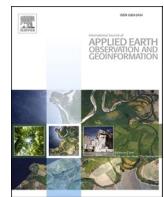




Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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



A joint learning Im-BiLSTM model for incomplete time-series Sentinel-2A data imputation and crop classification

Baili Chen ^{a,b,c}, Hongwei Zheng ^{a,b,c,*}, Lili Wang ^{a,b,c}, Olaf Hellwich ^d, Chunbo Chen ^{a,b,c}, Liao Yang ^{a,b,c}, Tie Liu ^{a,b,c}, Geping Luo ^{a,b,c}, Anming Bao ^{a,b,c}, Xi Chen ^{a,b,c}

^a State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China

^b University of Chinese Academy of Sciences, Beijing, China

^c Research Center for Ecology and Environment of Central Asia, Chinese Academy of Sciences, Urumqi, China

^d Department of Computer Vision & Remote Sensing, Technische Universität Berlin, 10587 Berlin, Germany



ARTICLE INFO

Keywords:

Crop classification
Multi-temporal
Data imputation
Joint learning
Bidirectional LSTM
Interpretation

ABSTRACT

Multi-temporal deep learning approaches can make full use of crop growth patterns and phenological characteristics, resulting in excellent crop classification performance in large areas. However, obtaining complete time-series remote sensing images during the growing season is challenging due to cloud contamination. Hence, given time-series multispectral data, it is important to impute missing data and accurately classify crops. A novel Imputation-BiLSTM model (Im-BiLSTM) was developed based on Bidirectional Long Short-term Memory network (BiLSTM) to jointly perform missing data imputation and crop classification. The Im-BiLSTM model regards missing data as network variables, which are efficiently updated during backpropagation. Im-BiLSTM treats the interaction between imputation and classification tasks, reducing the error and uncertainty caused by the separation operation of imputation to classification. Furthermore, we improved the interpretability of the Im-BiLSTM model by evaluating the importance of input features and visualizing hidden state units. In Shawan County, Xinjiang, China, we acquired a total of 10 Sentinel-2A images from April to October 2016, of which 3 images lost partial data due to cloud cover. The Im-BiLSTM model was applied to incomplete time-series data containing 10 time-steps for pixel-level crop classification, and the BiLSTM model was constructed based on cloud-free images for comparison. The performance of the proposed model was tested in four different cases of images missing and missing rates. The results showed that the classification of the Im-BiLSTM model outperformed the BiLSTM model, the overall accuracy was improved by a maximum of 4.2%, and the F1-scores of spring corn and tomato was improved by a maximum of 16.1% and 21.4%, respectively. Therefore, the Im-BiLSTM model can effectively improve classification performance by jointly imputing missing data. The imputation results (the coefficient of determination values range 0.4 ~ 0.9) indicated that the bands with the larger contribution to the classification had higher imputation accuracy. Feature importance evaluation showed that the Im-BiLSTM model captured the key phenological periods and features from time-series input. The visualization of hidden units demonstrated that the Im-BiLSTM model accumulated useful information over time, and the learned high-level features made the crops more separable than the original inputs.

1. Introduction

Accurate and timely crop classification can provide basic data for applications such as crop area estimation, yield forecasting, and food security assessment (Karthikeyan et al., 2020; Waldner et al., 2015). It has important scientific significance and practical value for various decision-making by the government and the private departments, including land adaptability evaluation, ecosystem services regulation,

agricultural subsidies and crop insurance (Estel et al., 2015; Johnson et al., 2014). Recently, the rapid development of earth observation satellites has produced a large number of remote sensing images with high temporal, spatial and spectral resolution, which has greatly promoted cost-effective, large scale and accurate crop classification (Moody et al., 2017; Song et al., 2017; Veloso et al., 2017).

Each crop has its unique phenological characteristics and seasonal growth patterns, which is the key to distinguishing different crops based

* Corresponding author.

E-mail address: hzheng@ms.xjb.ac.cn (H. Zheng).

on remote sensing (do Nascimento Bendini et al., 2019; Hu et al., 2019). The intra-class variability and inter-class similarity among crop spectra make it difficult to classify crops using only single-date images. Multi-temporal remote sensing technology has excellent classification performance because it can reflect the spectral difference of crop growth (Foerster et al., 2012; Vuolo et al., 2018). Vuolo et al. (2018) demonstrated that multi-temporal Sentinel-2 data effectively improved crop classification performance compared to single-date image. Piedelobo et al. (2019) focused on combining Sentinel-2 and Landsat-8 images to obtain dense time-series data for large-area crop classification. Bargiel (2017) proposed the phenological sequence patterns method based on the integration of time-series radar images and crop phenology information. However, formulating rules and models to extract temporal features from time-series data based on expert agronomic knowledge, such as statistical values and phenological metrics, is time-consuming and lacks flexibility and automation (Mou et al., 2018; Pelletier et al., 2019; Zhong et al., 2019). How to make full use of the sequential relationships and abundant seasonal patterns of time-series remote sensing observations is essential for identifying crops (Li et al., 2021).

Recently, deep learning has been widely explored in remote sensing applications due to its powerful end-to-end automatic extraction of high-level features from raw input data, reducing human feature engineering intervention (Kussul et al., 2017; Lei et al., 2021; Zhu et al., 2017). As one of the most prominent architectures of deep learning, Recurrent Neural Network (RNN) is good at processing sequence data and learning the correlation between time-series information (Zhong et al., 2019). Long Short-term Memory network (LSTM) is an improved structure of RNN, which alleviates the gradient vanishing problem of RNN and captures the long-term dependence of time-series data by introducing a gating mechanism (Hochreiter and Schmidhuber, 1997). Furthermore, LSTM has a global view by introducing a bidirectional flow (BiLSTM), which can learn time-series data from forward and backward directions, achieving state-of-the-art performance in crop classification. Zhou et al. (2019) achieved multi-temporal crop classification using the LSTM model in southern China with complex geographic environment and planting patterns. Zhao et al. (2020) proposed an improved model based on LSTM and convolutional neural network to deeply synergize Sentinel-2 optical and Sentinel-1 SAR data to obtain dense time-series data. Xu et al. (2020) developed an attention-based LSTM model for multi-temporal crop classification, and verified that the model had spatial generalization in different regions and outperformed other machine learning models. Crisóstomo de Castro Filho et al. (2020) compared LSTM, BiLSTM and traditional machine learning models in rice crop detection, and the results showed that the BiLSTM model had better performance. In addition to high performance, the interpretability of deep learning models is another critical aspect, which provides users with reliability. Recently, some interpretability methods of the LSTM model have been proposed. For example, perturbation analysis of time-series input (Campos-Taberner et al., 2020), attention mechanisms (Pruthi et al., 2019), gradient-based computation (Xu et al., 2020), hidden units visualization (Pérez-Suay et al., 2020).

However, the acquisition of complete time-series images during crop growing season is challenging due to cloud cover. Cloud contamination causes image data missing, especially during critical phenological periods, which limits the accuracy of crop classification (Yan et al., 2021; Zhao et al., 2020). Currently, many studies usually adopt a two-step strategy, that is, first impute missing data and then perform crop classification. Conventional interpolation methods could be divided into three groups: spatial, temporal, and spatiotemporal (Chen et al., 2017; Zhou et al., 2022). The spatial interpolation method uses pixels adjacent to the missing area to fill in the missing data, which is not suitable for larger missing areas (Cheng et al., 2014; Shen et al., 2015). The temporal interpolation method uses the adjacent date image data to impute missing data, which requires time-series data fluctuate regularly, without mutation and with small time intervals (Julien and Sobrino, 2010; Roy and Yan, 2020). Spatial-temporal data fusion based on multi-

source satellite images has been used to obtain complete dense time-series images, such as MODIS and Landsat (Moreno-Martínez et al., 2020), Optical and SAR images (Gao et al., 2021; Zhou et al., 2019). This method is computationally intensive and time-consuming. The above interpolation methods have certain assumptions and limitations, and cannot guarantee the quality of imputed missing data, which may degrade the downstream classification performance. Furthermore, these imputation-then-classification strategies ignore the interaction between imputation and classification tasks, which may result in suboptimal results.

In many real applications, imputation of missing data can be considered a secondary task, while classification application performed on imputed dataset is the primary purpose. In fact, we cannot directly assess the quality of the imputation data due to missing original data. During the imputation process, when the classification application as the main task achieves better performance, it in turn improves the accurate estimation of the missing data. Thus, a joint learning strategy was proposed to perform simultaneously missing data imputation and classification application in a multi-task learning framework (MTL) (García-Laencina et al., 2013). Cao et al. (2018) proposed to perform imputation and prediction applications simultaneously in a bidirectional-RNN, and the results showed that the model achieved higher performance in both imputation and downstream prediction tasks. Kim et al. (2021) proposed a novel end-to-end MTL framework based on generative adversarial network and RNN to impute missing data and perform downstream classification jointly in time-series data, achieving state-of-the-art performance on real-world datasets. This MTL jointly strategy has achieved success in some fields (Che et al., 2018; Ma et al., 2020), but has not been validated in crop classification with incomplete time-series remote sensing data.

This study proposed a novel Im-BiLSTM model, which jointly performs missing data imputation and crop classification. We applied the Im-BiLSTM model to Shawan County, Xinjiang, China, which mainly includes 5 crops. Our study aims to improve crop classification performance based on incomplete Sentinel-2A time-series data. In the Im-BiLSTM model, missing data is regarded as network variables, which are effectively updated during error backpropagation. The Im-BiLSTM model optimized the imputation and classification errors together and alleviated the error from imputation to classification, making the classification more accurate. Furthermore, we improved the Im-BiLSTM model interpretability by evaluating input features importance and visualizing hidden state units.

2. Materials and methods

2.1. Study area

Shawan County ($84^{\circ}98' - 86^{\circ}09'E$, $44^{\circ}12' - 45^{\circ}02'N$) (Fig. 1) is located in northern Xinjiang, China. The region has a temperate continental arid climate, with a mean annual temperature of $6.3 - 6.9^{\circ}C$, a mean annual precipitation of $140 - 350$ mm, and a mean annual evapotranspiration of $1500 - 2000$ mm. Shawan County implements single-season planting and is one of the largest agricultural counties in Xinjiang, with a cultivated area of 360,000 ha. The main crops include cotton, winter wheat, spring wheat, spring corn and tomato, among which cotton has the largest planting area.

