Remote Sensing of Environment 221 (2019) 430–443

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Deep learning based multi-temporal crop classification[☆]

Liheng Zhong[a,*], Lina Hu[b], Hang Zhou[c]

[a] Department of Water Resources, State of California, Sacramento, CA 95814, United States
[b] Department of Sociology, Tsinghua University, Beijing 100084, China
[c] Descartes Labs, Inc., Santa Fe, NM 87501, United States

ARTICLE INFO

Keywords:
Crop classification
Deep learning
Artificial neural network
Convolutional neural network
Multi-temporal classification
Landsat

ABSTRACT

This study aims to develop a deep learning based classification framework for remotely sensed time series. The experiment was carried out in Yolo County, California, which has a very diverse irrigated agricultural system dominated by economic crops. For the challenging task of classifying summer crops using Landsat Enhanced Vegetation Index (EVI) time series, two types of deep learning models were designed: one is based on Long Short-Term Memory (LSTM), and the other is based on one-dimensional convolutional (Conv1D) layers. Three widely-used classifiers were also tested for comparison, including a gradient boosting machine called XGBoost, Random Forest, and Support Vector Machine. Although LSTM is widely used for sequential data representation, in this study its accuracy (82.41%) and F1 score (0.67) were the lowest among all the classifiers. Among non-deep-learning classifiers, XGBoost achieved the best result with 84.17% accuracy and an F1 score of 0.69. The highest accuracy (85.54%) and F1 score (0.73) were achieved by the Conv1D-based model, which mainly consists of a stack of Conv1D layers and an inception module. The behavior of the Conv1D-based model was inspected by visualizing the activation on different layers. The model employs EVI time series by examining shapes at various scales in a hierarchical manner. Lower Conv1D layers of the optimized model capture small scale temporal variations, while upper layers focus on overall seasonal patterns. Conv1D layers were used as an embedded multi-level feature extractor in the classification model which automatically extracts features from input time series during training. The automated feature extraction reduces the dependency on manual feature engineering and pre-defined equations of crop growing cycles. This study shows that the Conv1D-based deep learning framework provides an effective and efficient method of time series representation in multi-temporal classification tasks.

1. Introduction

Seasonality is one of the most prominent characteristics of vegetation. Multi-temporal remote sensing is an efficient source of time series observations to monitor growing dynamics for vegetation classification (Rogan et al., 2002; Xie et al., 2008). As remotely-sensed time series are being generated at an unprecedented scale and rate from an expanding collection of platforms, a key aspect towards the goal of classification is how to use time series to fully utilize the wealth of seasonal patterns and sequential relationships. As a consequence, there has been an increasing amount of interest in time series representation, which extracts features from time series to retrieve useful information on vegetation growing stages and conditions.

1.1. *Traditional approaches for temporal feature representation*

In multi-temporal remote sensing, various methods have been developed to process and analyze vegetation index (VI) time series. Some studies directly used original VI values during different periods of the year as the main input to rule-based classification algorithms like decision trees (Lloyd, 1990; Friedl et al., 1999; Wardlow and Egbert, 2008). The direct use of time series is simple but effective for vegetation types with distinctive temporal characteristics like rice (Xiao et al., 2005, Xiao et al., 2006, Dong et al., 2015, G. Zhang et al., 2015). In these methods, the sequential relationship of multi-temporal observations was not explicitly considered, which might ignore some useful information in time series inputs.

[☆] Disclaimer: The views expressed in this abstract are those of the author, and not of the State of California.
[*] Corresponding author at: Room 233, 1416 9th Street, Sacramento, CA 95814, United States.
E-mail address: lihengzhong@berkeley.edu (L. Zhong).

https://doi.org/10.1016/j.rse.2018.11.032
Received 8 March 2018; Received in revised form 2 November 2018; Accepted 22 November 2018
0034-4257/ Published by Elsevier Inc.

