Title
End-to-end fusion of hyperspectral and chlorophyll fluorescence imaging to identify rice stresses

Short title: Deep fusion to identify rice stresses

Authors
Chu Zhang¹,†, Lei Zhou²,†, Qinlin Xiao³,4, Xiulin Bai³,4, Baohua Wu³,4, Na Wu³,4, Yiyong Zhao⁵, Junmin Wang⁶, Lei Feng³,4,*

Affiliations
1. School of Information Engineering, Huzhou University, Huzhou 313000, China.
2. College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, China.
3. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China.
4. Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou 310058, China.
5. Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China.
6. Institute of Crop Science and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China.
†These authors contributed equally to this work.
*Corresponding author. Email: Lfeng@zju.edu.cn

Abstract
Herbicides and heavy metals are hazardous substances of environmental pollution, resulting in plant stress and harming humans and animals. Identification of stress types can help trace stress sources, manage plant growth, and improve stress-resistant breeding. In this research, hyperspectral imaging (HSI) and chlorophyll fluorescence imaging (Chl-FI) were adopted to identify the rice plants under two types of herbicides stresses (butachlor (DCA) and quinclorac (ELK)) and two types of heavy metal stresses (cadmium (Cd) and copper (Cu)). Visible/near-infrared spectra of leaves (L-VIS/NIR) and stems (S-VIS/NIR) extracted from HSI and chlorophyll fluorescence kinetic curves of leaves (L-Chl-FKC) and stems (S-Chl-FKC) extracted from Chl-FI were fused to establish the models to detect the stress of the hazardous substances. Novel end-to-end deep fusion models were proposed for low-level, mid-level and high-level information fusion to improve identification accuracy. Results showed that the high-level fusion-based convolutional neural network (CNN) models reached the highest detection accuracy (97.7 %), outperforming the models using a single data source (<94.7 %). Furthermore, the proposed end-to-end deep fusion models required a much simpler training procedure than the conventional two-stage deep learning fusion. This research provided an efficient alternative for plant stress phenotyping, including identifying plant stresses caused by hazardous substances of environmental pollution.
1. Introduction

Rice (*Oryza sativa* L.) is one of the main staple foods in the world. With the development of breeding techniques, high yield and stress-resistant varieties have been developed and promoted for planting to increase the rice yield. However, the yield will reach the plateaus when the planting fields no longer increase. On the other hand, climate instability, biotic and abiotic stresses show great threats to rice production. Although stress-resistant varieties have been bred, they can only deal with a few stresses. These varieties also suffer from the complex growth environment, resulting in yield and quality loss.

Herbicides and heavy metals are the main hazardous substances causing environmental pollution, resulting in the abiotic stress of plants. Generally, the rice plants are exposed to herbicides or heavy metals, which might not cause death to the plants, and the plants under the threat of relatively low concentrations of herbicides and heavy metals might not show obvious differences in the symptoms. Under this situation, it is hard to identify the stress types of plants. Moreover, the stresses will affect plant growth, and different plant organs may respond differently. The acquisition of phenotyping traits of different organs can provide complementary information of plants, improving the precision and stability of plant growth status monitoring. Accurately and automatically identifying abiotic stress types can obtain plant growth information and find the optimal solution to deal with the stresses. Knowing what happens to the plants and how to treat them makes it possible to minimize the influence of abiotic stresses to stabilize the yield and quality.

High-throughput phenotyping of plants can help obtain a large number of phenotyping traits in digital and automatic manners. The development of modern analytical techniques has made great contributions to high-throughput plant phenotyping (Campbell et al., 2018; Shakoor et al., 2017; Singh et al., 2021; Song et al., 2021; Yang et al., 2020). The numerous data acquired by the analytical instruments contained various information relating to plant growth status. The plant phenotyping traits can evaluate the plant growth status. High-throughput phenotyping also helps study the relationship between phenomics and genomics (Mir et al., 2019; Xiao et al., 2022). Hyperspectral imaging (HSI) (Liu et al., 2020; Mishra et al., 2017; Saric et al., 2022; Zhang & Zhang, 2018) and chlorophyll fluorescence imaging (Chl-FI) (Gorbe & Calatayud, 2012; Luisa Perez-Bueno et al., 2019; Zhang & Zhang, 2018) are two widely used techniques for high-throughput plant phenotyping, providing different information on plant growth. HSI and Chl-FI have been studied for heavy metals and herbicides stresses (Chu et al., 2022; Li et al., 2018; Wang et al., 2020; Weber et al., 2017).