2.2. Data

2.2.1. Field survey data

We conducted a field survey on land use types in July 2016, and obtained the spatial distribution information of crops in Shawan County. GPS was used to record the longitude and latitude location of each plot, and each land type was assigned a specific label (a total of 20 labels). However, some labels only applied to a few parcels, except for the 5 crop types of interest, other land types (a few fruit and vegetable parcels,

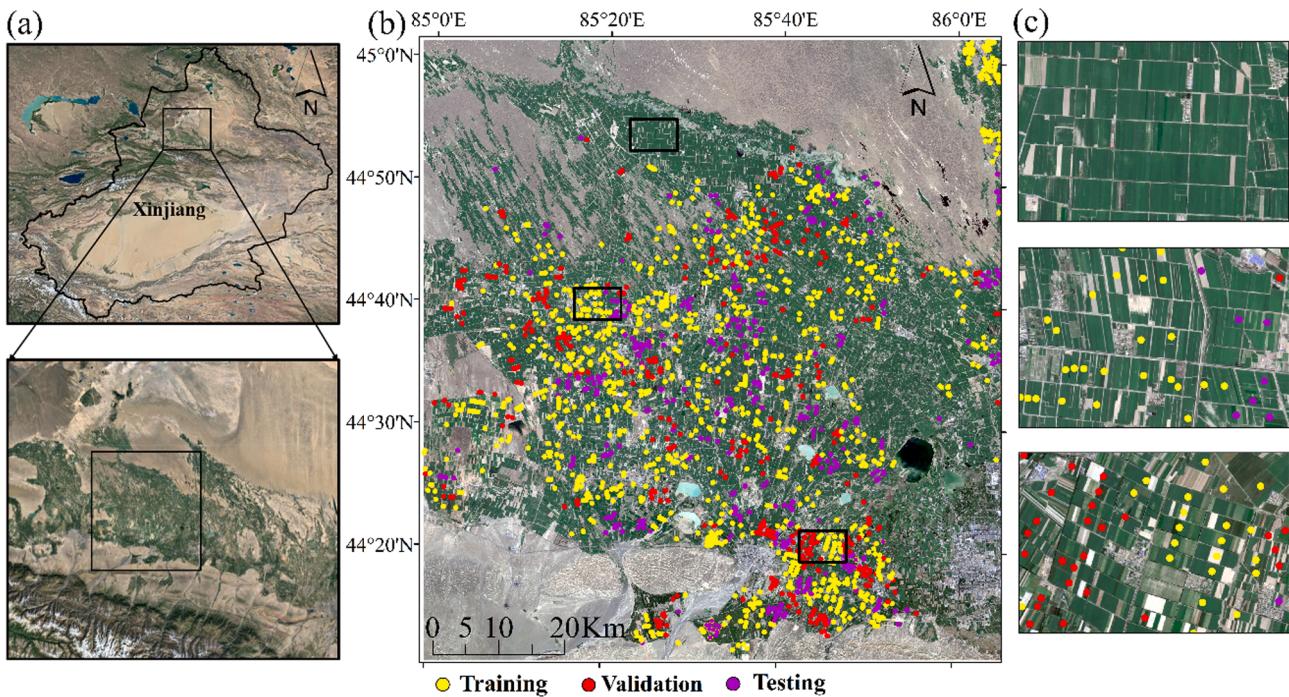


Fig. 1. (a) The image from Google Earth indicates the location of the study area, (b) a true-color composite image of Shawan County based on the Sentinel-2A optical image on August 10, and dots represent sampling parcels, and (c) detailed display of land parcels.

water bodies, residential areas, roads, forests and wastelands, etc.) were merged into the “Other” label. A total of 2697 parcels were obtained, including 867 cotton plots, 362 winter wheat plots, 111 spring wheat plots, 448 spring corn plots, 376 tomato plots and 533 “Other” plots. Our study focused on the pixel-level crop mapping. In order to meet the principle of independent and identical distribution, we used the block-wise random sampling method to divide the training, validation and testing sets at a ratio of 70%:15%:15% (Rußwurm and Körner, 2017). The number of parcels and pixels is shown in Table 1.

We also investigated the growth phenology of 5 crops. Each crop type has its specific phenological sequence and stages with different duration (Fig. 2). Winter wheat has the longest growth cycle. It is usually sown in September each year, turns green in March-April of the following year, and matures and harvested in July. Spring wheat is sown at the end of March and early April each year, and harvested at the end of July. Cotton, spring corn and tomato are spring crops, sown in early April. Cotton has the longest growing cycle and cannot be harvested until October. Tomato generally completes its life cycle in late August, and spring corn usually ends in mid-September. It is helpful to classify crops by analyzing crop phenological differences.

2.2.2. Sentinel-2A time-series images

The European Space Agency provides free access to Sentinel-2A images. Sentinel-2A satellite provides multispectral images with 13 bands, and its revisit period is 10 days. The specificity of Sentinel-2A

images can be found in Drusch et al. (2012). In order to obtain relatively complete time-series images during the growing season (from April to October), each month should include 1–2 scene images (Fig. 2). We obtained a total of 10 images on April 22, May 22, June 11, July 21, July 28, August 10, August 30, September 6, September 19, and October 6, 2016. However, images on July 21, August 30, and September 19 were heavily covered by clouds (Fig. 3), resulting in data loss. We performed atmospheric correction on 10 images to obtain bottom-of-atmosphere reflectance data and performed cloud masking. Then 10 optical bands including Blue, Green, Red, Red-edge1, Red-edge2, Red-edge3, Near-infrared1 (NIR1), Near-infrared2 (NIR2), Shortwave infrared1 (SWIR1) and Shortwave infrared2 (SWIR2) of each image were used as the input features for the Im-BiLSTM model. We used the Nearest Neighbor Interpolation method to resample the 20 m spatial resolution band to 10 m.

2.3. Joint imputation and classification by the Im-BiLSTM model

2.3.1. Pixel-level time-series multi-spectral samples

In the acquired time-series Sentinel-2A images, each pixel has two attributes of time and spectrum. A time-series multi-spectral pixel with D features of time length T was denoted as $X = (x_1, x_2, \dots, x_T) \in R^{T \times D}$, where $x_t \in R^D$ represents the D features observed at time-step t. D features denote the spectral reflectance of 10 bands, and T represents 10 time-steps from April 22 to October 6 in turn, as shown in the complete

Table 1

The number of parcels and pixels of the land type used for classification.

Crop type	Training			Validation			Testing		
	Parcel	Pixel	Proportion (%)	Parcel	Pixel	Proportion (%)	Parcel	Pixel	Proportion (%)
Cotton	586	88,991	36.5	142	31,557	38.0	139	28,739	39.4
Winter wheat	243	25,450	10.5	63	8,629	10.4	56	7,751	10.6
Spring wheat	78	10,817	4.4	20	4,176	5.0	13	2,949	4.0
Spring corn	301	32,518	13.4	80	11,029	13.3	67	7,378	10.1
Tomato	258	21,056	8.6	60	7,641	9.2	58	6,574	9.0
Other	378	64,690	26.6	78	20,001	24.1	77	19,482	26.7
Total	1844	243,522	100	443	83,033	100	410	72,873	100

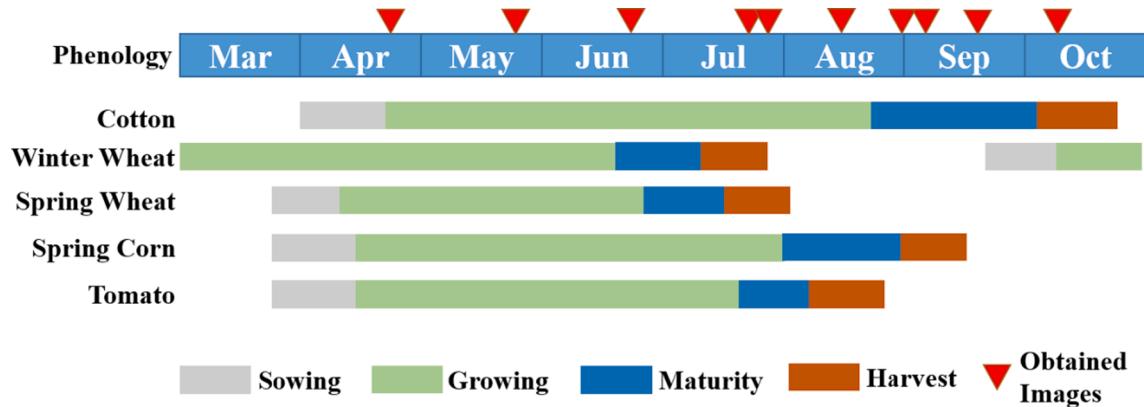


Fig. 2. Phenological sequence of 5 crops in Shawan County. Red triangle represents the Sentinel-2A optical image acquired monthly.

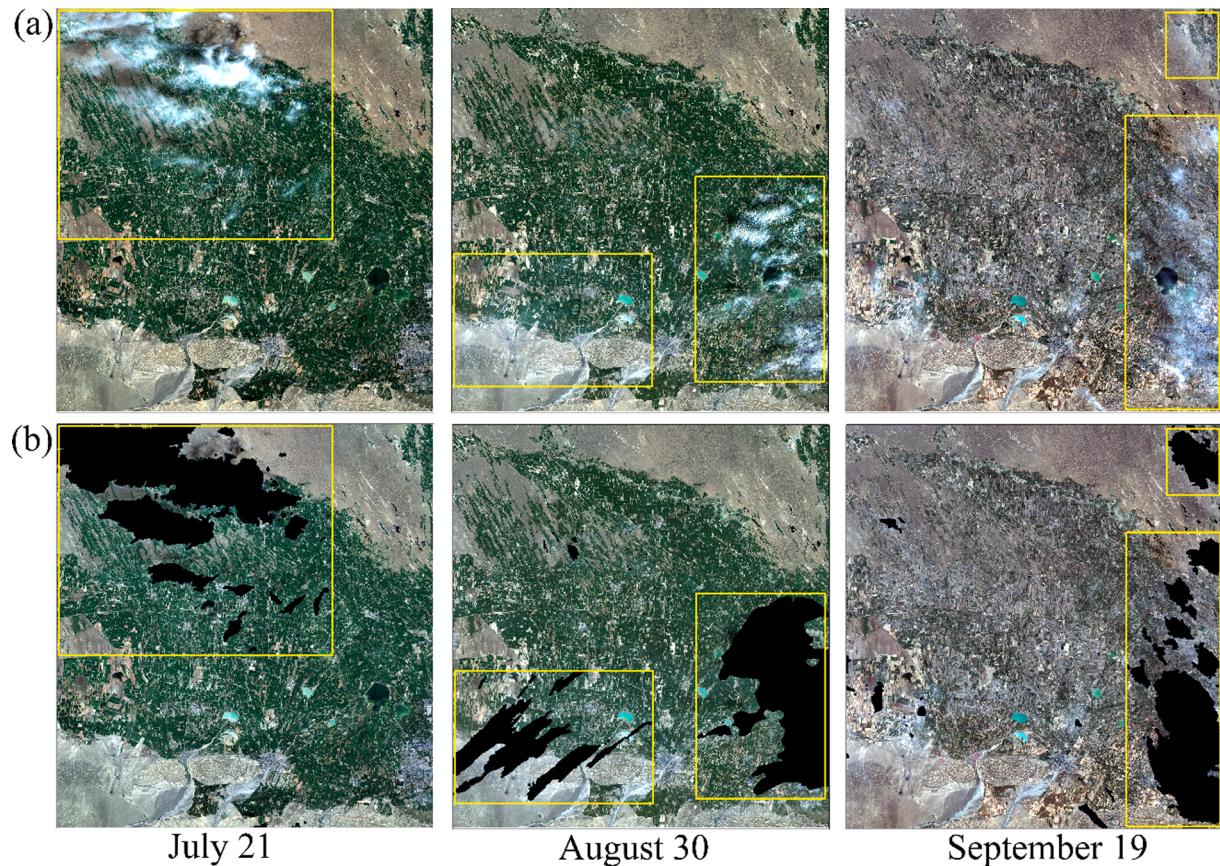


Fig. 3. Missing data for July 21, August 30, and September 19 images due to cloud contamination. (a) Optical images before cloud mask processing, and (b) after the cloud mask processing, all band data in the masked area are lost and denoted as Nodata in computing.

case of Fig. 4(a).

