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500 and 3000 samples of each variety were selected from the full training set. The test performance of all models was evaluated on the same hold-out test set.

The spectral range of 441–948 nm (Spectral range 1) for visible/near-infrared system and 975–1646 nm (Spectral range 2) for near-infrared system were analyzed. The average spectra of 4 varieties of rice seeds in two different spectral ranges are shown in Figure 3a,b, and corresponding average spectra preprocessed by second derivative (polynomial order: 2, smoothing points: 7) are presented in Figure 3c,d.

In the two spectral ranges, only slight differences existed among the average spectra of different varieties. As seen from the second derivative spectra in the Spectral range 1, more obvious differences could be observed at the wavelengths of 451, 456, 463, 478, 678, 698 and 716 nm. These wavelengths were mainly related to the color variations of the rice seeds. In the Spectral range 2, the wavelengths at 1136, 1177, 1251, 1325, 1386, 1440, 1494 and 1619 nm showed greater differences. The wavelengths at 1136, 1177 and 1251 nm are attributed to the second overtone of C–H stretching mode [26]. The wavelength at 1386 nm (around 1387 nm) might be attributed to the first overtone of stretching and anti-symmetric O–H bond [27]. The wavelengths at 1440 and 1494 nm are attributed to water [26,28]. The wavelength at 1619 nm (around 1620 nm) might be attributed to first overtone of C–H group

absorption [29]. However, for both preprocessed and unpreprocessed average spectra, the spectral differences among different seed varieties were not significant.
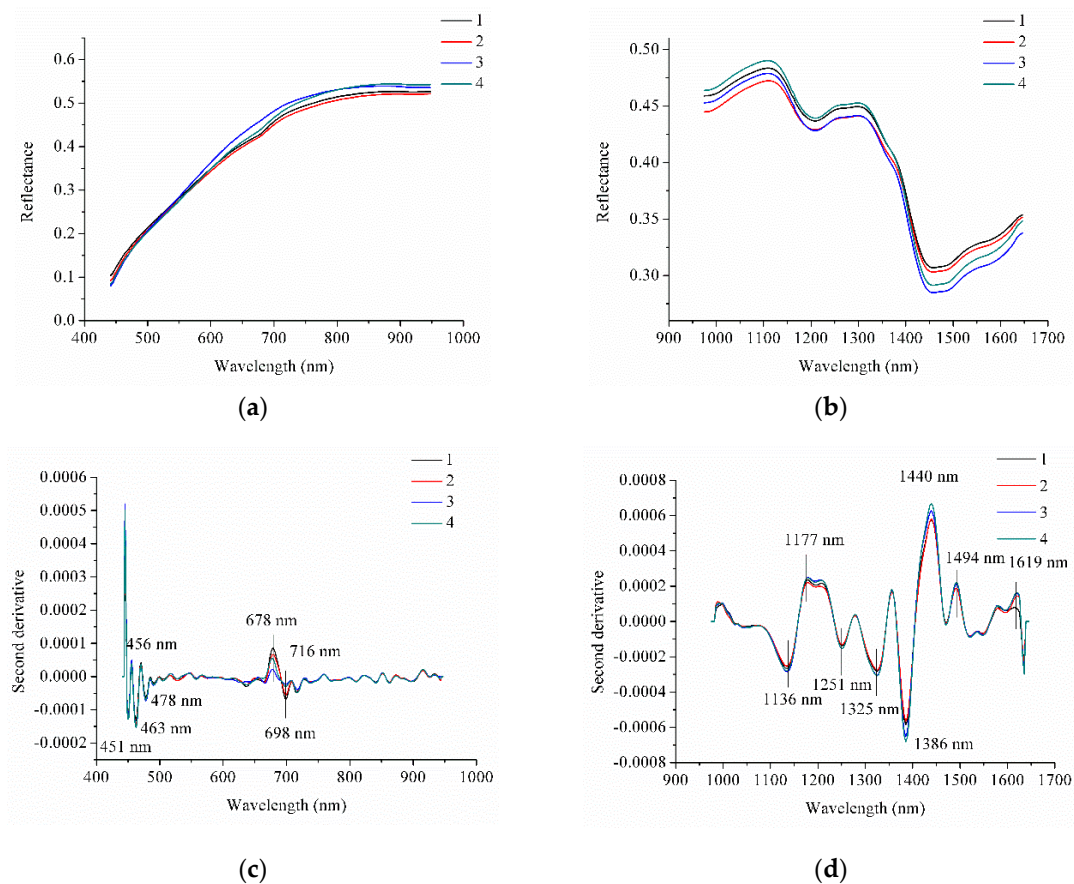

Figure 3. Average spectra of 4 varieties of rice seeds in two different spectral ranges, (**a**) Spectral range 1: 441–948 nm, (**b**) Spectral range 2: 975–1646 nm; the corresponding average spectra preprocessed by second derivative, (**c**) Spectral range 1, (**d**) Spectral range 2.