HSI integrates spectroscopy technique and imaging technique, capturing spectral information of pixel within the image. The spectral information is related to the chemical compositions and physiological and biochemical reactions. Image information relates to external information, such as plant structure, color, morphological features, etc. Full-range spectra, feature wavelengths and spectral indices derived from full-range spectra are the mostly used for analysis. Chl-FI has also been used for high-throughput phenotyping of plants. It captures chlorophyll fluorescence signals from the samples, which relates to photosynthesis. Chlorophyll fluorescence kinetics parameters are mostly used for analysis, and some studies have used chlorophyll fluorescence spectra for analysis (Kasampalis et al., 2021; Marques da Silva et al., 2020; Mishra et al., 2009).
Previous studies have proved the effectiveness of HSI and Chl-FI for plant stress phenotyping individually. The combination of HSI and Chl-FI has also been studied for plant stress phenotyping (Feng et al., 2018; Kumar et al., 2021; Li et al., 2010). However, most of these studies have analyzed HSI and Chl-FI separately (Feng et al., 2018; Kumar et al., 2021; Li et al., 2010; Naumann et al., 2008). Since HSI and Chl-FI acquired different phenotyping traits based on different principles, the fusion of the features of HSI and Chl-FI for plant stress types identification can be explored to use complementary information of different phenotyping traits. Information fusion has been widely used for integrating multi-modal or multi-sensor data to improve analysis performances for different purposes with higher precision and reliability. In the review of (Bauriegel & Herppich, 2014), the authors discussed the potential of information fusion of HSI and Chl-FI. In the review of (Singh et al., 2018), the author also discussed the potential of information fusion for plant stress phenotyping. Various studies have shown the good performances of information fusion for plant phenotyping, including the fusion of data acquainted by different types of sensors (representing different techniques) (Berdugo et al., 2014; Feng et al., 2020; Feng et al., 2022; Moshou et al., 2005; Moshou et al., 2014; Rischbeck et al., 2016; Salve et al., 2022; Xu et al., 2021; Zhou et al., 2021).

Deep learning has been the hottest topic in machine learning and artificial intelligence. With the ability to learn deep and representative features from big data, deep learning has been used in various fields, including plant phenotyping. Deep learning has also been used for high-throughput phenotyping plant phenotyping (Gao et al., 2020; Jiang & Li, 2020; Kamarudin et al., 2021; Noon et al., 2020; Saleem et al., 2019; Singh et al., 2018). According to previous studies, shallow CNN models can work well on one-dimensional (1D) spectra data (Chu et al., 2022; Zhang et al., 2019). To our knowledge, no previous studies have used the Chl-FKC as inputs of CNN. The 1D Chl-FKC is similar to the VIS/NIR spectra. Due to the significant feature learning ability, DL can fuse features from VIS/NIR spectra and Chl-FKC to fully reveal the phenotyping information of plants. Researchers have conducted deep learning-based information fusion for plant phenotyping (Kamarudin et al., 2021; Scholl et al., 2021; Weng et al., 2020).

In general, there are three different levels of information fusion, including low-level fusion (fusion of original data), mid-level fusion (fusion of features extracted by feature extractor) and high-level fusion (decision fusion) (Alofi, 2017; Azcarate et al., 2021; Khaleghi et al., 2013). However, most existing information fusion models are built based on a two-stage training procedure, requiring individual feature extractors and classifiers. The first step is to train a feature extractor or directly use a manually defined one to produce the features for fusion. The second step is to train another model for discrimination based on the fused features obtained in the first step (Berdugo et al., 2014; Feng et al., 2020; Salve et al., 2022). These information fusion models are complex and need manual intervention. Recent studies have developed end-to-end deep fusion models for applications, combining feature extraction and modeling in one model (Jomaa et al., 2020; Stahlschmidt et al., 2022). The features are automatically learned, fused and fed into the classifier. The end-to-end deep fusion models have simpler training procedures and are more applicable for real-world applications.

This study focused on identifying different types of abiotic stresses using HSI and Chl-FI techniques. The specific objectives were to: 1) identify two types of herbicides stresses (butachlor (DCA) and quinclorac (ELK)) and two types of heavy metal stresses (cadmium (Cd) and copper (Cu)); 2) explore the performances of the stress type identification using...
different techniques (HSI and Chl-FI) and different organs (leaves and stems); 3) explore the stress type identification using the three levels end-to-end deep fusion to fuse the information acquired by different techniques (HSI and Chl-FI) and different organs (leaves and stems).

2. Materials and Methods

2.1 Sample preparation

The rice variety used in this study was Zhongheyou 4, provided by the Institute of Crop Science and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences. The rice seeds were sowed onto the seedbed. The rice seedlings were transplanted to the laboratory for stress one month later. The rice seedlings were transplanted in plug trays with nutrient soil. Regular water and fertilizer management were conducted. After one week of transplantation, the seedlings were used for treatments with different stresses.

Two different herbicides were used, including butachlor (DCA) and quinclorac (ELK). Two heavy metals were used, including copper (Cu) and cadmium (Cd). For butachlor stress, butachlor solutions with 50% active constituent were purchased from the local pesticide shop, and the dosage of used butachlor solutions was 0 (control, CK), 1.5, 3 and 6 ml/L. For quinclorac stress, quinclorac powders with 99% active constituent were purchased, and the dosage of used quinclorac powder were 0 (CK), 0.56, 1.12 and 2.24 g/L. For Cu and Cd stresses, the concentration of Cu and Cd were 0 (CK), 10, 30 and 60 μm/L. The different concentrations of herbicides and heavy metals were used to include more sample variations under one type of stress. In all, the number of samples under the treatments of CK, Cd, Cu, DCA and ELK were 240, 360, 360, 268, 358, respectively.