2.3.2. Data missing patterns

As shown in Fig. 3, data missing occurred in different areas of the 3 images, and the same area of the images on August 30 and September 19. Therefore, the time-series pixel has the following four missing patterns: 1) data missing only on July 21, 2) data missing only on August 30, 3) data missing only on September 19, and 4) data missing both on August 30 and September 19, as shown in Fig. 4. In our experiment, we first obtained the pixels with complete time-series from the labeled plots of 10 images. Then in the complete time-series dataset, for the 4 missing patterns observed on the ground, the random data missing rate was set to 30% for July 21, August 30 and September 19, and the simultaneous

data missing rate was set to 10% for August 30 and September 19, respectively. In addition, we also manually set more missing cases, including missing 3 to 5 images. We identified additional three missing cases (2, 3, 4) in Table 2 by evaluating which image dropout combinations had the most impact on the classification performance of the BiLSTM model. We then set 30%, 50% and 70% missing rates for each image of these three cases respectively, thus adding 9 experiments to evaluate the performance of the Im-BiLSTM model under more missing patterns and higher missing rates. We randomly eliminated the corresponding proportional observation data from the dataset as the imputation truth-value for evaluating the imputation accuracy. Finally, these data with complete and incomplete patterns were input to the Im-BiLSTM model to perform imputation and classification tasks. We

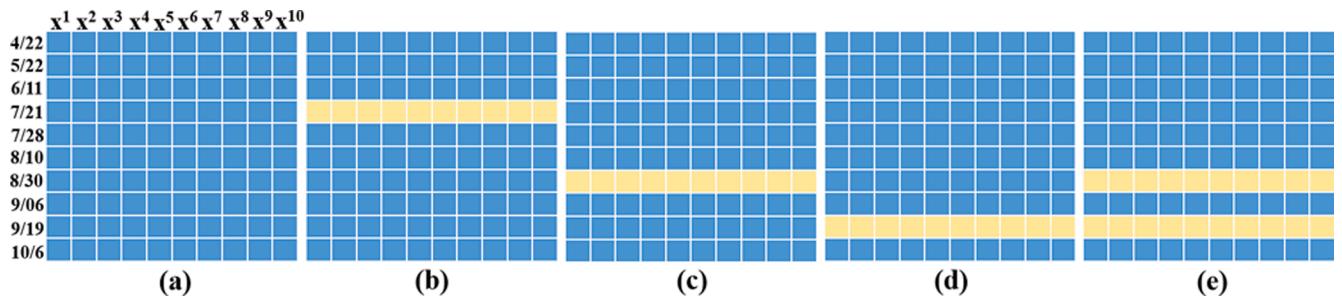


Fig. 4. The representation pattern of the pixel-level time-series multispectral sample. (a) Complete case without data missing, (b) data missing on July 21, (c) data missing on August 30, (d) data missing on September 19, and (e) data missing on August 30 and September 19.

Table 2
The cases of image data missing.

Missing cases	Case1	Case 2	Case 3	Case 4
Date	July 21	July 21	June 11	May 22
	August 30	August 30	July 21	June 11
	September 19	September 06	August 30	July 21
Missing rate (MR)	Ground case	30%; 50%; 70%	30%; 50%; 70%	30%; 50%; 70%

need to build a mask vector m_t to inform the Im-BiLSTM model of the specific missing pattern of input data.

$$m_t = \begin{cases} 0, & \text{if } x_t \text{ missed}, \\ 1, & \text{otherwise} \end{cases} \quad (1)$$

2.3.3. The Im-BiLSTM model architecture

In this section, we proposed the Imputation Bi-directional LSTM (Im-BiLSTM) model, which jointly performs missing data imputation and crop classification. Fig. 5(a) shows the Im-BiLSTM model architecture: (1) the input layer is composed of pixel-level time-series multispectral data containing 10 time-steps, (2) a BiLSTM layer with 64 hidden units, (3) the regression layer imputes all features in the time-series input through a fully connected layer, and (4) in the output layer, a fully connected layer with Softmax function to perform crop classification. Next, we specifically introduce the working principle of the Im-BiLSTM model.

As shown in Fig. 5(a), in the Im-BiLSTM model, missing data is regarded as variables in the network, which are directly learned from input data and effectively updated when the error passes during back-propagation. The Im-BiLSTM model can learn the time-series input from the forward ($x_1 \rightarrow x_{10}$) and backward ($x_{10} \leftarrow x_1$) directions. At time-step t , Im-BiLSTM can obtain an imputed vector \hat{x}_t based on the forward $t - 1$ steps and the backward $10 - t$ steps inputs. If x_t is actually observed, we directly calculate its imputation error l_t and pass x_t to the next time-step. Otherwise, since the next time-step imputation is related to the current time-step variables, we replace x_t with the imputed \hat{x}_t and pass \hat{x}_t to the next time-step. Therefore, the imputation error of the missing variables \hat{x}_t depends on the imputation error of the true observations at the next time-step. Thus, the imputation errors obtained in the forward and backward directions can be backpropagation along the opposite direction to optimize the imputation accuracy.

The gating mechanism adaptively determines the extent to which the LSTM retains the previous state and memorizes the current information. Each LSTM cell consists of three gates (Fig. 5(b)), the forget gate, the input gate and the output gate, which control the cell state C_t and the hidden state h_t . The cell state indicates the useful information accumulated over time and the hidden state indicates the information output from the current to the next time-step. At each time-step, the LSTM cell

reads input data x_t and previous hidden state h_{t-1} . A regression layer is used to impute the input data to obtain the complete input vector x_t^c and the mean absolute error function is used to calculate the imputation error. Initialize the initial hidden state to an all-zero tensor.

$$\hat{x}_t = W_x h_{t-1} + b_x \quad (2)$$

$$x_t^c = m_t \odot x_t + (1 - m_t) \odot \hat{x}_t \quad (3)$$

$$l^i = \frac{\sum_{t=1}^T (m_t \odot |\hat{x}_t - x_t|)}{\sum_t m_t} \quad (4)$$

where W_x and b_x are learnable weight matrices, l^i denotes imputation error, T represents the total time-step.

The forget gate vector determines how much previous information is retained:

$$f_t = \sigma(W_f \bullet [h_{t-1}, x_t^c \odot m_t] + b_f) \quad (5)$$

The input gate vector determines how much of the current time-step information is stored in the memory units:

$$i_t = \sigma(W_i \bullet [h_{t-1}, x_t^c \odot m_t] + b_i) \quad (6)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_c \bullet [h_{t-1}, x_t^c \odot m_t] + b_c) \quad (7)$$

The output gate vector determines how much of the current time-step information is output:

$$o_t = \sigma(W_o \bullet [h_{t-1}, x_t^c \odot m_t] + b_o) \quad (8)$$

$$h_t = o_t \odot \tanh(C_t) \quad (9)$$

where $W_f, W_i, W_c, W_o, b_f, b_i, b_c$ and b_o are learnable weight matrices, and σ and \tanh are sigmoid and hyperbolic tangent activation functions. The \odot operation denotes elementwise multiplication. The \circ operation indicates the concatenate operation, which can inform the model at which time-step data is lost. In the BiLSTM model, the hidden features from the forward and backward directions [$[h_t^{forward}, h_t^{backward}]$] are grouped together to obtain the final hidden vector h_t .

In the output layer, a fully connected layer with the Softmax function (Eq. (10)) receives hidden feature vector h_t to produce a normalized prediction score p_i . The class with the highest probability is the predicted class.

$$p_i = \frac{\exp(u_i h_t)}{\sum_{j=1}^K \exp(u_j h_t)} \quad (10)$$

where u_i and u_j are the output projection matrix, K denotes the number of classes.

The cross-entropy function (Eq. (11)) calculated the loss between prediction and true label. In the Im-BiLSTM model, our main task is to conduct crop classification, so the weight of classification loss should be more important than imputation loss to ensure optimal classification

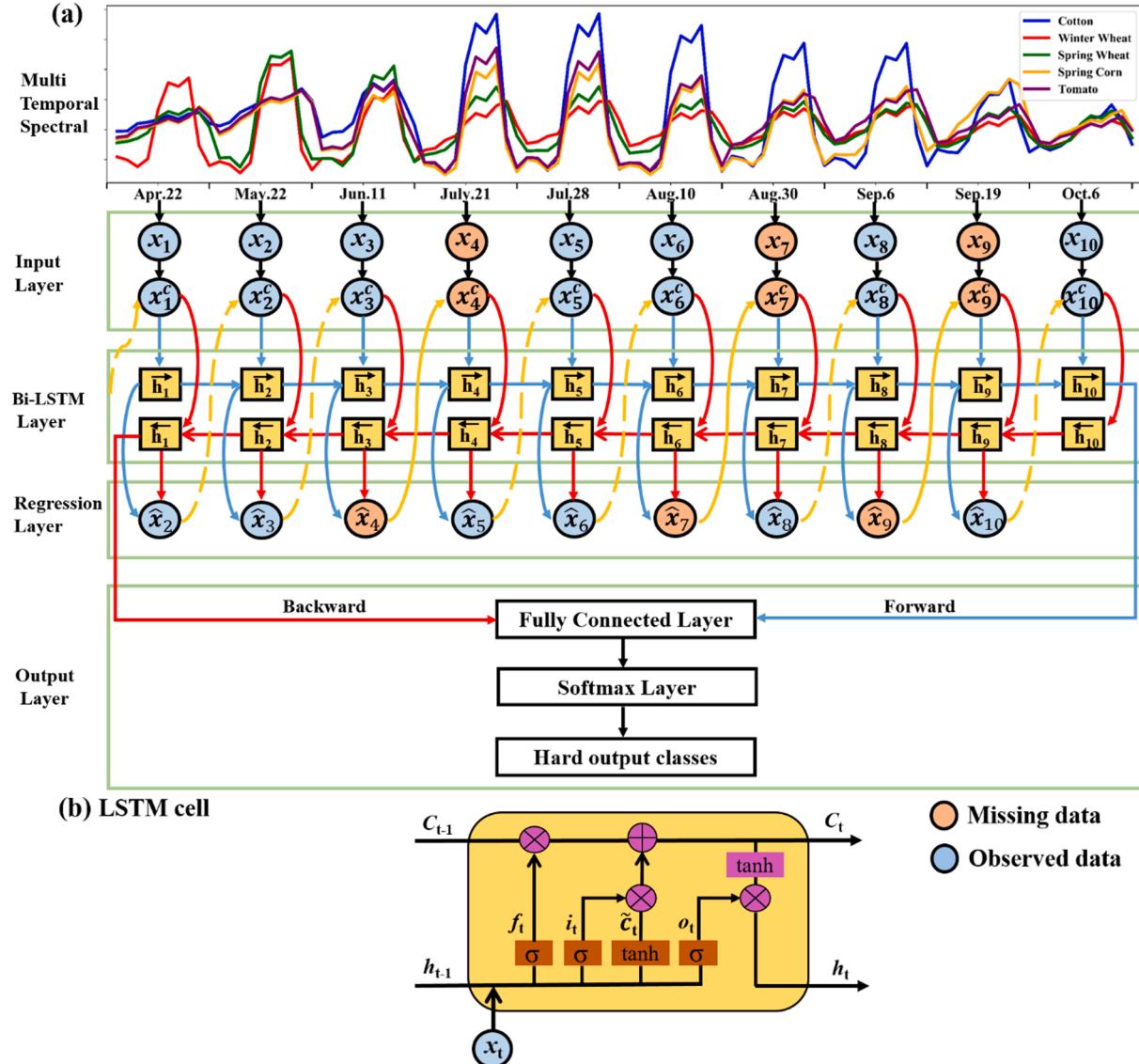


Fig. 5. (a) The architecture of the Im-BiLSTM model. The orange circle indicates that features at that time-step are missing. (b) The network structure of LSTM cell.

accuracy. We used the Adam optimizer to optimize the accumulated loss of classification and imputation, a learning rate of 0.001, batch size of 512, and dropout rate of 0.3.

$$\ell^c = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K P_{ij}^{true} \log(p_{ij}^{pred}) \quad (11)$$

$$\ell^{sum} = \lambda_1 \ell^i + \lambda_2 \ell^c \quad (12)$$

where N is the number of samples, K is the number of categories, i is the sample index, j is the category index, P_{ij}^{true} and p_{ij}^{pred} denote the true label and predicted label probability. λ_1 and λ_2 represent the weight of imputation and classification loss, and are set to 0.4 and 0.6.