3.2. Classification Results of Different Models

Discriminant models were built by KNN, SVM and CNN models. To evaluate the effectiveness of these models, training sets with different number of samples were used. Training sets with 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500 and 3000 samples selected from the total training samples of each rice seed variety were used to build discriminant models, and the test set for above-mentioned models were all the same (2664, 2394, 1933, and 1916 rice seeds of Xiushui 134, Zhejing 99, Zhongjiazao 17 and Zhongzao 39) for different number of training samples. The overall results are presented in Figure 4.

As shown in Figure 4, the classification accuracy of the training and test set of different models increased with the increase in the number of samples in the training set. Some fluctuations could be observed from the classification accuracy of the training set. When the number of samples came to a critical value, the increase rate of classification accuracy was small. For different models, the critical values were different. The models using 3000 training samples performed best since large training set is expected to contain a more extensive set of different types of feature combinations and feature values than small training sets.
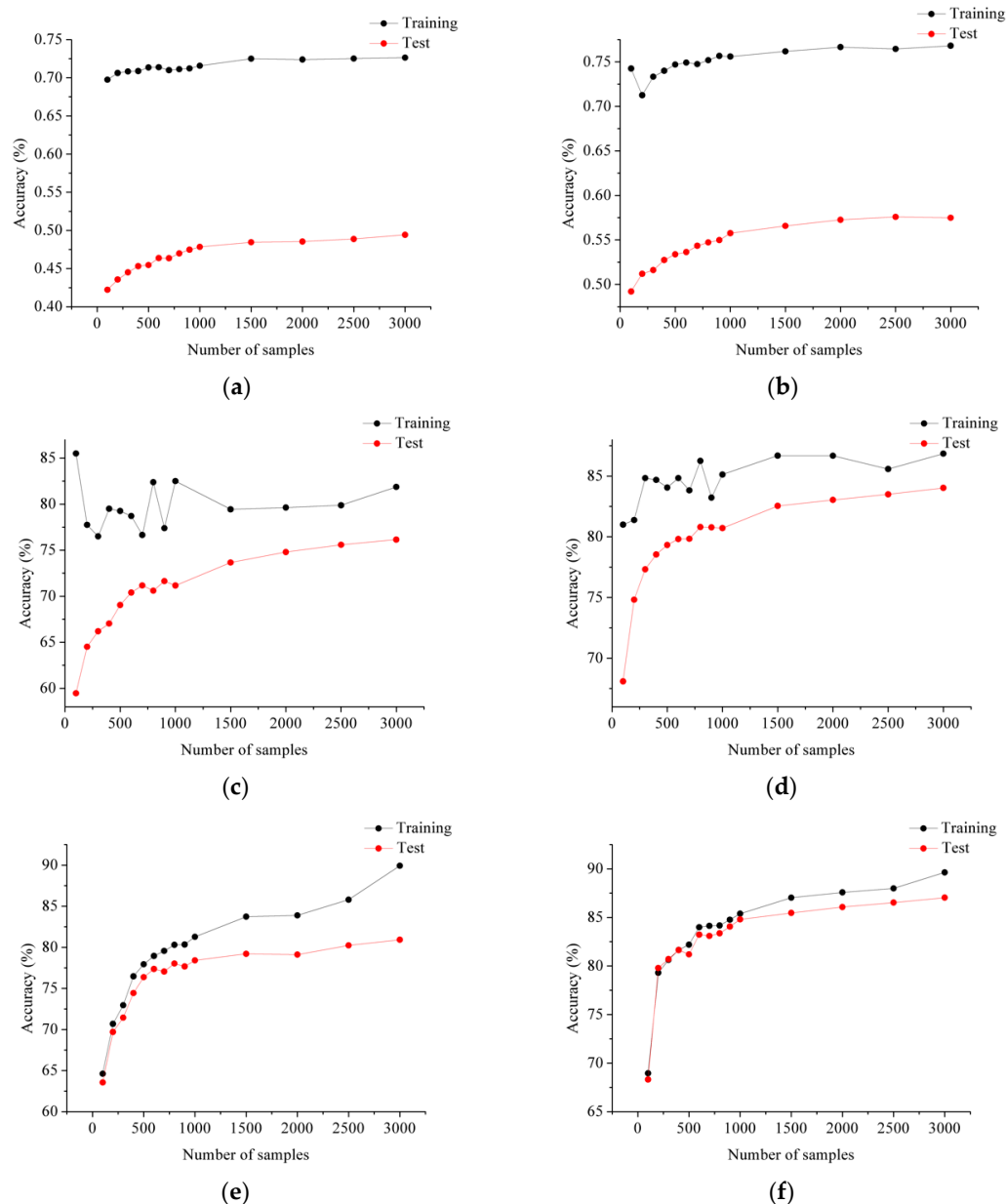
Figure 4. Results of discriminant models using different number of samples (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500 and 3000) in the training set. (**a**) Results of KNN models using different number of samples for Spectral range 1; (**b**) Results of KNN models using different number of samples for Spectral range 2; (**c**) Results of SVM models using different number of samples for Spectral range 1; (**d**) Results of SVM models using different number of samples for Spectral range 2; (**e**) Results of CNN models using different number of samples for Spectral range 1; (**f**) Results of CNN models using different number of samples for Spectral range 2.