A more common way of processing multi-temporal VI data is to extract temporal features or phenological metrics from the time series. Simple statistics- or threshold-based methods employed global statistics and selected threshold values to calculate metrics such as the maximum VI, time of peak VI, and onset of green-up to characterize VI time series (Reed et al., 1994; Vina et al., 2004; Brown et al., 2010; Walker et al., 2014; Walker et al., 2015), which may improve classification accuracy compared to using original VI values (Simonneaux and Francois, 2003; Simonneaux et al., 2008). More complex temporal feature extractors like pre-defined mathematical equations or models have been widely used for multi-temporal classification and vegetation phenology studies. In these studies, VI time series were represented by a function or a set of functions, for example, Fourier transform (Olsson and Eklundh, 1994, Verhoef et al., 1996, Evans and Geerken, 2006, M. Zhang et al., 2008, Geerken, 2009), wavelet transform (Sakamoto et al., 2005; Sakamoto et al., 2006; Galford et al., 2008), Savitzky-Golay filter (J. Chen et al., 2004, Shao et al., 2016), linear regression (Funk and Budde, 2009), spline fitting (Bradley et al., 2007; Bradley and Mustard, 2008), Hidden Markov Model (Siachalou et al., 2015), Kalman Filter (Vicente-Guijalba et al., 2014), manually-defined shapes (Sakamoto et al., 2010; Sakamoto et al., 2011; Sakamoto et al., 2013; Sakamoto et al., 2014), or a combination of function with different forms (Wang et al., 2012). Among the curve-fitting functions, the logistic/sigmoid function proposed by Badhwar (1984) has gained popularity for its robustness and convenience to derive phenological features in characterizing vegetation dynamics and growing cycles (X. Zhang et al., 2003, Fisher et al., 2006, Beck et al., 2006, Fisher and Mustard, 2007, Dannenberg et al., 2015, Xin et al., 2015, Gonsamo and Chen, 2016).

While existing approaches of temporal feature extraction provide many choices to represent vegetation dynamics, in practice it is not an easy task to find an effective and suitable approach. Some problems are:

- The manual work of model design and feature extraction relies on human experience and domain knowledge. Features from general models may not be adequate for specific tasks compared to features directly learned from data and driven by tasks in an end-to-end fashion.
- Manual feature engineering is also tedious and time-consuming. Human intervention is usually required to deal with changing environmental and weather conditions. It is hard for human knowledge to simultaneously account for the complex factors such as interclass similarity, intra-class variability, atmospheric conditions, and light-scattering mechanisms (Zhu et al., 2017).
- The fixed form of pre-defined models and associated mathematical assumptions limit the flexibility to handle disparate temporal patterns. For approaches based on curve-fitting or empirical seasonal patterns, it is difficult to choose a curve function for all the vegetation types particularly in a diverse landscape (Zhong et al., 2011).

In short, a proper way of temporal feature representation is desired to fully utilize the sequential information in time series, but most existing feature extractors come with limitations in automation and flexibility. An ideal temporal feature extractor could be trainable to automatically adapt for specific tasks while bringing minimal mathematical constraints to handle complex and varying patterns, just as the way experienced human experts learn to interpret and identify vegetation types based on the shape of seasonal VI profiles. The human learning process to perceive and recognize temporal patterns appears quite different from aforementioned traditional methods. Human experts do not simply check values in the series by time steps like some decision-tree-based methods. They neither limit their understanding to a fixed mathematical model nor a group of models. One interesting fact is that even for experts it is almost impossible to clearly list all the rules of pattern identification, making it difficult to develop classification algorithms at the human knowledge level (Clancey, 1983; Ripley, 2007). The active field of deep learning often provides solutions to such tasks and shows great potential in feature representation of remotely-sensed time series (Zhu et al., 2017).

1.2. Remote sensing studies using deep learning models

1.2.1. Deep learning architectures

Inspired by the learning processing of human beings, artificial neural networks (ANNs) employ a general structure of connected units to learn feature representation exclusively from data and reduce task-specific and explicit-rule-based programming. Deep learning models, or deep ANNs with more than two hidden layers, provide sufficient model complexity to learn feature representations from data in an end-to-end regime instead of manual feature engineering based on human experience and prior knowledge (LeCun and Bengio, 1995). In recent years, deep learning was considered as a breakthrough technology in machine learning and data mining including the research field of remote sensing. Image classification studies particularly benefit from deep learning due to its flexibility in feature representation, automation by expert-free end-to-end learning, and computational efficiency. With deep learning models features are automatically extracted for given classification tasks without pre-defined feature crafting algorithms by building a part of the deep ANNs as autoencoders (Y. Chen et al., 2014, W. Li et al., 2016, Wan et al., 2017, Mou and Zhu, 2018, Mou et al., 2018a), or feature selectors (Zou et al., 2015).