2.4. The Im-BiLSTM model interpretation

2.4.1. Input feature importance evaluation

In order to understand the Im-BiLSTM model, we calculated the derivative value of the input feature based on the gradient back-propagation to represent the input feature importance (Rübwurm and Körner, 2020; Xu et al., 2021). The Im-BiLSTM model is a nonlinear differentiable function, and can map the input tensor \mathbf{X} to the

corresponding label y . During the training process, starting from the maximum prediction score $\hat{y} = \max(p_i)$, we further backpropagate its gradients through the entire network to each feature of the input tensor \mathbf{X} . Therefore, the gradient $\frac{\partial \hat{y}}{\partial x}$ indicates the contribution of the input feature to the final classification. The positive gradient indicates the predicted category score is positively correlated with the input features, while the negative gradient is the opposite. Zero gradient means that the input feature has no effect on the prediction.

2.4.2. Hidden feature analysis

We analyze how information flows in the recurrent network by considering the expected response of hidden units to each time-step input (Ming et al., 2017). Ming et al. (2017) proved that the numerator of the Softmax function (Eq. (10)) could be decomposed into the product of multiple factors using the method proposed by Murdoch and Szlam (2017).

$$\exp(U_i h_T) = \exp\left(\sum_{t=1}^T U_i(h_t - h_{t-1})\right) = \prod_{t=1}^T \exp(U_i \Delta h_t) \quad (13)$$

where $\exp(U_i \Delta h_t)$ represents the multiplicative contribution of the input

at time-step t to the predicted probability of category i , and $\Delta h_t = h_t - h_{t-1}$ can be expressed as the model's response to the input at time-step t .

Although h_t is calculated by non-linear transformation of h_{t-1} and x_t , Δh_t can be determined by x_t at the current time-step when previous time-step hidden state h_{t-1} is given. Therefore, Δh_t can reflect the degree to which hidden units are affected by the input x_t . We use the expected response calculated from all samples as a stable measure:

$$S(x) = \frac{\sum_{n=1}^N E(\Delta h_t | x_t^n)}{N} \quad (14)$$

where N denotes the number of samples. The expected response $S(x)$ can be positive or negative, and a larger absolute value indicates that the hidden state unit is more sensitive to the input x_t .

The t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton, 2008) is applied to qualitatively visualize the separability between different crops. The t-SNE is a commonly used dimensionality reduction technique for high-dimensional data, which can project high-dimensional features into a 2-dimensional space, making similar samples closer and heterogeneous samples farther. The Im-BiLSTM model can learn high-level hidden features from input data. The t-SNE separability analysis of raw input features and hidden features provides an overall structural cognition of the model learned representation.

2.5. Evaluation

We calculated the confusion matrix, overall accuracy and F1-score to evaluate the classification performance. The overall accuracy (OA) is the ratio of the number of correctly classified samples to the number of the total samples. The F1-score is the harmonic mean of producer's accuracy and user's accuracy (Zhong et al., 2019). In addition, the coefficient of determination (R^2) and the root mean square error (RMSE) are used to evaluate the imputation performance of the Im-BiLSTM model.

3. Results

3.1. Spectral reflectance time-series profile analysis of crops

The spectral reflectance time-series profiles aggregated from specific crop samples are shown in Fig. 6, illustrating their potential to identify different crops through separable time windows. According to similar phenological patterns, five crops can be divided into two clusters for analysis: 1) cotton, spring corn and tomato, 2) winter wheat and spring wheat. As shown in Fig. 6(a), for the visible bands, Red-edge1 band and SWIR bands, there was a large overlap between cotton, spring corn, and tomato from April 22 to August 10, while showed more difference from August 30 to September 6. On August 30 and September 6, tomatoes were harvested, and the bare soil information reflected in the tomato plots was markedly different from that of cotton and spring corn. On September 6, the buffer of spring corn had little overlap with that of other two crops, especially the Red and SWIR2 bands. From Red-edge2 to NIR2 bands, the cotton buffer was significantly different from spring corn and tomato during July 21 to September 6.

As shown in Fig. 6(b), for the visible bands, Red-edge1 band and SWIR bands, there was a large difference between winter wheat and spring wheat from April 22 to May 22. On April 22, winter wheat would be the only green in this period and spring wheat was just emerged. On May 22, spring wheat was at jointing stage and the reflectance difference between the two crops became smaller but still separable. After winter wheat and spring wheat were harvested in late July and early August, the bare soil information reflected by different plots varies greatly, so the buffers of the two crops overlapped greatly. From Red-edge2 to NIR2 bands, winter wheat and spring wheat had a large overlap throughout the time-series.

3.2. The Im-BiLSTM model performance analysis

3.2.1. Classification performance

Based on the 4 missing cases in Table 2, the Im-BiLSTM model imputed the missing data and used the imputed complete time-series data containing 10 time-steps for crop classification, while the BiLSTM model only used the time-series data composed of images without cloud coverage to classify crops. We calculated the confusion matrices (Fig. 7), overall accuracy (OA) and per-class F1-score on the testing set (Tables 3 and 4). The BiLSTM model had an OA of 96.9% in the complete dataset, and 95.4%, 94.9%, 93.2%, and 92.0% in missing cases 1 to 4, respectively. Compared with BiLSTM, the OA of the Im-BiLSTM model was improved in all the 4 missing cases, with a maximum improvement of 4.2%, which occurred in case 4 (MR: 30%). Except for cotton, winter wheat and "Other", the F1-score of the remaining classes in the Im-BiLSTM model was significantly improved, especially spring corn and tomato. In case 3, compared with BiLSTM model, the F1-score of spring corn and tomato obtained by the Im-BiLSTM model increased by 11.2% (MR: 30%), 9.8% (MR: 50%), and 7.1% (MR: 70%) respectively, and 15.2% (MR: 30%), 11.9% (MR: 50%), and 7.4% (MR: 70%) respectively. In case 4, the F1-score of spring corn and tomato obtained by Im-BiLSTM model increased by 16.1% (MR: 30%), 13.9% (MR: 50%), and 11.2% (MR: 70%) respectively, and 21.4% (MR: 30%), 17.3% (MR: 50%), and 11.9% (MR: 70%) respectively. The results showed that the Im-BiLSTM model introduced a regression layer to impute missing data, so the increased temporal information further improved the classification accuracy. Overall, for the 4 missing cases, the Im-BiLSTM model performed better at a low missing rates (30%, 50%) and improved less at a high missing rates (70%). Fig. 8 showed the spatial classification map of crops in Shawan County obtained by Im-BiLSTM model. Without segmentation of the plots, the Im-BiLSTM model recognized not only different plot types but also had clear boundaries between plots.

3.2.2. Imputation performance

In our Im-BiLSTM model, crop classification is the main task, and imputation of missing data is the secondary task. In model joint training, classification guides the imputation process, and the imputation provides high-quality missing values, which helps achieve optimal performance on the classification task. What is the imputation performance of missing data when the classification accuracy is optimal? We used the R^2 to evaluate the imputation accuracy of missing data for each image on the testing set, and density scatter plots to visualize the bands with maximum and minimum imputation accuracy for each image, as shown in Fig. 9. On May 22, the Blue, Green, Red and Red-edge1 bands had the highest imputation accuracy, followed by the SWIR1 and SWIR2 bands. The imputation accuracy was similar for July 21 and August 30, the Red-edge2, Red-edge3, NIR1 and NIR2 bands had the highest imputation accuracy, followed by the Red and SWIR2 bands. On September 06, the Red-edge2 to the NIR2 bands had the highest imputation accuracy. Imputation accuracy was lower on June 11 and lowest on September 19. The imputation accuracy of the bands at different time-steps was different, which can be further illustrated by the model interpretability.

3.3. Interpreting the learning behavior of the Im-BiLSTM model

3.3.1. Feature importance evaluation

The feature importance evaluation results calculated by the gradient-based method showed that the Im-BiLSTM model captured important phenological periods and features. Fig. 10(a) showed that the period around August 30 and September 6 was important, as the time-series profiles of cotton, spring corn and tomato (Fig. 6(a)) were separable during this period. At that time, cotton was in boll opening stage, spring corn was in mature stage, and tomato was harvested, thus the temporal profiles associated with the crop growth conditions showed significant differences. For cotton, the model assigned high positive importance values to the Red-edge2, Red-edge3, NIR1, and NIR2 bands on August

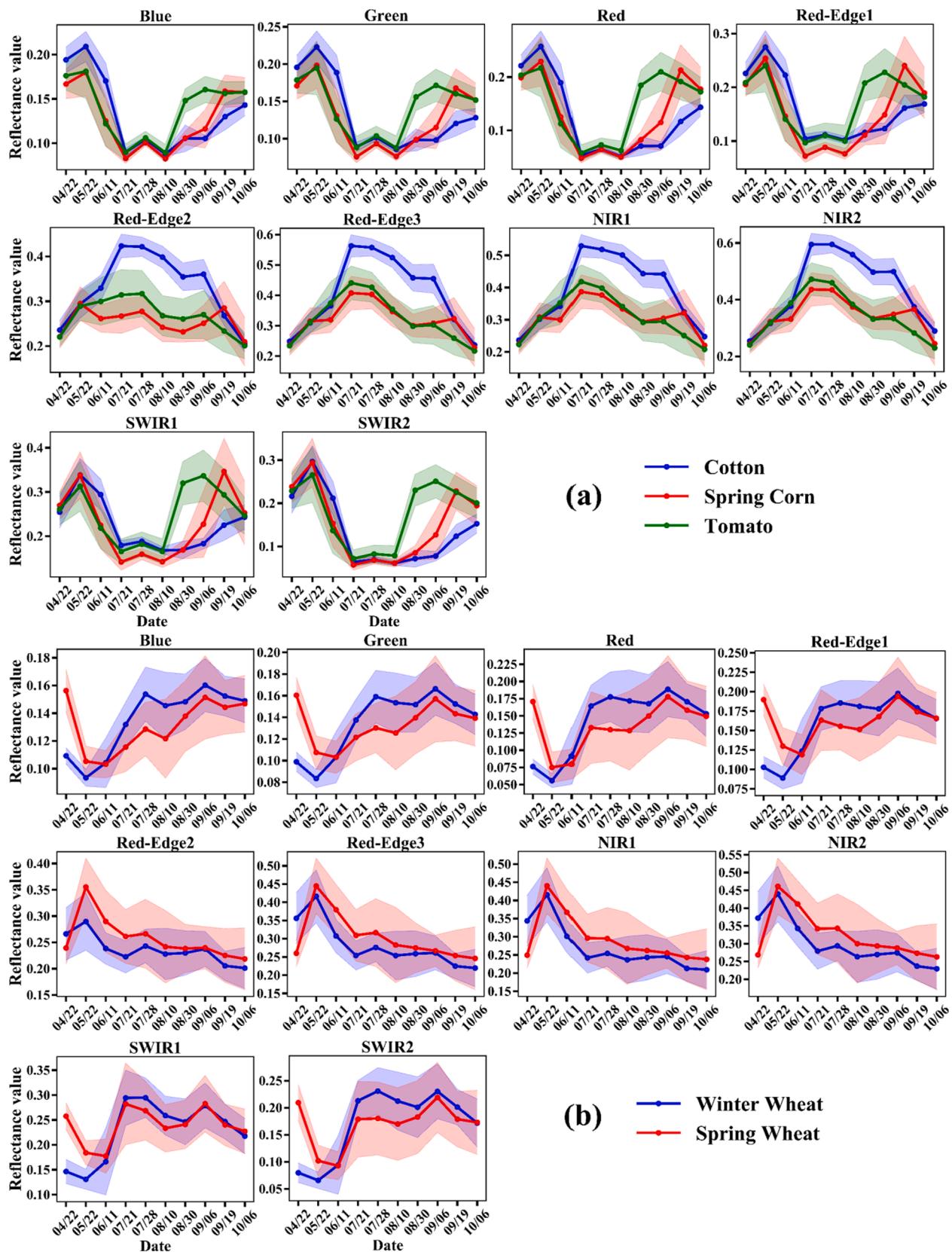


Fig. 6. The spectral reflectance time-series profiles. (a) Cotton, spring corn and tomato, and (b) winter wheat and spring wheat. The X-axis represents the image date, and the Y-axis represents the band reflection value. The solid line represents the average reflectance value, and the buffer refers to a standard deviation from average value.