It can be seen in Figure 5 that the KNN model performs much worse than the other two models, and the test accuracy of models using Spectral range 1 and 2 are both lower than 60%. For KNN, SVM and CNN models, the classification accuracy of the training and test sets increased along with the increase of the training samples, although some fluctuations existed in the classification accuracy of the training sets (especially KNN and SVM models). KNN models using different number of training samples obtained relatively worse results than the corresponding SVM and CNN models. For SVM and CNN models, the classification results were close when the same number of samples was used. With

the increase of training samples, the performances of CNN models were better than the corresponding SVM models. The CNN model using 3000 samples in the Spectral range 2 obtained the best results. The classification accuracy of the training and test set was 89.6% and 87.0%, respectively.
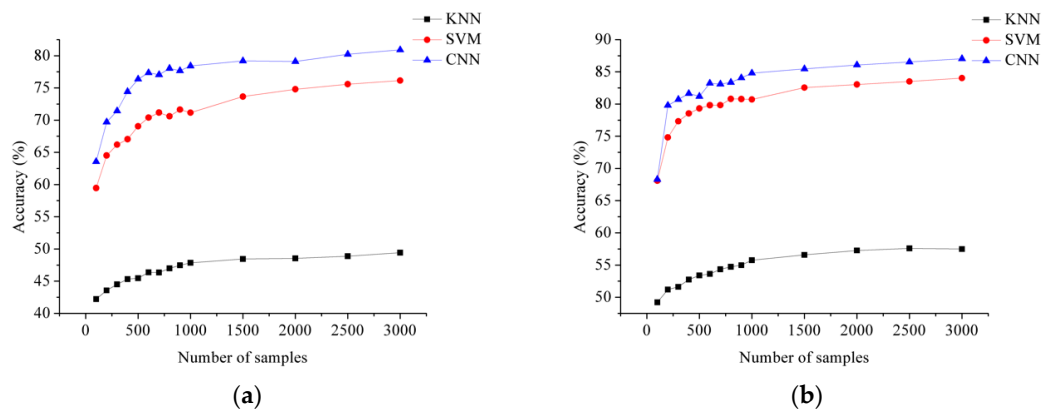

Figure 5. Test performance of KNN, SVM and CNN on the fixed hold-out test set. (**a**) Results of models using Spectral range 1; (**b**) Results of models using Spectral range 2.

For all methods, models trained with reflectance values in Spectral range 2 performed better than models trained with reflectance values in Spectral range 1. As seen from Figure 3, no typical spectral peaks or minimums could be found from the average spectra in the Spectra range 1, while obvious spectral peaks or minimums were found from the average spectra in the Spectra range 2. The differences in second derivative spectra in the Spectral range 1 were mainly attributed to the color variations of the rice husk, and the differences in second derivative spectra in the Spectral range 2 were mainly attributed to the chemical compositions of rice seeds. This might be the reason why models in Spectral range 1 performed a little worse than those in the Spectral range 2.