Convolutional neural network (CNN) is one of the most successful network architecture in deep learning methods. The learning process of CNNs is computational efficient and insensitive to shift in data like image translation, making CNN a leading model to recognize 2D patterns in images (Krizhevsky et al., 2012). In remote sensing studies, 2D CNNs have been widely used to extract spatial features from the dimensions of width and height for object detection and semantic segmentation of high resolution images (X. Chen et al., 2014, Penatti et al., 2015, F. Hu et al., 2015, Kampffmeyer et al., 2016, Sherrah, 2016, Audebert et al., 2018, Maggiori et al., 2017, W. Li et al., 2017, Volpi and Tuia, 2017, Marcos et al., 2018, Marmanis et al., 2018). Another major application of CNNs is hyperspectral image classification, in which CNNs were used to extract spatial-spectral features, through either 1D convolution across the spectral dimension (W. Hu et al., 2015), 2D across the spatial dimensions (Yue et al., 2015; Zhao and Du, 2016; Mou et al., 2018a), or 3D across the spectral and the spatial dimensions simultaneously (Y. Li et al., 2017). Kussul et al. (2017) found that 2D convolution in the spatial domain achieved slightly higher accuracy in crop classification than 1D convolution in the spectral domain. Guidici and Clark (2017) concatenated hyperspectral images from three seasons and applied 1D convolution to the spectral domain for land cover classification. In these studies, convolutional layers in CNNs play a role as feature extractors mostly for the spatial or the spectral domains, but rarely for the temporal domain of remotely sensed time series.

Recurrent neural networks (RNN) are another category of ANNs that extend the conventional networks with loops in connections (Connor et al., 1994; Zaremba et al., 2014). RNNs are specialized for sequential data analysis and have recently shown to be successful in several remote sensing applications. Lyu et al. (2018) adopted a RNN to leverage the sequential properties of multispectral data, such as spectral correlations and band-to-band variability across the spectral dimension. Mou and Zhu (2018) used an encoder to produce multi-level convolutional feature maps from shallow to deep, and a RNN as the decoder to recursively collect multi-scale features and aggregate features sequentially into a high-resolution semantic segmentation image. RNNs are capable of representing data in continuous dimensions with sequential dependency, and the most popular use of RNNs is in the temporal domain to extract features from multi-temporal observations.

1.2.2. RNNs for time series representation

Because of their capability of analyzing sequential data, RNNs are often considered as a natural candidate to learn the temporal relationship in image time series and model land cover changing patterns (Mou et al., 2018b). Variants of RNNs are usually used for improved learning efficiency, and the most well-known variant is the long short-term memory (LSTM), a special RNN unit that represents temporal dependency at various time-spans with gated recurrent connections (Hochreiter and Schmidhuber, 1997). Lyu et al. (2016) employed a LSTM model to learn a joint spectral-temporal feature representation from a bi-temporal image pair for change detection. Mou et al. (2018b) extended the work by using 2D convolutional layers as spatial feature extractors to provide inputs to LSTM, and reported that by employing temporal dependency, the proposed network achieved better results than traditional change detection algorithms based on simple image differencing or stacking. Rußwurm and Körner (2017) used LSTM to extract dynamic temporal features from a longer image sequence to classify crop types. In their recent study, Rußwurm and Körner (2018) improved the architecture by adding 2D convolutional layers as spatial feature extractors and connecting recurrent cells in a bidirectional manner to reduce temporal biases towards later images. Jia et al. (2017) used LSTM to learn the pattern of land cover transitions and predict land cover at each time step as a sequential output. These studies also compared the classification results by LSTM-based models with other approaches to show the advantage of using LSTM as a temporal feature extractor. For example, Lyu et al. (2016) found that LSTM achieved ~95% accuracy compared to ~80% by Support Vector Machine (SVM) and ~70% by Decision Tree in multiple experiments. Mou et al. (2018b) also conducted a series of experiments, in which the combined use of CNN and LSTM resulted in ~98% accuracy that was superior to SVM (~95%) and Decision Tree (~85%) Rußwurm and Körner (2017) reported 90.6% accuracy by a multi-temporal LSTM model, which was slightly higher than CNN (89.2%) and much higher than SVM (40.9%).