After the first time of stress, regular water and fertilizer management were conducted. The efficiency of hyperspectral image acquisition and chlorophyll fluorescence image acquisition makes it unable to acquire a lot of samples a day. Thus, the four stresses were conducted on four successive days. After one week, the image acquisition of the corresponding stresses was conducted on another 4 successive days for the second batch of samples. As for hyperspectral imaging acquisition and chlorophyll fluorescence image acquisition, both leaves and stems were cut from the seedling and used for image acquisition separately.

2.2 Hyperspectral image acquisition

An assembled hyperspectral imaging system (as described previously in (Kong, Zhang, Cao, et al., 2018)) was used to acquire hyperspectral images of rice leaves and stems. The system covers the spectral range from 380 to 1030 nm, integrated with a spectrograph, a camera with lens and tungsten halogen light source. The samples could be placed on a conveyer belt driven by a stepper motor. The hyperspectral images were collected using line-scan mode and calibrated by black-white calibration.

\[
I = \frac{I_{\text{raw}} - I_{\text{black}}}{I_{\text{white}} - I_{\text{black}}} \tag{1}
\]
where $I$ denotes the calibrated spectral image. $I_{raw}$, $I_{black}$ and $I_{white}$ denote the raw image, black image and white image.

The system parameters were adjusted to acquire clear and non-deformable images. The camera exposure time was set as 0.027s, the speed of the moving plate was set as 3.4 mm/s, and the distance between the lens and the moving plate was adjusted to 26 cm. It was expected to cover the tested plant as much as possible, the plant was cut into segments and put symmetrically under the camera (on the conveyor belt) for line-scan image collection.

2.3 Chlorophyll fluorescence image acquisition

A pulse amplitude modulated chlorophyll fluorescence imaging system (FluorCam FC800, Photon Systems Instruments, Brno, Czechia) (as described previously in (Cen et al., 2017)) was used to acquire chlorophyll fluorescence images. This chlorophyll fluorescence imaging system consists a CCD camera (1392 × 1040 resolution) with an industry lens SV-H1.4/6 (VS Technology, Tokyo, Japan). The light source is formed by five light emitting diode (LED). An elevating table (HTVS120, SPL, Hangzhou, China) is used to adjust the distance between the samples and the lens. The image acquisition procedure was the same as (Yao et al., 2018), while the system parameters differed. When acquiring the chlorophyll fluorescence images, the actinic lights, saturating flashes, exposure time and sensitivity were adjusted 90 %, 75 %, 33.3 μs and 33.3 %. The segmented plants were placed inside a dark room (the area that the camera could cover) for chlorophyll fluorescence image acquisition.

2.4 Spectra data extraction and chlorophyll fluorescence kinetic curves extraction

2.4.1 Spectral curves extraction

For hyperspectral images, leaves of a sample are identified as the region of interest (ROI) of leaves (LROI), and the stem of a sample is also identified as an ROI (SROI). Considering the fact that the head and the end of the spectra contained obvious noises caused by the instrument, only the visible/near-infrared (VIS/NIR) spectra in the range of 454-957 nm (396 wavebands) were used for analysis. The pixel-wise spectra within the ROI were preprocessed by wavelet transform (wavelet function Daubechies 8 with decomposition level 3). All pixel-wise spectra within each ROI were averaged as the spectrum of the samples. The VIS/NIR spectra of leaves (L-VIS/NIR) and stems (S-VIS/NIR) were obtained for each rice plant, respectively. For chlorophyll fluorescence images, the leaves and the stem were identified as ROIs (similar to ROI definition in hyperspectral images), respectively. Pixel-wise Chl-FKC within each ROI were averaged as the chlorophyll fluorescence kinetic spectrum. The Chl-FKC of leaves (L-Chl-FKC) and stems (S-Chl-FKC) were obtained for each rice plant, respectively. Maximum normalization was conducted to preprocess the VIS/NIR spectra and the Chl-FKC for further modeling procedure.

2.4.2 Dataset preparation

To establish models, the class labels of the samples in the control group (CK), the class labels of the samples under the stresses of Cd, Cu, DCA and ELK were assigned as 0, 1, 2, 3 and 4, respectively. A typical dataset split approach in deep learning area was adopted.
For each group, the samples were randomly split into the training, validation and testing sets (Kuutti et al., 2021; Lim et al., 2020; Sun et al., 2022), using the ratio of 4:1:1. The samples in the training, validation and testing sets were the same for each single data source. The details of the dataset were listed in Table 1.

Table 1. Descriptions of the dataset

<table>
<thead>
<tr>
<th>Data source</th>
<th>Number of features</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Training</td>
</tr>
<tr>
<td>L-VIS/NIR</td>
<td>396</td>
<td>1060</td>
</tr>
<tr>
<td>S-VIS/NIR</td>
<td>396</td>
<td>1060</td>
</tr>
<tr>
<td>L-Chl-FKC</td>
<td>286</td>
<td>1060</td>
</tr>
<tr>
<td>S-Chl-FKC</td>
<td>286</td>
<td>1060</td>
</tr>
</tbody>
</table>

2.5 Classification models

2.5.1 Conventional machine learning methods

Support vector machine (SVM) is a widely used pattern recognition method (Burges, 1998). For linearly separable issues, linear classifier is developed. For non-linear classification, SVM maps the original data into high dimension by kernel functions and establish hyperplanes to maximally classify the closest training samples of different classes. In this study, SVM was used to compare deep learning approaches.