Tables 1 and 2 show the confusion matrices of SVM and CNN models using 3000 training samples of each rice seed variety. The performances of KNN models were much worse than SVM models and CNN models, so the confusion matrices are not presented. As shown in Tables 1 and 2, rice seeds of variety 1 (Xiushui 134) and 2 (Zhejing 99) were more likely to be misclassified as each other, while rice seeds of variety 3 (Zhongjiazao 17) and 4 (Zhongzao 39) showed the similar phenomenon. This might be caused by the difference in rice types. Rice seeds of variety 1 and 2 are japonica rice, while variety 3 and 4 are indica rice. Compared with SVM, CNN produces less misclassified samples in each variety. In both spectral ranges, CNN outperformed SVM in overall accuracy and per-class accuracy. The results indicated that CNN was effective for spectral data analysis and could improve the general classification accuracy.

Table 1. Confusion matrices of SVM and CNN models using 3000 training samples of each rice seed variety in the Spectral range 1.

		Training Set					Test Set				
		1	2	3	4	Accuracy (%)	1	2	3	4	Accuracy (%)
	1	2281	606	74	39	76.0	1849	693	80	42	69.4
	2	461	2350	145	44	78.3	467	1735	120	72	72.5
SVM	3	42	139	2578	241	85.9	36	127	1582	188	81.8
	4	20	75	290	2615	87.2	46	62	192	1616	84.3
	total					81.9					76.1
	1	2530	462	8	0	84.3	1979	643	33	9	74.3
	2	347	2631	16	6	87.7	455	1841	68	30	76.9
CNN	3	11	21	2804	164	93.5	12	80	1673	168	86.5
	4	0	0	166	2826	94.2	12	21	169	1714	89.5
	total					89.9					80.9

Table 2. Confusion matrices of SVM and CNN models using 3000 training samples of each rice seed variety in the Spectral range 2.

		Training Set					Test Set				
		1	2	3	4	Accuracy (%)	1	2	3	4	Accuracy (%)
SVM	1	2723	275	1	1	90.8	2367	289	5	3	88.9
	2	129	2840	20	11	94.7	163	2179	34	18	91.0
	3	1	31	2450	518	81.7	5	44	1498	386	77.5
	4	1	20	570	2409	80.3	3	14	459	1440	75.2
	total					86.9					84.0
CNN	1	2836	164	0	0	94.5	2453	209	1	1	92.1
	2	141	2846	10	3	94.9	177	2179	29	9	91.0
	3	0	8	2560	432	85.3	4	28	1584	317	81.9
	4	0	4	482	2514	83.8	0	8	372	1536	80.2
	total					89.6					87.0

4. Discussion

Machine learning methods play essential roles in interpreting spectral data for different spectroscopy techniques and spectral imaging techniques. Deep learning is now a hot topic in AI, and convolutional neural networks (CNNs) are one of the most popular deep learning models. Generally, deep learning methods were used to analyze two-dimensional images. In this study, CNN was applied to one-dimensional spectra, and the CNN models obtained good performances. When compared with KNN and SVM models, CNN model performed better. The results of this study indicate that CNN could be used to handle spectral data effectively, which brings new alternatives for spectral data analysis. Acquarelli et al. (2017) proposed a CNN architecture for vibrational spectroscopic data analysis, and partial least squares-discriminant analysis (PLS-DA), logistic regression and KNN were used for comparison. The results showed that CNN model yielded satisfactory results. In most cases, CNN performed best, and in some other cases, CNN was not the best model [14]. Liu et al. (2017) proposed a CNN architecture to classify preprocessed and unpreprocessed Raman spectra. CNN models obtained the best performances when compared with KNN, SVM, gradient boosting, random forest and correlation [15]. These researches together with the results in this manuscript indicated that CNN could be used to deal with one-dimensional spectral data.