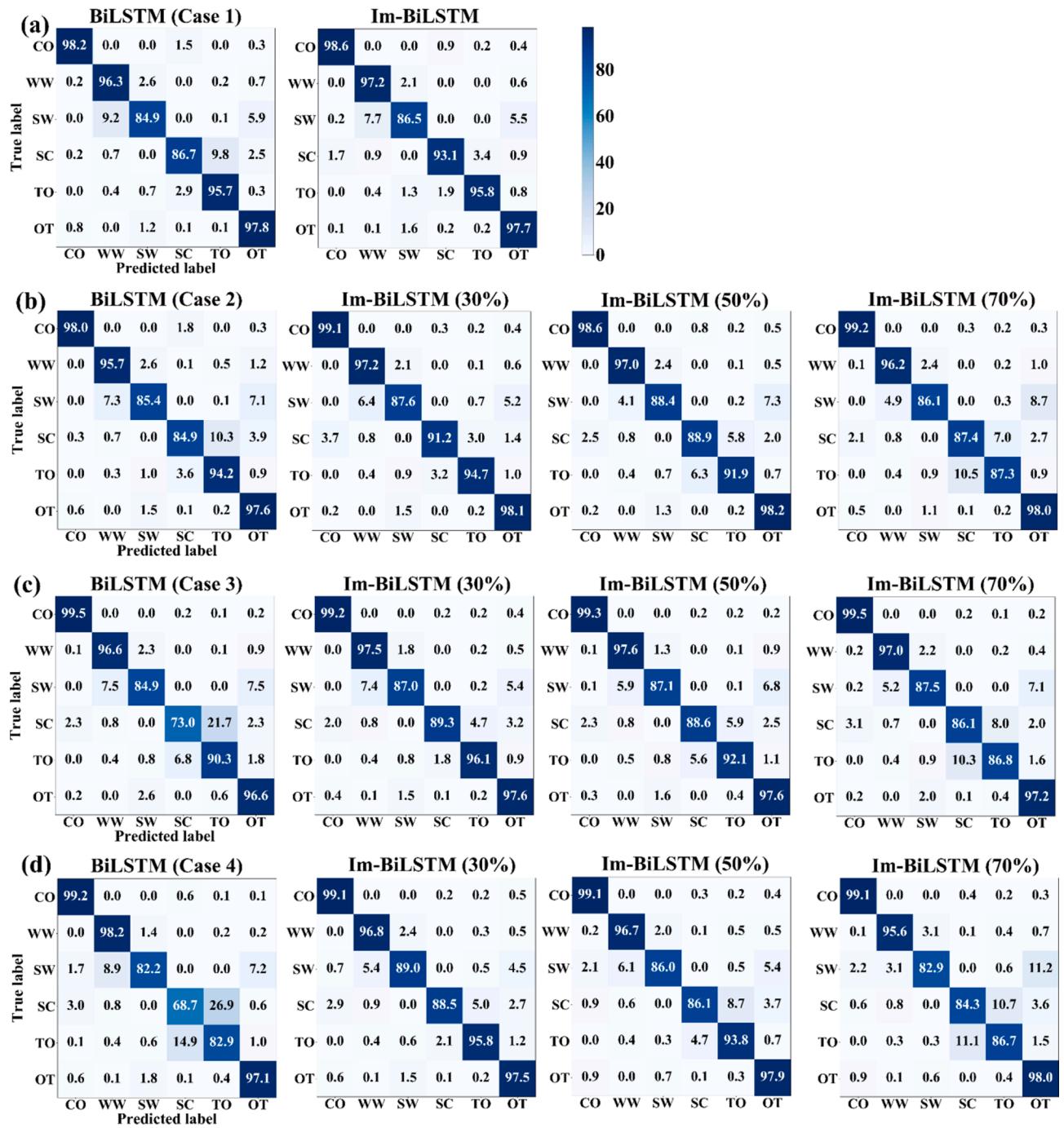


Fig. 7. (a-d) Confuse matrices of the Im-BiLSTM model and BiLSTM model under 4 missing cases, and the values in brackets indicate the missing rate for each image. The diagonal value indicates the class-wise accuracy, namely cotton (CO), winter wheat (WW), spring wheat (SW), spring corn (SC), tomato (TO) and Other (OT).

30, corresponding to the separability of the time-series profile in Fig. 6 (a). However, the model yielded high negative importance in Red and SWIR2 bands for cotton. By analyzing the features of training samples, we found that when the Red and SWIR2 bands have higher reflectance, cotton might be more easily misclassified into spring corn. For spring corn and tomato, the change of feature importance over time was opposite to that of cotton, and the model assigned high positive importance to Red and SWIR2 bands on September 6 and August 30, respectively. This was consistent with the separability window in Fig. 6 (a).

For winter wheat and spring wheat, the phenological difference was obviously different from the above three crops. Fig. 10(b) showed that April 22 is an important period for identifying winter wheat and spring

wheat. At that time, winter wheat was at jointing stage and spring wheat was at trileafing stage, thus the temporal profiles were significantly different (Fig. 6(b)). The feature importance changes of winter wheat and spring wheat were opposite, and the model assigned high importance values in Red and SWIR2 bands and identified these 2 bands as important bands of winter wheat (negative value) and spring wheat (positive value) classification. When the reflectance values of the Red and SWIR2 bands on April 22 were higher, winter wheat may be misclassified as spring wheat. On August 10 and after, the gradient values of winter wheat and spring wheat were close to 0, indicating that the subsequent temporal features had little effect on identifying these 2 crops. This was because winter wheat and spring wheat have been harvested during this period, and the model received bare soil

Table 3

F1-score value for each class and overall accuracy (OA) in the complete time-series dataset and ground-truth missing case 1.

Class	F1-Score (%)		
	Complete		Case 1
	BiLSTM	Im-BiLSTM	
Cotton	98.6	98.9	98.8
Winter Wheat	96.4	96.4	95.8
Spring Wheat	85.9	84.6	84.5
Spring Corn	95.4	94.7	90.3
Tomato	97.2	94.1	89.3
Others	97.3	97.6	97.3
OA (%)	96.9	96.7	95.4

Table 4

F1-score value for each class and overall accuracy (OA) in the missing case 2 to 4.

Class	F1-Score (%)											
	Case 2			Case 3			Case 4					
	Im-BiLSTM			BiLSTM			Im-BiLSTM			BiLSTM		
	30%	50%	70%	0	30%	50%	70%	0	30%	50%	70%	0
Cotton	98.7	98.7	99.0	98.7	99.0	99.1	99.0	99.2	98.7	98.9	99.0	98.7
Winter Wheat	96.8	97.1	96.6	95.9	96.6	97.0	96.9	96.3	96.6	96.7	96.5	96.8
Spring Wheat	85.7	86.4	85.5	83.7	85.8	86.1	84.2	81.3	86.5	87.4	84.7	82.9
Spring Corn	94.2	91.5	90.0	88.7	93.5	92.1	89.4	82.3	93.0	90.8	88.1	76.9
Tomato	93.8	90.1	86.4	87.9	93.1	89.8	85.3	77.9	92.4	88.3	82.9	71.0
Others	97.7	97.4	97.0	96.6	97.0	97.1	97.0	96.4	97.0	97.0	96.5	97.5
OA (%)	96.6	95.9	95.3	94.9	96.4	96.0	95.2	93.2	96.2	95.7	94.6	92.0

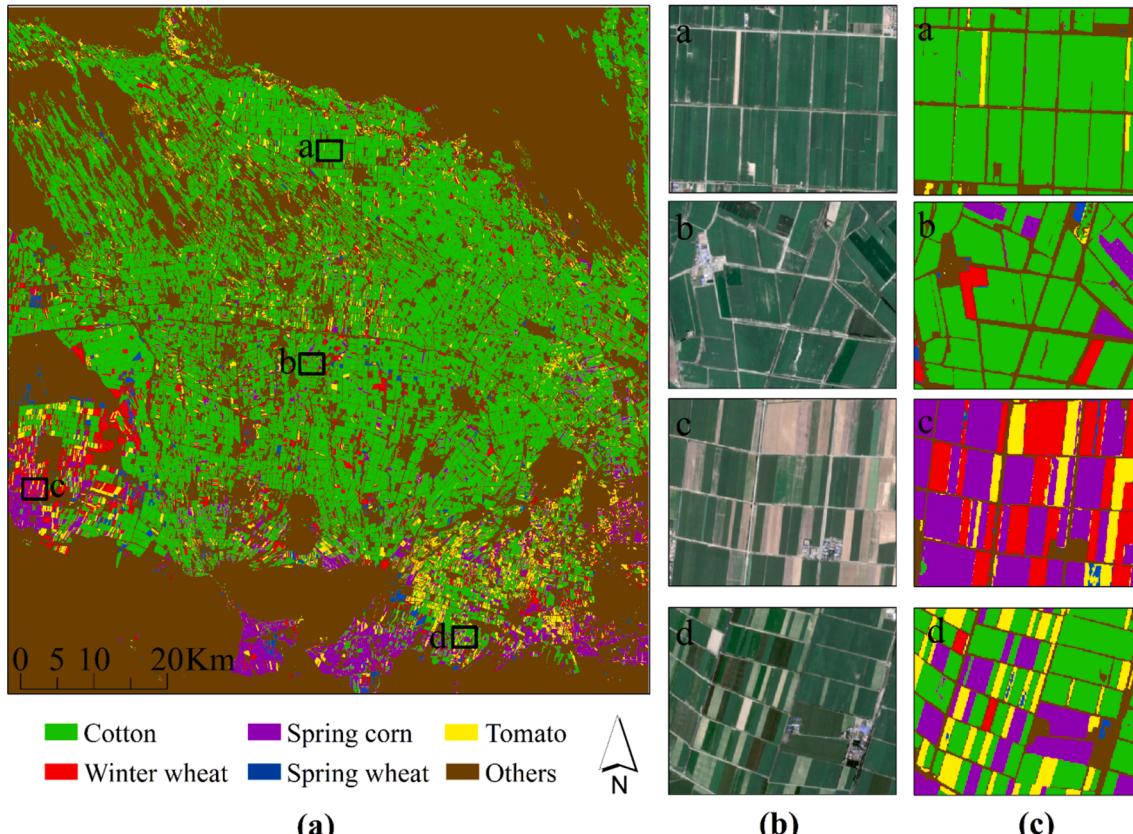


Fig. 8. (a) Crop spatial classification map obtained by the Im-BiLSTM model, (b) Zoom-in of the optical image on August 10, and (c) Zoom-in of the classification map.

information that has no effect on classification.