2.5.2 Deep learning method

Convolutional neural network (CNN) is a widely used deep learning algorithm. The extracted VIS/NIR spectra and Chl-FKC were used to identify the stress types in this study.

A 1D CNN architecture was designed as the base classifier for processing each single data source. This model was defined as CNN-S. Figure 1 shows the shallow CNN architectures used in this study. The first part was an attention layer, which operated based on equation (1).

\[
Y_{attention} = \theta_{RELU} \left( W_1 \cdot \theta_{RELU} \left( W_2 \cdot X + b_1 \right) + b_2 \right) \cdot X
\]  

where \( W_1, W_2, b_1 \) and \( b_2 \) are the trainable parameters (weights and bias) in the attention layer, \( \theta_{RELU} \) is the activation function, \( X \) denotes the input. The second part was a 1D-convolution block consisting of three 1D convolution layers (kernel size of 3, stride of 1, activated using RELU). To reduce the dimension of features, a max pooling layer (pool size of 2, stride of 2) was added after the first convolution layer. The third part was a fully connected neural network including three dense layers. The numbers of neurons were 512, 128, 5, respectively, using RELU activation. Furthermore, a batch normalization layer was added before each convolution and dense layer. The SoftMax Cross-Entropy was utilized as the loss function.
2.6 Fusion strategies

Leaves and stems showed different physiochemical characteristics under stresses. During the seedling stage of rice, the leaves and stems are all green, and the VIS/NIR spectra and Chl-FKC of leaves and stems are quite similar (Figs 4, 5). Since HSI and Chl-FI acquired different phenotyping traits based on different principles, and each technique can provide limited plant phenotyping information. Thus, if the fusion of HSI and Chl-FI can help identify rice stress types is worthy of investigation. There were two organs (leaves and stems) and two techniques (HSI and Chl-FI). Five different fusion strategies of data were conducted according to Table 2. It should be noted that L-VIS/NIR, S-VIS/NIR, L-Chl-FKC and S-Chl-FKC of the same sample were used for fusion.

Table 2. The fusion strategies used in this study

<table>
<thead>
<tr>
<th>Fusion strategy</th>
<th>L-VIS/NIR</th>
<th>S-VIS/NIR</th>
<th>L-Chl-FKC</th>
<th>S-Chl-FKC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fusion 1</td>
<td>✓‡</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Fusion 2</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fusion 3</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Fusion 4</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Fusion 5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

‡ The symbol ✓ means that the corresponding features were used for fusion, and the symbol × means that the corresponding features were not used for fusion.

This study explored three different levels of information fusion (low-level, mid-level and high-level). For each information fusion level, end-to-end deep fusion models were developed. For low-level fusion, the original data of each type of features were concatenated directly as a long vector and performed the classification task by the CNN-S model according to the fusion strategies in Table 2.

For mid-level fusion, an end-to-end CNN architecture for information fusion was proposed to fuse different types of original spectra for plant stress types identification. Each type of

Figure 1. The architecture of CNN model with single type of inputs. ‘1D-Conv’ means 1D convolution, ‘Max Pool’ denotes max pooling layer.
original features was fed into the corresponding feature extractor, and the model required only a one-stage training. The fusion strategies for the end-to-end CNN were based on Table 2. The mid-level fusion model fused the deep features for further classification. In Figure 2, the end-to-end CNN model for mid-level fusion using all the four types of features is presented. Four single CNN-S models were employed for processing different data sources. Specially, the last dense layer of the CNN-S was removed. Therefore, the 128-dimension output (4 vectors from the second dense layer of each CNN-S sub-model) was considered as the deep features. They were weighted and concatenated for fusion.

The mid-level fusion could be described using equation (2).

\[ Y_{mid} = \varphi_{Concat} \left( \theta_{\text{RELU}} \left( W_1 \right) F_1, \theta_{\text{RELU}} \left( W_2 \right) F_2, \theta_{\text{RELU}} \left( W_3 \right) F_3, \theta_{\text{RELU}} \left( W_4 \right) F_4 \right) \tag{2} \]

where \( \varphi_{Concat} \) means that this function concatenates all inputs to generate a vector; \( F_1, F_2, F_3, F_4 \) are the deep features from L-VIS/NIR, S-VIS/NIR, L-Chl-FKC, S-Chl-FKC, respectively.

![Figure 2. The architecture of end-to-end CNN model for mid-level fusion using all the four types of features. HSI denotes hyperspectral imaging, FL denotes fluorescence, 1D-Conv means 1D convolution. \( W_1 \) to \( W_4 \) were the weights for fusion.]

For high-level fusion, another end-to-end CNN model was designed (shown in Figure 3). In this paper, high-level fusion could also be understood as decision fusion, which provides the final decision on classification based on the output of different classifiers. Different from the mentioned mid-level fusion model, the high-level fusion model used four complete CNN-S models to process the four data sources. Then, the outputs of the four sub-models (four vectors with a shape of 5*1) were concatenated and further fed into another dense
layer with five neurons. It should be pointed out that a new loss function was proposed especially for this deep decision fusion model, see equation (3).

\[
\text{Loss} = \ell(y, y_F) + \ell(y, y_1) + \ell(y, y_2) + \ell(y, y_3) + \ell(y, y_4)
\]

where \(y_F\) denotes the output of the whole decision fusion-based deep learning model, \(y_1, y_2, y_3, \) and \(y_4\) are the output of four sub-models, \(\ell\) denotes the Cross-Entropy loss. This loss function could simultaneously promote the final classification loss and that of the sub-models as small as possible. Moreover, due to the constraint of loss function on the output of sub models, the fused \(y_1, y_2, y_3, y_4\) were forced to approach the expected probability distribution. Thus, the combination of them was a typical decision fusion.