As can be seen from the confusion matrices, models performed better in distinguishing two rice varieties of different types (e.g., rice 1 and rice 3) than in distinguishing two rice varieties of the same type (e.g., rice 1 and rice 2). Moreover, models of Spectral range 1 performed better in classification of rice varieties 3 and 4 (both indica rice) than Spectral range 2, while models of Spectral range 2 performed better in classification of rice varieties 1 and 2 (both japonica rice) than Spectral range 1. This indicates a potential model improvement by combining two spectral ranges. However, hyperspectral images of the two spectral ranges were scanned by two different hyperspectral cameras. To combine these two spectral ranges, works should be done to match rice seeds in two corresponding hyperspectral images. This might be time-consuming in practical use. In this study, only 4 varieties of rice seeds were used. Since CNNs have achieved high performance in classification of thousands of image classes, classification of more rice seed varieties using a CNN might also be possible. But as a cost, large amounts of samples should be collected for each rice seed variety. More varieties of rice seeds and even more crop species should be studied in future research.

The influence of the number of training samples was also studied. Generally, performance of machine learning methods increases with the number of training samples. The lack of training samples limits the test performance of trained models. While test performance increase with the number of training samples, the improvement might not be significant after a certain point due to redundancy of information contained in training samples. Moreover, the sample collection procedure could be time-consuming. Thus, it is important to balance performance and cost of models. As shown in Figure 5, overall, the model performance improved with the number of training samples for KNN,

SVM and CNN. When the number of training samples increased, the performances of the CNN models were superior to the corresponding SVM models. Deep learning methods could learn features automatically, and more samples would allow more potential feature combination to be explored. For models trained with less than 1500 samples, performance improved significantly with the number of training samples. For models trained with more than 1500 samples, performance improvements were not significant. In practical applications, models should be built to identify much more rice varieties. To achieve high model performance with a reasonable sample collection cost, it is better to keep a hold-out test set, and then gradually collect samples for training until test accuracy does not change significantly.

5. Conclusions

In this study, hyperspectral imaging of two different spectral range was used to identify rice seed varieties. Performance of three machine learning methods, namely KNN, SVM and CNN were tested. Models built with reflectance values in Spectral range 2 performed better than models built with reflectance values in Spectral range 1. The influence of the number of training samples was also studied. As the size of training set increase, CNN models outperformed the other two models. The result showed that with the help of fast sample collection using hyperspectral imaging, CNN could be an effective method for spectral data classification. In future research, more rice varieties should be studied to extend the use of CNNs in spectra data analysis.

Author Contributions: Z.Q., J.C., C.Z. and Y.H. conceived and designed the experiments; Y.Z., S.Z., J.C. and C.Z. performed the experiments; Y.Z., J.C. and C.Z. analyzed the data; Y.H. and Z.Q. contributed reagents/materials/analysis tools; Z.Q., J.C. and C.Z. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Huebner, F.R.; Hussain, A.; Lookhart, G.L.; Bietz, J.A.; Bushuk, W.; Juliano, B.O. Discrimination of sister-line IR rice varieties by polyacrylamide gel electrophoresis and reversed-phase high-performance liquid chromatography. *Cereal Chem.* **1991**, *68*, 583–588.
2. Wu, J.; Yu, H.; Dai, H.; Mei, W.; Huang, X.; Zhu, S.; Peng, M. Metabolite profiles of rice cultivars containing bacterial blight-resistant genes are distinctive from susceptible rice. *Acta Biochim. Biophys. Sin.* **2012**, *44*, 650–659. [CrossRef] [PubMed]
3. Zhang, X.; Liu, F.; He, Y.; Li, X. Application of hyperspectral imaging and chemometric calibrations for variety discrimination of maize seeds. *Sensors* **2012**, *12*, 17234–17236. [CrossRef] [PubMed]
4. Yang, S.; Zhu, Q.B.; Huang, M.; Qin, J.W. Hyperspectral image-based variety discrimination of maize seeds by using a multi-model strategy coupled with unsupervised joint skewness-based wavelength selection algorithm. *Food Anal. Methods* **2017**, *10*, 424–433. [CrossRef]
5. Sun, J.; Lu, X.; Mao, H.; Jin, X.; Wu, X. A method for rapid identification of rice origin by hyperspectral imaging technology. *J. Food Process Eng.* **2017**, *40*, e12297. [CrossRef]
6. Kong, W.; Zhang, C.; Liu, F.; Nie, P.; He, Y. Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis. *Sensors* **2013**, *13*, 8916–88927. [CrossRef] [PubMed]
7. Manley, M. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chem. Soc. Rev.* **2014**, *43*, 8200–8214. [CrossRef] [PubMed]
8. Dai, Q.; Sun, D.-W.; Xiong, Z.; Cheng, J.-H.; Zeng, X.-A. Recent advances in data mining techniques and their applications in hyperspectral image processing for the food industry. *Compr. Rev. Food Sci. Food Saf.* **2015**, *13*, 891–905. [CrossRef]
9. Dai, Q.; Cheng, J.-H.; Sun, D.-W.; Zeng, X.-A. Advances in feature selection methods for hyperspectral image processing in food industry applications: A review. *Crit. Rev. Food Sci. Nutr.* **2015**, *55*, 1368–1382. [CrossRef] [PubMed]