3.3.2. Hidden state units analysis

The LSTM cell at each time-step has 64 hidden units. We calculated the expected response of each hidden unit to each time-step input to analyze the information flow in the recurrent network. The larger the absolute value of the expected response, the more sensitive the model is to this time-step input, which was displayed in the heatmap (Fig. 11(a) and Fig. 12(a)). We also calculated the absolute average overall expected response of hidden units at each time-step to quantitatively visualize the heatmap, as shown in Fig. 11(b) and Fig. 12(b). The Im-BiLSTM model selectively aggregated important information from each time-step input and stored it in hidden units. As the time-steps increased, the overall response gradually increased to the maximum and then began to decrease, indicating that the model had learned enough important representations from the previous time-steps to

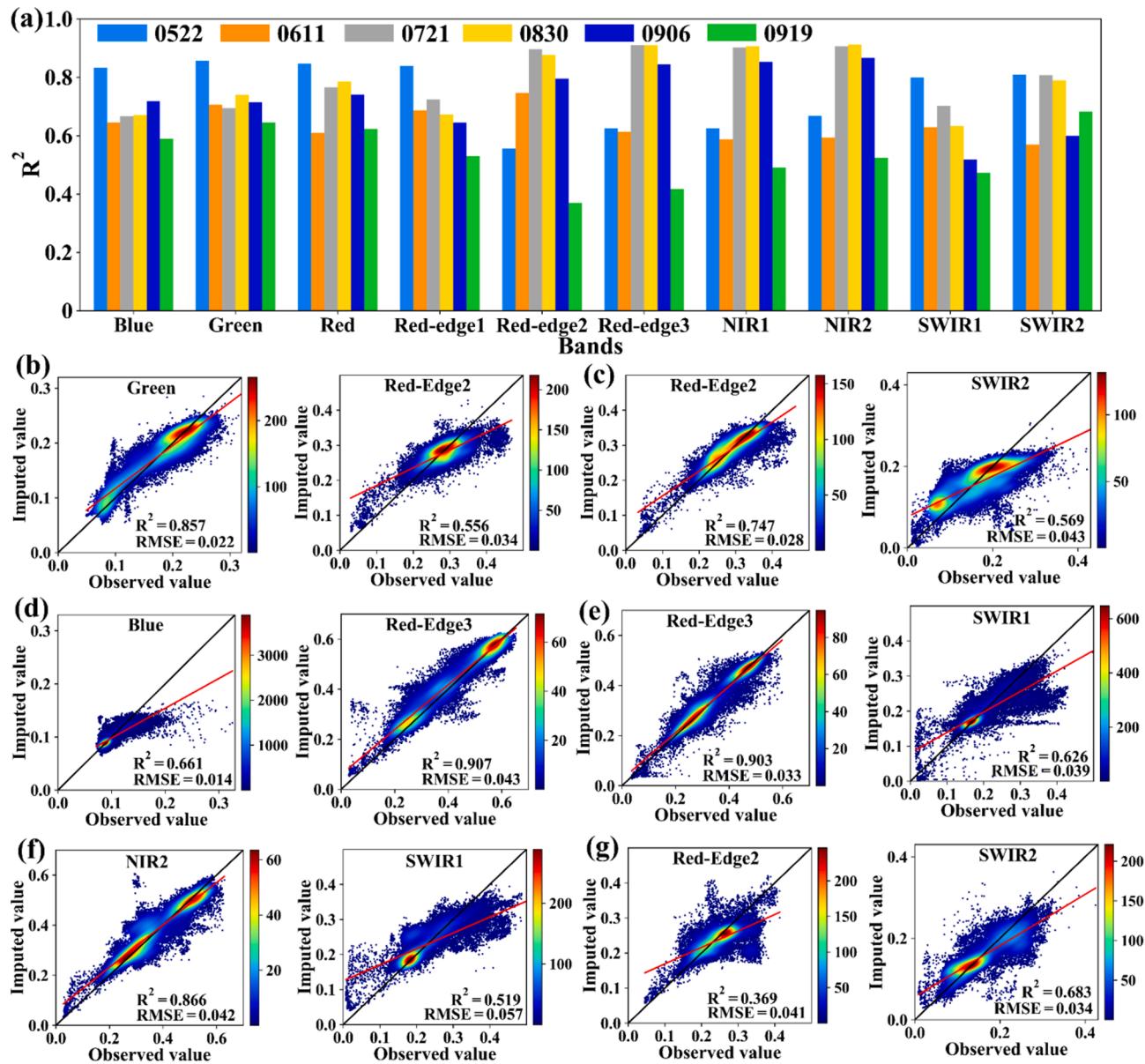


Fig. 9. (a) The bar chart shows the imputation accuracy of the missing data obtained by the Im-BiLSTM model on the 6 missing images, respectively. (b-g) Density scatter plots visualize the bands with maximum and minimum imputation accuracy on each missing image, including May 22, June 11, July 21, August 30, September 6, and September 19.

identify crops, so the later input features had little impact on classification. The response of hidden units to temporal input can be revealed by the corresponding important phenological events of crops, such as the transition of emergence, jointing, maturity, and harvest. As shown in Fig. 11(a), for cotton, spring corn, and tomato, the maximum response period of hidden units was from July 28 to September 6, corresponding to the late growing season. As shown in Fig. 11(b), for these three crops, the overall response was small and increased slowly from April 22 to July 21, and then rapidly increased to the maximum from July 28 to August 30. Cotton was in blooming stage on August 10 and spitting stage on August 30, so the overall response increased rapidly due to the sudden transition of phenological events. For spring corn, the overall response reached its maximum on August 30 and September 6, when the corn was in the mature stage. Tomato was still growing on August 10, but was harvested on August 30, so the information received by Im-BiLSTM abruptly changed from vegetation to bare soil, and the overall response increased significantly.

As shown in Fig. 12(a), for winter wheat and spring wheat, the period

of maximum hidden unit response was from July 28 to August 10. As shown in Fig. 12(b), from April 22 to May 22, winter wheat transitioned from jointing to booting stage, while spring wheat transitioned from three-leaf to jointing stage, so the overall response of winter wheat increased more. Winter wheat and spring wheat were gradually harvested from July 28 to August 10, when the information received by the model changed from vegetation to bare soil, so the response of these two crops reached its maximum during this period. The overall response gradually decreased after August 10, indicating that the model had learned important representations that can identify winter wheat and spring wheat from the accumulated temporal features from April 22 to August 10, while the subsequent time-steps inputs had little effect.

For each of the 6 land types, 5000 samples were randomly selected from the dataset, and the t-SNE algorithm was used to perform dimensionality reduction separability analysis. Fig. 13(a-j) showed the dimensionality reduction results that the raw input features gradually accumulated along 10 time-steps (from 10 to 100 dimensions). The high-level features (128 dimensions) learned by the LSTM cell at the last time-

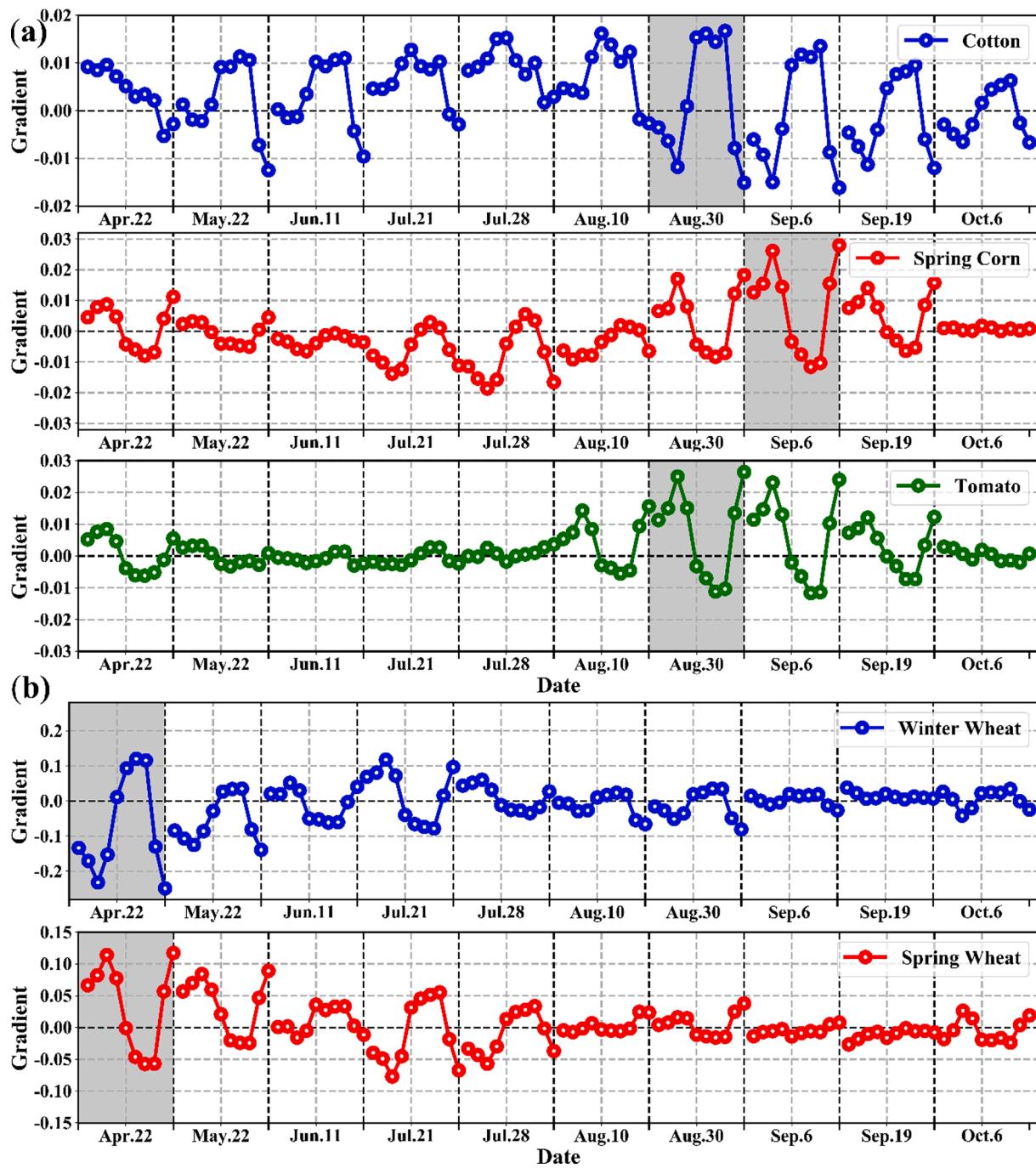


Fig. 10. The distribution of the input features importance based on the average gradient value for each crop. (a) Cotton, spring corn and tomato, and (b) winter wheat and spring wheat. Each date contains 10 features represented by circles, are listed in order of Blue, Green, Red, Red-edge1, Red-edge2, Red-edge3, NIR1, NIR2, SWIR1 and SWIR2 bands. The positive gradient indicates the feature is positively correlated with predicted class score, while negative gradient is opposite. Zero gradient indicates the feature has no effect on the prediction.

step in the forward and backward directions were dimensionally reduced to provide an overview for hidden units of Im-BiLSTM model (Fig. 13(k)). As shown in Fig. 13(a-j), from April 22 to October 6, the separability of the 6 land types became more obvious with the accumulation of temporal features, but it was still challenging to identify different crops using only original features. Fig. 13(k) showed the dimensionality reduction result of hidden features, and samples of the same class were clustered better than raw features. The results showed that the Im-BiLSTM model has the powerful ability to learn important features from time-series input to identify crops.