![Diagram](image)

**Figure 3.** The architecture of end-to-end CNN model for end-to-end high-level fusion using all the four types of features. HSI denotes hyperspectral imaging, FL denotes fluorescence, 1D-Conv means 1D convolution. \(y_1\) to \(y_4\) were the outputs of the single model.

To evaluate the superiority of the proposed methods, two-stage mid- and high-level fusion approaches were also used for comparison. In this study, only the two-stage fusion of all the four types of features were implemented. As for mid-level fusion, features extraction, feature fusion and modeling were conducted separately. The features were learned and extracted from CNN-S models using each single type of features, and all the features extracted from the four types of features were concatenated. The fused features were then fed into a CNN-S model. As for high-level fusion, the predicted probability distribution vectors of the CNN-S models using four data sources were first extracted. Then, the averaged probability distribution vectors were used to make the final decision.

2.7 Model performance evaluation and software
The classification accuracy was evaluated by the ratio of the number of correctly classified samples to the number of total samples. SVM was conducted on scikit-learn (version:1.0.1) in python 3.7, and CNN was conducted using the MXNET framework (Amazon; Seattle, WA, United States) with python 3.7.

3. Results

3.1 Profiles of VIS/NIR spectra and Chl-FKC

Figure 4 shows the average VIS/NIR spectra of the leaves (Figure 4(A)) and stems (Figure 4(B)) of the healthy rice plants and the plants under the stresses of Cd, Cu, DCA and ELK. Figure 5 shows the average Chl-FKC of the leaves and stems (Figure 5(A)) and stems (Figure 5(B)). As shown in Figures 4 and 5, typical hyperspectral profiles of plants can be found for leaves and stems of rice plants, as well as the Chl-FKC. Differences can be found for leaves and stems under different stresses. However, no particular regulation (for example, which stress has higher reflectance or fluorescence intensity) could be found. There were variations during the measurement caused by the samples, instruments and measurement conditions, etc. Thus, it was difficult to identify the stress types by observing the differences in spectra curves or Chl-FKC. Further investigation should be conducted for stress types identification.

(A)
Figure 4. Average VIS/NIR spectra of the leaves (A); average VIS/NIR spectra of the stems (B)
3.2 Model establishment using a single type of features

The full VIS/NIR spectra and Chl-FKC of leaves and stems were used as inputs of SVM and CNN-S models. Table 3 shows the statistical results of the classification models. To construct SVM and CNN-S models, the training sets and validation sets were used, and the optimal models were selected according to the classification performances of the validation sets. For SVM, the kernel function was chosen as ‘rbf’. The optimization ranges of parameters C and g were both \([10^{-8}, 10^{-7}, 10^{-6}, ..., 10^{0}, 10^{7}, 10^{8}]\). For all deep learning models in this study, the number of epochs was set as 500. A scheduled learning rate was used by starting with 0.005 for the first 100 epochs, which was reduced to one tenth by every 100 epochs. The batch size was set as 128. The statistical results of SVM and CNN-S models are shown in Table 3.

Both CNN-S and SVM models performed well for L-VIS/NIR and S-VIS/NIR. CNN-S models obtained slightly better results than the corresponding SVM models, with classification accuracy of the training, validation and testing sets all over 90%. For L-Chl-FKC and S-Chl-FKC, both CNN-S models and SVM models failed to obtain satisfactory results, with the classification accuracy of the training, validation and testing sets over 70%. CNN-S models and SVM models obtained close results. It could be noted that overfitting occurred for both SVM and CNN models using chlorophyll fluorescence induction kinetic spectra. For VIS/NIR spectra, the classification models of leaves performed better than the corresponding models of stems, indicating that leaves were more suitable for rice stress types identification when using HSI. For Chl-FKC, the classification models of leaves obtained close results to those of stems.

Table 3. The classification results of SVM and CNN models using L-VIS/NIR, S-VIS/NIR, L-Chl-FKC, S-Chl-FKC

<table>
<thead>
<tr>
<th>Dataset type</th>
<th>Model</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Training</td>
</tr>
</tbody>
</table>

(B)
<table>
<thead>
<tr>
<th>Method</th>
<th>Model</th>
<th>L-VIS/NIR</th>
<th>S-VIS/NIR</th>
<th>L-Chl-FKC</th>
<th>S-Chl-FKC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SVM</td>
<td>100</td>
<td>96.2</td>
<td>92.9</td>
<td>99.9</td>
</tr>
<tr>
<td></td>
<td>CNN</td>
<td>98.8</td>
<td>98.4</td>
<td>100</td>
<td>99.9</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>100</td>
<td>91.7</td>
<td>92.0</td>
<td>75.8</td>
</tr>
<tr>
<td></td>
<td>CNN</td>
<td>95.1</td>
<td>92.0</td>
<td>94.7</td>
<td>76.2</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>87.9</td>
<td>91.3</td>
<td>74.3</td>
<td>75.5</td>
</tr>
<tr>
<td></td>
<td>CNN</td>
<td>92.7</td>
<td>92.0</td>
<td>94.7</td>
<td>76.2</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>94.7</td>
<td>75.5</td>
<td>75.5</td>
<td>76.2</td>
</tr>
<tr>
<td></td>
<td>CNN</td>
<td>92.0</td>
<td>76.2</td>
<td>76.2</td>
<td>76.2</td>
</tr>
</tbody>
</table>