1.3. The motivation of this study

This study aims to develop deep learning models as the temporal feature representation approach to identify crop growth patterns and classify crop types. Deep learning models were used as an integrated framework that simultaneously extracts features and performs classification following the end-to-end principle to eliminate the need for handcrafted features and pre-defined functions. Classification results of the proposed models were compared with those from some leading machine learning classifiers. The classification problem, identifying summer crop types using VI time series, is a task that human experts are relatively good at and could effectively evaluate classification results with proper visualization and mapping settings. The major contributions of this study include:

- **Exploring the feasibility of using CNNs for temporal feature representation in multi-temporal classification**. When processing time series, a RNN or its variants like LSTM is often deemed as a natural starting point as RNNs were initially proposed to analyze sequential data. In contrast, CNNs were commonly used in spatial and spectral domains but hardly across the temporal dimension in remote sensing studies, although theoretically 1D convolution is effective and computationally efficient to recognize temporal patterns at multiple scales. In this study, the input data are time series of a single vegetation index (EVI), so the information used for classification is almost purely temporal. By comparing convolutional architectures with LSTM-based models and conventional approaches, we propose that CNNs have great potential in temporal feature representation and should be considered as a candidate for multi-temporal image classification tasks.
- **Visualizing activations of deep neural networks at various temporal scales and interpreting how the deep learning model works**. One common criticism of ANNs is that they are relatively "black box". In remote sensing studies the visualization of deep learning models is challenging and rarely done with a few exceptions. Guidici and Clark (2017) analyzed activation maps created by the convolutional layer to interpret what the classifier learned from the spectral dimension, and highlighted local spectral regions excited by the convolutional layer to explore the distinctive spectral nature of the classes. Rußwurm and Körner (2018) visualized LSTM cell activations and showed how information aggregated over the sequence. Mou and Zhu (2018) utilized class activation maps to visualize learned spatial features from local to holistic levels. In our current study, we focused on the visualization methods for temporal patterns with various lengths from ~monthly to seasonal scales and from lower to higher layers. Two state-of-art methods were attempted: one is finding the special pattern in an input series that activates certain filters in certain layers, and the other is generating an input series that maximize the activation on a given filter of a layer. The visualization for the first time sheds light on how each part of the designed CNN recognizes patterns of crop seasonality and connects learned features to crop phenology and farming practices.

The paper is organized as 5 sections. Following the introduction in Section 1, Section 2 describes the methods to develop deep learning architectures for classification, optimize models, and compare various deep learning models with other classifiers, and the results are presented in Section 3. Section 4 then interprets the behavior of the optimized model to understand how it works and discusses the feasibility of using deep learning models as the representation method for time series in classification applications. Finally Section 5 concludes the paper.

2. Method and materials

2.1. Study area

The study area is Yolo County, California, US. Yolo County mostly locates in the Central Valley and is a major area for agricultural production. Mountainous areas on the west side of the county are unsuitable for cultivation and were excluded from further analysis using a Digital Elevation Model. The east side of the county has flat terrain and ideal soil and irrigation conditions (Fig. 1). Precipitation is about 400 mm per year and mostly occurs in winter as a characteristic of Mediterranean climate.

2.2. Data

2.2.1. Land use survey data

California Department of Water Resources (CDWR) conducts county-level land use survey regularly and we used the latest survey for Yolo County in year 2014. Land parcels in the county were visited by staff in regional offices of CDWR. During the visits a variety of land use attributes were recorded including the general crop category, the specific crop type, mixed cropping types, tree crop age, cover crop, irrigation methods, and other special conditions. A label was created for each observed combination of attributes and 267 unique labels are present in the shapefile of the survey. Some rare labels are only applicable to a few parcels, so it is necessary to merge detailed labels into crop categories of interest. In this study, crop categories were defined according to California Water Plan (CDWR, 2009). Irrigated crop water consumption in the rainless summer is the major concern of water planning, and the land use survey focused on the dry season which is also the growing season for most crops. A total of 13 summer crop categories were selected for classification and analysis, occupying about a half of the study area (Table 1). All non-crop land use types were aggregated into a general category called Other (OT).