4. Discussion

4.1. Joint learning performance overview of the Im-BiLSTM model

To overcome the limitations of cloud contamination on multi-temporal crop classification, this study proposed the Im-BiLSTM model, which jointly performs missing data imputation and crop classification. In the 4 missing cases, the Im-BiLSTM model showed better classification performance than the BiLSTM model using only cloud-free images, especially for the less distributed spring corn and tomato crops. When images were continuously lost and the data missing rate ranges

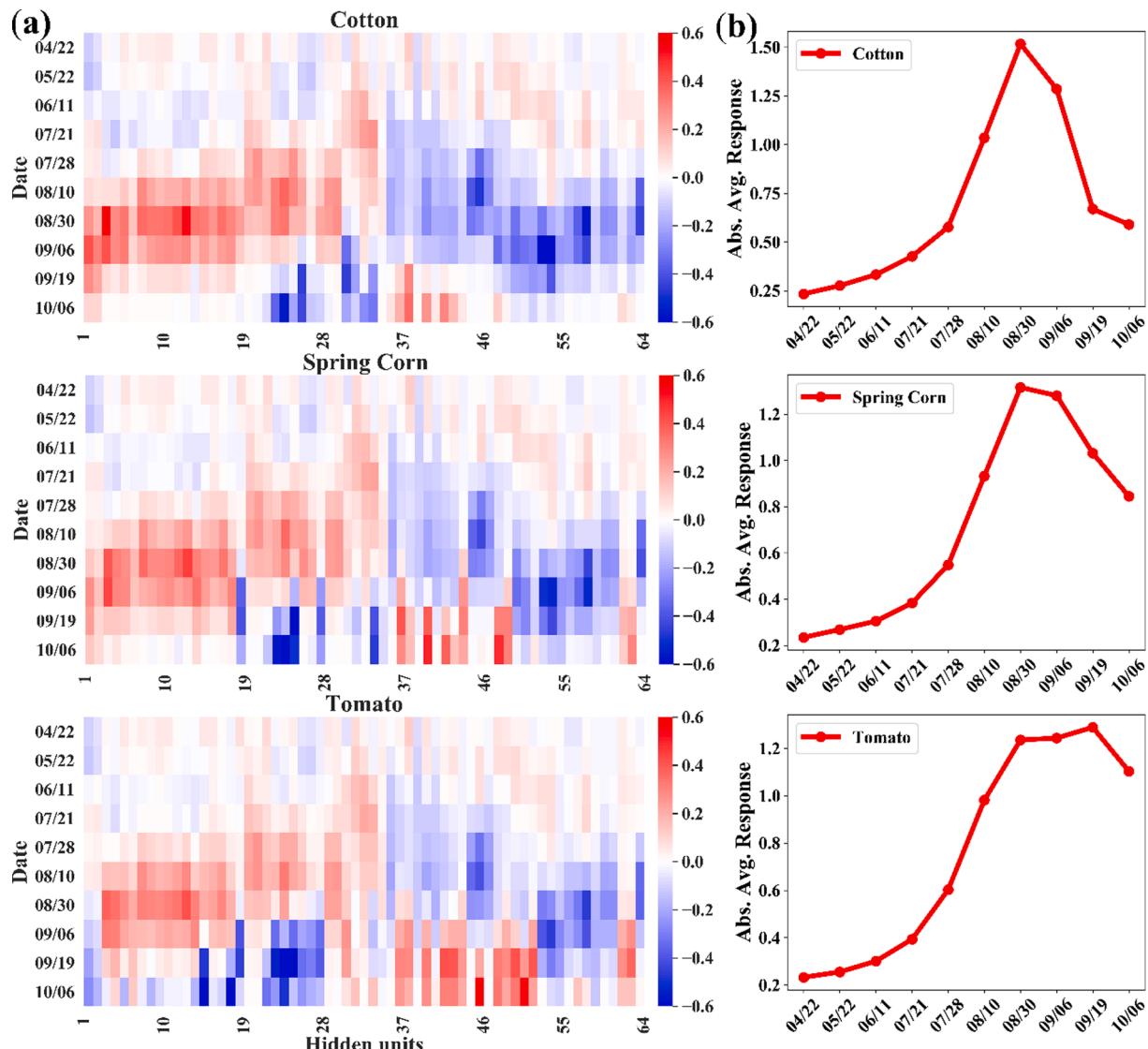


Fig. 11. Cotton, spring corn and tomato. (a) The heatmap represents the expected response of each hidden unit to each time-step input. (b) The curve quantitatively visualizes the response heatmap of hidden unit over time based on the absolute average overall expected response value of all hidden units in each time-step.

30%~50%, the performance of the Im-BiLSTM model was still relatively stable. However, when the missing rate of image data was too high (70%), the improvement of classification performance was not obvious because there was not enough data for Im-BiLSTM model learning. For the imputation performance, the periods and bands with high imputation accuracy were consistent with the key dates and bands obtained from the feature importance evaluation. From the analysis of hidden units, the response of the model to the input data on September 19 was declining, indicating that this time-step input had little effect on the model. In the Im-BiLSTM model, the imputation accuracy depended on crop type and key phenological periods, and the band with the larger contribution to the classification had higher imputation accuracy. The result showed that Im-BiLSTM adaptively learned important features from the time-series data and forgot useless information, which greatly improved the computational efficiency and reduced the noise introduced by useless feature imputation.

In the Im-BiLSTM model, the interaction between imputation and classification tasks significantly reduces the uncertainty and error caused by the previous two-step strategy of imputing-then-classifying. Xu et al. (2021) directly filtered pixels without sufficient temporal information, which was not suitable for contaminated images during critical phenological periods. Cai et al. (2018) used the Savitzky-Golay

algorithm to impute the missing data of Landsat time-series, and found that it had litter impact on the final classification. Gao et al. (2017) fused MODIS-Landsat data to obtain complete time-series data for crop classification, but the consistency between multi-source data should be considered. Certain assumptions and additional computational costs are needed to impute missing data, and the quality of imputed data cannot guarantee the final classification accuracy. Missing data is regarded as variables in the Im-BiLSTM network, which can be directly learned by data-driven without any assumptions. In the Im-BiLSTM model, the missing data imputation directly oriented to the ultimate classification goal, with high computational efficiency and better classification performance.

4.2. The possible improvements of the current study

The Im-BiLSTM model focuses on pixel-based time-series multispectral information and ignores the spatial correlation. We can improve the model structure to take advantage of the spatio-temporal information, such as the combination of convolution and recurrent networks (Qiao et al., 2021). In our experiment, we use 10 bands of the Sentinel-2A image as model input, but not all bands are useful. Reducing feature redundancy can reduce model complexity and improve model

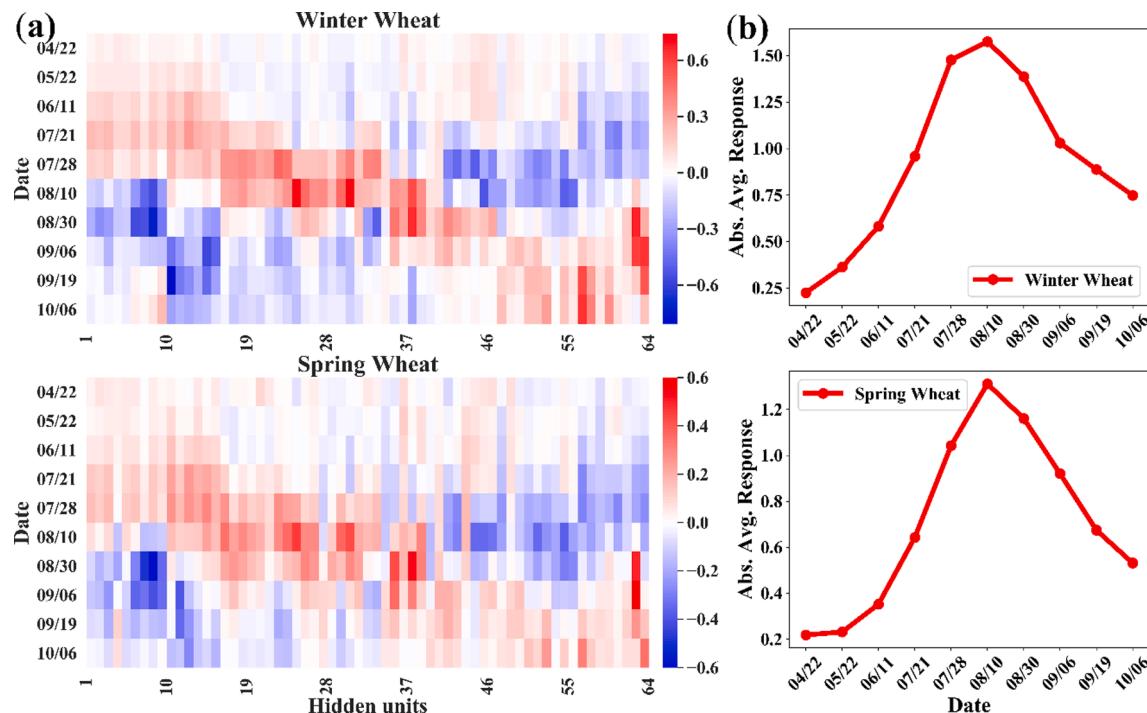


Fig. 12. Winter wheat and spring wheat. (a) The heatmap represents the expected response of each hidden unit to each time-step input. (b) The curve quantitatively visualizes the response heatmap of hidden unit over time based on the absolute average overall expected response value of all hidden units in each time-step.

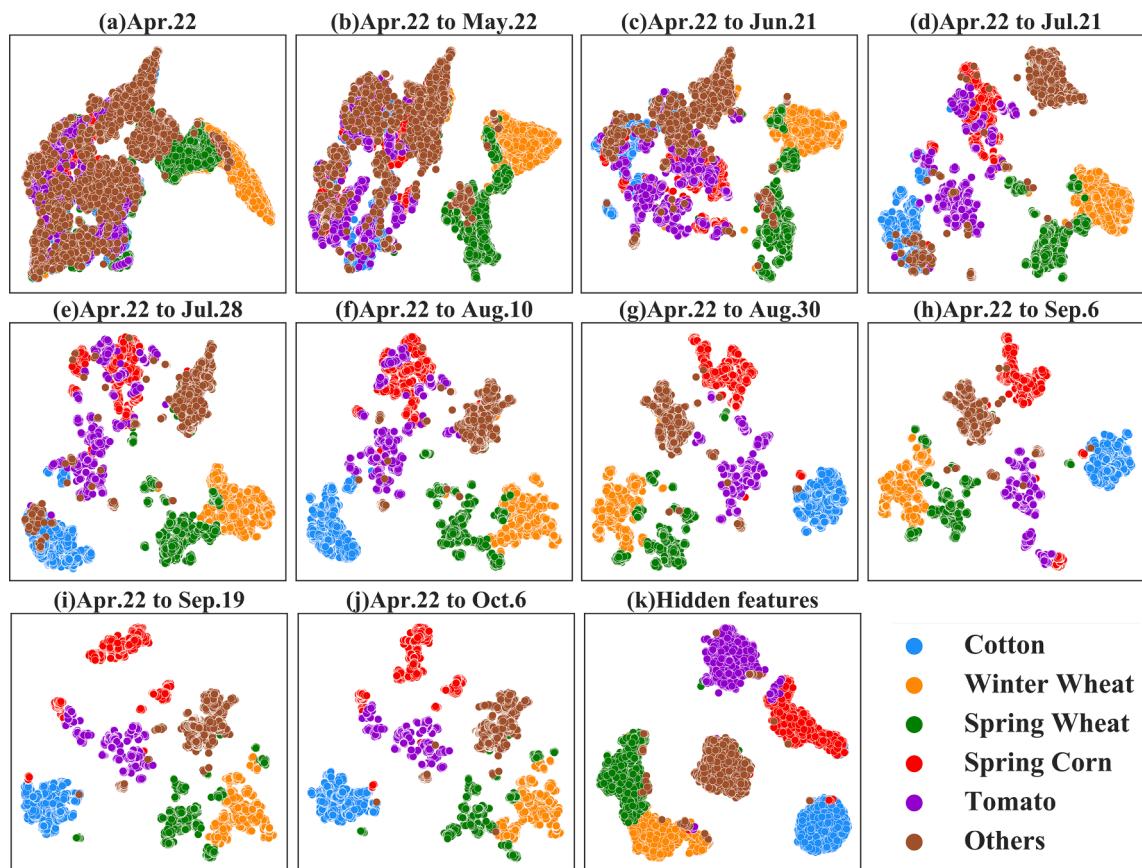


Fig. 13. Visualization of the separability of raw and hidden features based on t-SNE. (a-j) the dimensionality reduction results of the raw features accumulated along 10 time-steps, and (k) the dimensionality reduction result of the hidden features.

interpretability (Hu et al., 2019). In addition to feature importance evaluation and hidden feature visualization, we should improve the Im-BiLSTM model interpretability from more perspectives, so as to better understand the internal complex discriminant patterns and enhance reliability (Pérez-Suay et al., 2020; Xu et al., 2021).