Figure 6 shows the confusion matrix of CNN models using the four data types. No particular regulation could be found for the misclassification of samples among different classes. One sample under one type of stress could be misclassified as any other type of stress. The results of the confusion matrix showed the feasibility to identify the stress types of Cd, Cu, DCA and ELK using HSI and Chl-FI.
Figure 6. Confusion matrix of CNN-S models using VIS/NIR spectra of leaves and stems, Chl-FKC of leaves and stems, respectively. (A) is the confusion matrix of the training, validation and testing sets (from left to right) of CNN-S model using VIS/NIR spectra of leaves; (B) is the confusion matrix of the training, validation and testing sets (from left to right) of CNN-S model using VIS/NIR spectra of stems; (C) is the confusion matrix of the training, validation and testing sets (from left to right) of CNN-S model using Chl-FKC of leaves; (D) is the confusion matrix of the training, validation and testing sets (from left to right) of CNN-S model using Chl-FKC of stems.

3.3 Model establishment using fused datasets

For low-level fusion, the fused datasets were constructed according to the fusion strategies in Table 2. The fused datasets were also used as inputs of the CNN-S models with the same architectures as Figure 1. The parameter settings of the CNN-S models for fused datasets were the same as those of the CNN-S models for a single type of features. The results of the CNN models using the fused datasets are presented in Table 4. CNN-S model using the fusion of L-VIS/NIR and S-VIS/NIR obtain slightly better results than that using VIS/NIR spectra of leaves and stems individually. CNN-S model using the fusion of L-Chl-FKC and S-Chl-FKC obtained significantly better results than that using features of leaves and stems individually. For fusion of VIS/NIR spectra and chlorophyll fluorescence induction kinetic spectra of leaves, the performance of the CNN model was worse than that of the CNN-S model using L-VIS/NIR, and significantly better than that of the CNN-S model using L-Chl-FKC. A similar phenomenon could be found for stems. Classification models using the fusion of L-VIS/NIR and L-Chl-FKC showed better classification performances than the corresponding models using the fusion of S-VIS/NIR and S-Chl-FKC. CNN-S model using the fusion of VIS/NIR spectra and Chl-FKC of leaves and stems showed good performances, and the performances were close to the CNN-S model using L-VIS/NIR.

As for the end-to-end fusion approach for mid-level fusion, the original VIS/NIR spectra and Chl-FKC of leaves and stems were fed into an end-to-end deep learning fusion model (shown in Figure 2) according to the fusion strategies in Table 2. By fusing L-VIS/NIR and S-VIS/NIR, the CNN model obtained good results, with classification accuracy of the training, validation and testing sets over 90%. CNN model using the fusion of L-Chl-FKC and S-Chl-FKC obtained worse results, with classification accuracy of the training, validation and testing sets over 80%. When L-VIS/NIR and L-Chl-FKC were fused, the corresponding CNN model obtained good performances, with classification of all the three sets over 90%. The classification accuracy of the training, validation and testing sets in the CNN model using the fusion of S-VIS/NIR and S-Chl-FKC was over 80%. When VIS/NIR spectra and Chl-FKC of leaves and stems were all fused, the corresponding CNN model...
obtained the best performances, with classification accuracy of all the three sets over 95%. The CNN model using the fusion of L-VIS/NIR and S-VIS/NIR obtain slightly better results than that using L-VIS/NIR and S-VIS/NIR individually. The CNN model using the fusion of L-Chl-FKC and S-Chl-FKC obtained equivalent or better results than that using L-Chl-FKC and S-Chl-FKC individually. As for the two-stage fusion approach of mid-level fusion, the deep features extracted from the VIS/NIR spectra and Chl-FKC of leaves and stems by the corresponding CNN-S models were all used for fusion. Better performances were obtained, with the classification accuracy of the training, validation and testing sets all over 97%.

As for high-level fusion, an end-to-end CNN model for decision fusion was also applied. The classification results were the best, and the classification accuracy of the training, validation and testing sets was 100 %, 97.4 % and 97.7 %. A two-stage high-level fusion approach was also used for comparison, and close results were obtained. The classification accuracy of the training, validation and testing sets was 100 %, 97.7 % and 97.7 %.

The high-level fusion showed the best performances of rice stress types identification compared with low-level fusion and mid-level fusion. High-level fusion depended on the outputs of each single model. The results of the classification models using each single type of features illustrated the potential of different types of features for rice stress types identification. The modeling performances varied for each type of features. High-level fusion could reduce the interference by the defects of different decision models using different features, producing excellent performances. The overall results illustrated the effectiveness of high-level fusion for rice stress types identification.