2.2.2. Landsat imagery

The main remotely sensed data input is Landsat surface reflectance, available from Landsat Level-2 products developed and distributed by

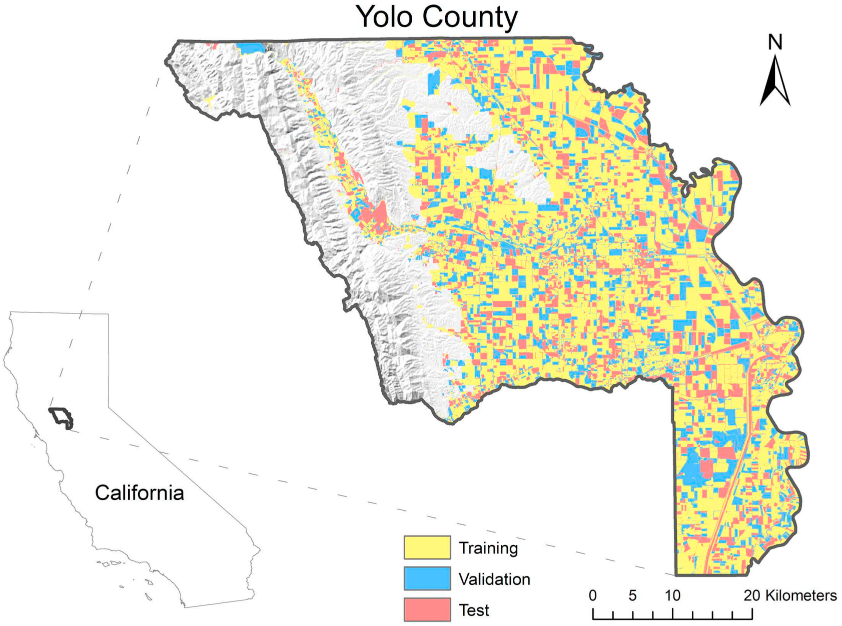


Fig. 1. The study area is the flat area of Yolo County, California. Mountainous areas shown in hillshade were excluded from the analysis. The study area is split into training, validation, and test sets in a parcel-wise random manner.

the USGS (WWW1, n.d). Level-2 surface reflectance of the year 2014 is derived from Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and Landsat 8 Operational Land Imager (OLI). ETM+ and OLI sensors provide multi-spectral optical bands at 30 m resolution, and both include bands for the calculation of Enhanced Vegetation Index (EVI) (Huete et al., 2002). The study area is covered in the footprint of Path 44 and Row 33. For the year 2014, a total of 37 Level-2 images of the scene were retrieved to create image time series, 19 from ETM+ and 18 from OLI. Gaps exist on processed images due to cloud, aerosol and the scan line corrector failure of Landsat 7. 20 images have less than 5% cloud cover, and 12 images have over 20% cloud cover. Landsat 7 ETM + images have an additional ~5% gap due to the scan line corrector failure.

EVI time series were calculated using blue, red, and near infrared surface reflectance from multi-temporal Landsat images (Huete et al., 2002). Although multi-spectral bands contain extra spectral information which is useful in crop type classification, our experiment employed the simplicity of one-dimensional EVI time series so that: i) experts experienced in crop phenology and time series interpretation are able to evaluate the classifier performance and possible sources of uncertainty, and ii) the analysis can focus on the contribution of time series information by effectively visualizing multi-temporal patterns and features generated by classification algorithms.

To generate time series with equal lengths and intervals, we used all possible Landsat overpassing dates in 2014 as the time steps. Each Landsat has a revisit frequency of every 16 days, and Landsat 7 and Landsat 8 orbits are offset to allow 8-day repeating coverage. The time steps are from DOY (day of year) 5, the first overpassing date, to DOY 365 with 8-day intervals, and all input time series have the same length of 46. Gaps caused by no observation, bad data or cloud cover were linearly interpolated using the nearest valid values before and after the time step. Linear interpolation is usually appropriate for short gaps (Kandasamy et al., 2013). We summarized the lengths of intervals between adjacent valid and usable observations. For all pixels in the full-year dataset, 65.2% of the intervals are 8-day long (the minimum length with two Landsat satellites), 18.6% are 16-day, 7.4% are 24-day, and 8.7% are 32-day or longer. Most of the long gaps occur in the cloudy winter. When focusing on the general growing season of summer crops from April to September, 74.1% of the intervals are 8-day, 16.1% are 16-day, 8.5% are 24-day, and 1.3% are 32-day or longer. Most gaps

Table 1
The current study used 14 categories for classification. The first 13 categories are from California Water Plan, and all other land uses were aggregated into the last category ("other").