5. Conclusion

This study proposed a novel Im-BiLSTM model for crop classification based on incomplete time-series Sentinel-2A data contaminated by cloud. The Im-BiLSTM model jointly performs missing data imputation and crop classification, which alleviates the error and uncertainty brought from imputation to classification, thereby making classification more accurate. Input feature importance evaluation and hidden units visualization combined with phenological knowledge can better understand the behavior of the Im-BiLSTM model, which significantly improves the deep learning model interpretability. In conclusion, this study deals with time-series data loss in multi-temporal crop classification from a new perspective. In addition to the specific application of this study, we believe that the model has great potential to be applied to other time-series analysis problems of remote sensing, including not only observation gaps caused by cloud cover, but also sensor faults and ground object occlusion.

CRediT authorship contribution statement

Baili Chen: Conceptualization, Data curation, Methodology, Writing – original draft, Writing – review & editing. **Hongwei Zheng:** Conceptualization, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing, Project administration. **Lili Wang:** Data curation. **Olaf Hellwich:** Funding acquisition. **Chunbo Chen:** Investigation. **Liao Yang:** Investigation. **Tie Liu:** Formal analysis. **Geping Luo:** Funding acquisition. **Anming Bao:** Funding acquisition, Project administration. **Xi Chen:** Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Open Project of Key Laboratory of Xinjiang Uygur Autonomous Region under Grant [2018D04027], the National Natural Science Foundation of China under Grant [41877012 & 41671108], the High-End Foreign Experts Project [2020-2023], the team project of the Chinese Academy of Sciences under Grant [2018-YDYLTD-002], the Strategic Priority Research Program of the Chinese Academy of Sciences, and Pan-Third Pole Environment Study for a Green Silk Road (Grant No. XDA20060303).

References

- Bargiel, D., 2017. A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sens. Environ.* 198, 369–383.
- Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sens. Environ.* 210, 35–47.
- Campos-Taberner, M., García-Haro, F.J., Martínez, B., Izquierdo-Verdiguier, E., Atzberger, C., Camps-Valls, G., Gilabert, M.A., 2020. Understanding deep learning in land use classification based on Sentinel-2 time series. *Sci. Rep.* 10, 1–12.
- Cao, W., Wang, D., Li, J., Zhou, H., Li, L., Li, Y., 2018. Brits: Bidirectional recurrent imputation for time series. *arXiv preprint arXiv:1805.10572*.
- Che, Z., Purushotham, S., Cho, K., Sontag, D., Liu, Y., 2018. Recurrent neural networks for multivariate time series with missing values. *Sci. Rep.* 8, 1–12.
- Chen, B., Huang, B.o., Chen, L., Xu, B., 2017. Spatially and Temporally Weighted Regression: A Novel Method to Produce Continuous Cloud-Free Landsat Imagery. *IEEE Trans. Geosci. Remote Sens.* 55 (1), 27–37.
- Cheng, Q., Shen, H., Zhang, L., Yuan, Q., Zeng, C., 2014. Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model. *ISPRS J. Photogramm. Remote Sens.* 92, 54–68.
- Crisóstomo de Castro Filho, H., Abílio de Carvalho Júnior, O., Ferreira de Carvalho, O.L., Pozzobon de Bem, P., dos Santos de Moura, R., Olino de Albuquerque, A., Rosa Silva, C., Guimarães Ferreira, P.H., Fontes Guimarães, R., Trancoso Gomes, R.A., 2020. Rice Crop Detection Using LSTM, Bi-LSTM, and Machine Learning Models from Sentinel-1 Time Series. *Remote Sensing* 12 (16), 2655.
- do Nascimento Bendini, H., Garcia Fonseca, L.M., Schwieder, M., Sehn Körting, T., Rufin, P., Del Arco Sanches, I., Leitão, P.J., Hostert, P., 2019. Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series. *Int. J. Appl. Earth Obs. Geoinf.* 82, 101872.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* 120, 25–36.
- Estel, S., Kuemmerle, T., Alcántara, C., Levers, C., Prischepov, A., Hostert, P., 2015. Mapping farmland abandonment and recultivation across Europe using MODIS NDVI time series. *Remote Sens. Environ.* 163, 312–325.
- Foerster, S., Kaden, K., Foerster, M., Itzler, S., 2012. Crop type mapping using spectral-temporal profiles and phenological information. *Comput. Electron. Agric.* 89, 30–40.
- Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., Prueger, J.H., 2017. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sens. Environ.* 188, 9–25.
- Gao, J., Yuan, Q., Li, J., Su, X., 2021. Unsupervised missing information reconstruction for single remote sensing image with Deep Code Regression. *Int. J. Appl. Earth Obs. Geoinf.* 105, 102599.
- García-Laencina, P.J., Sancho-Gómez, J.-L., Figueiras-Vidal, A.R., 2013. Classifying patterns with missing values using multi-task learning perceptrons. *Expert Syst. Appl.* 40 (4), 1333–1341.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Hu, Q., Sulla-Menashe, D., Xu, B., Yin, H., Tang, H., Yang, P., Wu, W., 2019. A phenology-based spectral and temporal feature selection method for crop mapping from satellite time series. *Int. J. Appl. Earth Obs. Geoinf.* 80, 218–229.
- Johnson, J.A., Runge, C.F., Senauer, B., Foley, J., Polasky, S., 2014. Global agriculture and carbon trade-offs. *Proc. Natl. Acad. Sci.* 111 (34), 12342–12347.
- Julien, Y., Sobrino, J.A., 2010. Comparison of cloud-reconstruction methods for time series of composite NDVI data. *Remote Sens. Environ.* 114 (3), 618–625.
- Karthikeyan, L., Chawla, I., Mishra, A.K., 2020. A review of remote sensing applications in agriculture for food security: crop growth and yield, irrigation, and crop losses. *J. Hydrol.* 586, 124905.
- Kim, J., Kim, T., Choi, J.-H., Choo, J., 2021. End-to-end multi-task learning of missing value imputation and forecasting in time-series data, 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, pp. 8849–8856.
- Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* 14 (5), 778–782.
- Lei, L., Wang, X., Zhong, Y., Zhao, H., Hu, X., Luo, C., 2021. DOCC: Deep one-class crop classification via positive and unlabeled learning for multi-modal satellite imagery. *Int. J. Appl. Earth Obs. Geoinf.* 105, 102598.
- Li, H., Wu, Y., Lu, H., Chen, S., 2021. Posterior Probability-Based Single Time-Series Model Combination Strategy for Predicting Multi-Temporal Crop Classification, 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). IEEE, pp. 26–31.
- Ma, Q., Li, S., Cottrell, G., 2020. Adversarial Joint-Learning Recurrent Neural Network for Incomplete Time Series Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Ming, Y., Cao, S., Zhang, R., Li, Z., Chen, Y., Song, Y., Qu, H., 2017. Understanding hidden memories of recurrent neural networks, 2017 IEEE Conference on Visual Analytics Science and Technology (VAST). IEEE, pp. 13–24.
- Moody, D.I., Brumby, S.P., Chartrand, R., Keisler, R., Longbotham, N., Mertes, C., Skillman, S.W., Warren, M.S., 2017. Crop classification using temporal stacks of multispectral satellite imagery. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIII. International Society for Optics and Photonics, p. 101980G.
- Moreno-Martínez, Á., Izquierdo-Verdiguier, E., Maneta, M.P., Camps-Valls, G., Robinson, N., Muñoz-Marí, J., Sedano, F., Clinton, N., Running, S.W., 2020. Multispectral high resolution sensor fusion for smoothing and gap-filling in the cloud. *Remote Sens. Environ.* 247, 111901.
- Mou, L., Bruzzone, L., Zhu, X.X., 2019. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* 57 (2), 924–935.
- Murdock, W.J., Szlam, A., 2017. Automatic rule extraction from long short term memory networks. *arXiv preprint arXiv:1702.02540*.
- Pelletier, C., Webb, G., Petitjean, F., 2019. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sensing* 11 (5), 523.
- Pérez-Suay, A., Adsuar, J.E., Piles, M., Martínez-Ferrer, L., Díaz, E., Moreno-Martínez, A., Camps-Valls, G., 2020. Interpretability of Recurrent Neural Networks in Remote Sensing, IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE 3991–3994.

- Piedelobo, L., Hernández-López, D., Ballesteros, R., Chakhar, A., Del Pozo, S., González-Aguilera, D., Moreno, M.A., 2019. Scalable pixel-based crop classification combining Sentinel-2 and Landsat-8 data time series: Case study of the Duero river basin. *Agric. Syst.* 171, 36–50.
- Pruthi, D., Gupta, M., Dhingra, B., Neubig, G., Lipton, Z.C., 2019. Learning to deceive with attention-based explanations. arXiv preprint arXiv:1909.07913.
- Qiao, M., He, X., Cheng, X., Li, P., Luo, H., Zhang, L., Tian, Z., 2021. Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks. *Int. J. Appl. Earth Obs. Geoinf.* 102, 102436.
- Roy, D.P., Yan, L., 2020. Robust Landsat-based crop time series modelling. *Remote Sens. Environ.* 238, 110810.
- Rußwurm, M., Körner, M., 2017. Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-spectral Satellite Images, 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1496–1504.
- Rußwurm, M., Körner, M., 2020. Self-attention for raw optical satellite time series classification. *ISPRS J. Photogramm. Remote Sens.* 169, 421–435.
- Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., Zhang, L., 2015. Missing information reconstruction of remote sensing data: A technical review. *IEEE Geosci. Remote Sens. Mag.* 3 (3), 61–85.
- Song, X.-P., Potapov, P.V., Krylov, A., King, L., Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, S.V., Hansen, M.C., 2017. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. *Remote Sens. Environ.* 190, 383–395.
- Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9.
- Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F., Ceschia, E., 2017. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens. Environ.* 199, 415–426.
- Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C., Ng, W.-T., 2018. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* 72, 122–130.
- Waldner, F., Fritz, S., Di Gregorio, A., Defourny, P., 2015. Mapping Priorities to Focus Cropland Mapping Activities: Fitness Assessment of Existing Global, Regional and National Cropland Maps. *Remote Sensing* 7.
- Xu, J., Yang, J., Xiong, X., Li, H., Huang, J., Ting, K.C., Ying, Y., Lin, T., 2021. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sens. Environ.* 264, 112599.
- Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., Huang, J., Li, H., Lin, T., 2020. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sens. Environ.* 247, 111946.
- Yan, S., Yao, X., Zhu, D., Liu, D., Zhang, L., Yu, G., Gao, B., Yang, J., Yun, W., 2021. Large-scale crop mapping from multi-source optical satellite imageries using machine learning with discrete grids. *Int. J. Appl. Earth Obs. Geoinf.* 103, 102485.
- Zhao, W., Qu, Y., Chen, J., Yuan, Z., 2020. Deeply synergistic optical and SAR time series for crop dynamic monitoring. *Remote Sens. Environ.* 247, 111952.
- Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* 221, 430–443.
- Zhou, Y., Luo, J., Feng, L.i., Yang, Y., Chen, Y., Wu, W., 2019. Long-short-term-memory-based crop classification using high-resolution optical images and multi-temporal SAR data. *GIScience & Remote Sensing* 56 (8), 1170–1191.
- Zhou, Y., Wang, S., Wu, T., Feng, L.i., Wu, W., Luo, J., Zhang, X., Yan, Na'na, 2022. Forward LSTM-based missing data reconstruction for time-series Landsat images. *GIScience & Remote Sensing* 59 (1), 410–430.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* 5 (4), 8–36.