Compared with end-to-end fusion models, the two-stage information fusion approaches were more complex. The features were extracted by the trained and optimized models for two-stage mid-level fusion. Moreover, one more classifier should be trained on the combined features. The whole procedure was complex, with more computation and manual intervention (here, 5 models should be trained and optimized). For two-stage high-level fusion, the predicted probability distribution vectors of the CNN-S model using each type of features were first extracted. A decision-making procedure was implemented to make the final decision. End-to-end models for information fusion showed simpler operation with only one-stage calculation.

As shown in Table 4, the models using the fusion of L-Chl-FKC and S-Chl-FKC obtained better results than those using L-Chl-FKC and S-Chl-FKC individually. The results indicated that combining Chl-FKC features could improve classification performances. Since the CNN-S models using L-VIS/NIR and S-VIS/NIR individually obtained good performances, the models using the fusion containing the features of L-VIS/NIR and S-VIS/NIR all obtained good performances. The results were not good when CNN-S models were built using L-Chl-FKC and S-Chl-FKC individually. The fusion of Chl-FKC features and the corresponding VIS/NIR features showed better results than the models using the corresponding Chl-FKC features, and obtained lower or close results than the models using the corresponding VIS/NIR features. Due to the gap between the performances of models using VIS/NIR spectra and corresponding Chl-FKC individually, the fusion of Chl-FKC features and the corresponding VIS/NIR features obtained the performances between the performances of the models using each type of features individually.
The fusion of all the features showed relatively better performances. By fusing all the types of features, the complementary information relating to rice stress types identification hidden in the obtained features could be revealed for rice stress types identification.

Table 4. The classification results of the CNN models of each fusion level

<table>
<thead>
<tr>
<th>Fusion level</th>
<th>Fusion strategy</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Training</td>
</tr>
<tr>
<td>Low-level</td>
<td>Fusion 1</td>
<td>99.9</td>
</tr>
<tr>
<td></td>
<td>Fusion 2</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Fusion 3</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Fusion 4</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Fusion 5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Fusion 5-TwoStage§</td>
<td>/</td>
</tr>
<tr>
<td>Mid-level</td>
<td>Fusion 1</td>
<td>97.6</td>
</tr>
<tr>
<td></td>
<td>Fusion 2</td>
<td>99.9</td>
</tr>
<tr>
<td></td>
<td>Fusion 3</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>Fusion 4</td>
<td>99.6</td>
</tr>
<tr>
<td></td>
<td>Fusion 5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Fusion 5-TwoStage</td>
<td></td>
</tr>
<tr>
<td>High-level</td>
<td>Fusion 1</td>
<td>99.6</td>
</tr>
<tr>
<td></td>
<td>Fusion 2</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Fusion 3</td>
<td>99.5</td>
</tr>
<tr>
<td></td>
<td>Fusion 4</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Fusion 5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Fusion 5-TwoStage¶</td>
<td>100</td>
</tr>
</tbody>
</table>

§ Low-level fusion directly fuses the raw data as the input of the model, which does not need a two-stage training. || Two-stage-based mid-level fusion model fuses the deep features extracted by single models (CNN-S). Another CNN classifier is trained for processing the fused features. ¶ Two-stage-based high-level fusion model fused 4 predicted probability distribution vectors (each includes 5 elements). The averaged probability distribution vectors are for making the final decision.

Furthermore, the confusion matrices of the three levels of information fusion using the Fusion 5 strategy were explored (shown in Figure 7). Although Cd and Cu were both heavy metal stresses, and DCA and ELK were herbicide stresses, no particular rules could be found for the misclassification of samples among different classes. The good performances of the classification of rice plants under different stress types showed that information fusion of phenotyping traits of leaves and stems acquired by HSI and Chl-FI had great potential for rice stress types identification.
Figure 7. Confusion matrices of the three levels of end-to-end information fusion using the Fusion 5 strategy. (A) is the confusion matrix of the training, validation and testing sets (from left to right) of low-level fusion; (B) is the confusion matrix of the training, validation and testing sets (from left to right) of mid-level fusion; (C) is the confusion matrix of the training, validation and testing sets (from left to right) of high-level fusion.

4. Discussion
The effectiveness of HSI and Chl-FI for high-throughput plant stress phenotyping has been widely verified in various studies (Ferreira de Sousa et al., 2017; Kalaji et al., 2018; Lowe et al., 2017; Luisa Perez-Bueno et al., 2019; Mahlein et al., 2018; Moustakas et al., 2021; Saric et al., 2022). In this study, VIS/NIR spectra acquired by HSI and chlorophyll fluorescence induction kinetic spectra acquired by Chl-FI of leaves and stems of rice plants were used to identify the stress types, and the results were promising.

In previous studies, SVM has been widely used as a conventional machine learning method to compare with deep learning methods for 1D spectra analysis. The results between SVM and deep learning methods were quite close (Fernandes et al., 2019; Qiu et al., 2018; Tarandeep et al., 2021; Yan et al., 2020). This study found similar trends for plant stress types identification using VIS/NIR spectra and chlorophyll fluorescence induction kinetic spectra of leaves and stems. The 1D VIS/NIR spectra and chlorophyll fluorescence induction kinetic spectra have simple data structures, and the potential of CNN models for deep feature learning could not be fully revealed. On the other hand, the number of samples was not large enough, which were suitable for SVM. The potential of CNN models for big data could not be fully revealed.