10. Wu, D.; He, Y.; Nie, P.; Cao, F.; Bao, Y. Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. *Anal. Chim. Acta* **2010**, *659*, 229–237. [CrossRef] [PubMed]

11. Anisur, R.; Byoung-Kwan, C. Assessment of seed quality using non-destructive measurement techniques: A review. *Seed Sci. Res.* **2016**, *26*, 285–305.

12. Mahesh, S.; Jayas, D.S.; Paliwal, J.; White, N.D. Comparing two statistical discriminant models with a back-propagation neural network model for pairwise classification of location and crop year specific wheat classes at three selected moisture contents using nir hyperspectral images. *Trans. ASABE* **2014**, *57*, 63–74.

13. Yang, X.; Hong, H.; You, Z.; Cheng, F. Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. *Sensors* **2015**, *15*, 15578–15594. [CrossRef] [PubMed]

14. Acquarelli, J.; Van, L.T.; Gerretzen, J.; Tran, T.N.; Buydens, L.M.; Marchiori, E. Convolutional neural networks for vibrational spectroscopic data analysis. *Anal. Chim. Acta* **2017**, *954*, 22–31. [CrossRef] [PubMed]

15. Liu, J.; Osadchy, M.; Ashton, L.; Foster, M.; Solomon, C.J.; Gibson, S.J. Deep convolutional neural networks for Raman spectrum recognition: A unified solution. *Analyst* **2017**, *142*, 4067–4074. [CrossRef] [PubMed]

16. Elmasry, G.; Kamruzzaman, M.; Sun, D.W.; Allen, P. Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: A review. *Crit. Rev. Food Sci. Nutr.* **2012**, *52*, 999–1023. [CrossRef] [PubMed]

17. Zhang, C.; Ye, H.; Liu, F.; He, Y.; Kong, W.; Sheng, K. Determination and visualization of ph values in anaerobic digestion of water hyacinth and rice straw mixtures using hyperspectral imaging with wavelet transform denoising and variable selection. *Sensors* **2016**, *16*, 244. [CrossRef] [PubMed]

18. Burges, C.J.C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [CrossRef]

19. Balabin, R.M.; Safieva, R.Z.; Lomakina, E.I. Near-infrared (NIR) spectroscopy for motor oil classification: From discriminant analysis to support vector machines. *Microchem. J.* **2011**, *98*, 121–128. [CrossRef]

20. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.

21. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

22. Clevert, D.; Unterthiner, T.; Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *arXiv* **2015**, arXiv:1511.07289.

23. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.

24. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision, Santiago, Chile, 7–13 December 2015; pp. 1026–1034.

25. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.

26. Westad, F.; Schmidt, A.; Kermit, M. Incorporating chemical band-assignment in near infrared spectroscopy regression models. *J. Near Infrared Spectrosc.* **2008**, *16*, 265–273. [CrossRef]

27. Marques, A.S.; Castro, J.N.F.; Costa, F.J.; Neto, R.M.; Lima, K.M.G. Near-infrared spectroscopy and variable selection techniques to discriminate pseudomonas aeruginosa strains in clinical samples. *Microchem. J.* **2016**, *124*, 306–310. [CrossRef]

28. Kinoshita, K.; Miyazaki, M.; Morita, H.; Vassileva, M.; Tang, C.; Li, D.; Ishikawa, O.; Kusunoki, H.; Tsenkova, R. Spectral pattern of urinary water as a biomarker of estrus in the giant panda. *Sci. Rep.* **2012**, *2*, 856. [CrossRef] [PubMed]

29. Aw, W.C.; Dowell, F.E.; Ballard, J.W.O. Using near-infrared spectroscopy to resolve the species, gender, age, and the presence of wolbachia infection in laboratory-reared drosophila. *G3* **2012**, *2*, 1057–1065. [CrossRef] [PubMed]