Moreover, variations can be found between VIS/NIR spectra and chlorophyll fluorescence induction kinetic spectra for stress types identification. As for HSI, the classification
performances of CNN models showed slightly better results than those of SVM models. As for Chl-FI, the classification performances of CNN models using Chl-FKC showed close results to those of SVM models. Models using VIS/NIR spectra performed better than those using chlorophyll fluorescence induction kinetic spectra. In some other studies, the performances for plant phenotyping using features of HSI were better than those using features of Chl-FI (Kasampalis et al., 2021; Kumar et al., 2021; Mahlein et al., 2019). Although spectral indices calculated from VIS/NIR spectra are widely used for analysis, the full range spectra are also widely used for analysis. Unlike VIS/NIR spectra, Chl-FI parameters calculated from Chl-FKC were more widely used for analysis in Chl-FI rather than the full range Chl-FKC. The results illustrated the potential of stress types identification using Chl-FKC, and more efforts should be made to improve the performances. VIS/NIR can reflect the physiological and biochemical changes under different stresses, and Chl-FI is an efficient technique to assess the status of plant photosynthesis. VIS/NIR spectra may provide more information than Chl-FKC.

Both leaves and stems showed promising results for plant stress types identification, and the performances of leaves were better than those of stems. In general, leaves were more likely to be used for plant phenotyping. Fewer studies have studied the high-throughput phenotyping of stems (Fan et al., 2019; Garcia-Martin et al., 2020; Jin et al., 2019; Kong, Zhang, Huang, et al., 2018; Nagasubramanian et al., 2018; Wang et al., 2011; Xiang et al., 2021; Zhang et al., 2021). The overall results showed that stems can be an efficient alternative for plant phenotyping in addition to leaves. Stems and leaves showed different phenotyping traits under different stresses. The good performances indicated that the combinations of leaves and stems have the great potential to provide more information for plant stress phenotyping.

HSI and Chl-FI acquired different phenotyping traits of leaves and stems. The fusion of VIS/NIR spectra features and Chl-FKC features combined different aspects of features of plants. The results showed that using the fusion of VIS/NIR spectra features and Chl-FKC features had the potential to improve the performances of plant stress types identification. In some other studies, the authors also fused data from different sensors to improve the performances of plant phenotyping (Huang et al., 2018; Maimaitijiang et al., 2017; Moshou et al., 2005).

With the advantage of deep learning on feature learning and classification tasks, end-to-end CNN models were designed to implement the information fusion of phenotyping traits acquired from leaves and stems by HSI and Chl-FI. The most widely used two-stage information fusion consisted of training or manually selecting features, concatenating features manually and training an independent classifier using fused features. The end-to-end deep fusion models combined all these steps into one model, and the whole procedure was implemented using one-stage training. These models were much simpler, and did not need manual intervention to extract or select appropriate features. By obtaining the phenotyping traits, these phenotyping traits could be directly fed into the end-to-end deep fusion models, which had great potential for real-world application.

With the development of advanced phenotyping techniques, various aspects of phenotyping traits can be obtained. How to fully reveal the information relating to plant growth status from various aspects of phenotyping traits is of importance. Information fusion provided an effective alternative to combine different aspects of phenotyping traits to reveal the phenotyping information related to plant growth status fully. The effectiveness of
information fusion for rice stress types identification that information fusion was promising for plant growth status evaluation. On the other hand, different organs have different phenotyping traits, even measured by the same techniques. Thus, different organs could be analyzed to evaluate the plant growth status. The combination of phenotyping traits of different organs will provide more information of plants, and how to effectively fuse the phenotyping traits of different organs remains a challenging issue. This study provided an efficient alternative for the fusion of the phenotyping traits of different organs for plant phenotyping, which could enhance the high-throughput phenotyping performances, including the abiotic stresses caused by environmental factors.

Acknowledgments

Author contributions: Chu Zhang: Conceptualization; Data curation; Formal analysis; Funding acquisition; Methodology; Supervision; Investigation; Writing - original draft; Writing - review & editing Lei Zhou: Data curation; Formal analysis; Methodology; Software; Validation; Visualization; Writing - original draft; Writing - review & editing Qinlin Xiao: Methodology; Investigation Xiulin Bai: Methodology; Investigation Baohua Wu: Formal analysis; Methodology; Na Wu: Validation; Yiying Zhao: Writing - original draft; Junmin Wang: Resources Lei Feng: Conceptualization; Funding acquisition; Project administration; Software; Supervision; Writing - review & editing

Funding: This study was supported by the National Natural Science Foundation of China (61705195), Shenzhen Science and Technology Projects (2021304895).

Competing interests: The authors declare that there is no conflict of interest regarding the publication of this article.

Data Availability: The data used in this study are available from the corresponding author upon reasonable request.

References


Lim, K., Pan, K., Yu, Z., & Xiao, R. H. (2020). Pattern recognition based on machine learning identifies oil adulteration and edible oil mixtures. *Nature Communications, 11*(1), Article 5353. [https://doi.org/10.1038/s41467-020-19137-6](https://doi.org/10.1038/s41467-020-19137-6)