Category code	Description	Number of parcels	Number of pixels	Areal percentage
RI	Rice	222	117,080	5.7%
SF	Safflower	131	40,252	1.9%
CR	Corn	120	21,609	1.0%
FI	Field crops other than SF and CR	372	102,064	4.9%
AL	Alfalfa	587	156,630	7.6%
PA	Pasture other than AL	402	62,901	3.0%
CU	Cucurbits (melons, squash, and cucumbers)	146	22,741	1.1%
TM	Tomatoes	614	175,990	8.5%
TR	Truck, nursery, and berry, other than CU and TM	249	18,370	0.9%
AP	Almonds and pistachios	522	95,812	4.6%
OR	Deciduous orchard other than AP	671	75,006	3.6%
SO	Subtropical trees (oranges, olives, etc.)	229	16,635	0.8%
VI	Vineyards	579	79,264	3.8%
OT	Other	14,932	1,084,213	52.4%
Total		19,776	2,068,567	100%

dropout is to avoid overfitting.

caused by missing data are short compared to the crop seasonality in the study area. We visually checked hundreds of plots of interpolation and decided that valid observations from the 37 images provide a sufficient temporal frequency to generate the input time series.

2.2.3. *Dataset partition*

Pixels of the whole study area were split into three datasets: training, validation, and test sets. The training set was used to train individual classification algorithms. The validation set was used to select the optimal hyper-parameters. The final classification results were evaluated with the test set. Dataset partition needs to follow two principles: i) these sets are independent from each other, and ii) the class distributions in all sets are similar (Rußwurm and Körner, 2017). In most croplands, pixels in the same parcel are very homogenous and highly correlated. Allocating pixels in a parcel to different sets will violate the principle of independency. Splitting the study area into relatively large sub-regions is not proper either. Many crop types spread unevenly in the study area, making it hard to delineate sub-regions with similar class distributions. Therefore, the basic spatial unit for dataset partition should be neither too small nor too large. We used land parcels as the basic unit to strike a balance between the two principles. Individual parcels usually differ in farming practices and management and crop types in different parcels can be seen as relatively independent from each other. We further grouped parcels using grids at 1 km interval so that parcels in the same grid are considered as a whole in dataset partitioning. Each parcel group was randomly assigned to one of the three sets approximately following the ratio of 60%:20%:20% (Fig. 1). There are 1,221,396 pixels and corresponding EVI time series in the training set, 431,920 in the validation set, and 415,251 in the test set.

2.3. *Classification*

2.3.1. *Neural network classifiers*

We tested three major types of deep neural networks to classify EVI time series: the first is plain deep ANNs called Multi-Layer Perceptron (MLP), in which neurons in a layer are fully connected to all neurons in neighboring layers, the second is one-dimensional convolutional neural network (Conv1D), which is a special form of CNN, and the third is LSTM in the RNN family. MLP is the simplest form of ANNs and can be used as a base to evaluate the classification performance without specialized architectures. Conv1D and LSTM represent two disparate but effective strategies to represent sequential data. Conv1D employs one-dimensional filters to capture the temporal pattern or shape of the input series. Conv1D layers can be stacked so that lower layers focus on local features and upper layers summarize more general patterns to a larger extent. An LSTM unit is designed to "remember" values over arbitrary time intervals, long or short. The use of LSTM improves the efficiency of depicting temporal patterns at various frequencies, which is a desirable feature in the analysis of crop growing cycles with different lengths.

Hyper-parameters of MLP include the number of hidden layers and the number of neurons in each layer. The search range of layer number was from 1 to 5. All the layers were set to have the same number of neurons, and values of 64, 128, 256, 512, and 1024 were tested. The MLP model was optimized by grid-searching the two hyper-parameters. Because each hidden layer in MLP is fully connected to the previous and the next layers, model complexity grows rapidly with more layers and neurons. The total number of parameters in the tested MLP architectures ranges from several thousands to millions.

The optimization of the Conv1D- and the LSTM-based models is more complex. Because of the versatility of the specialized architectures, there is no standard procedure to search for the optimal combination of hyper-parameters and various types of layers. In this study, the implementation of Conv1D was combined with pooling layers, fully-connected layers and dropout. Convolutional filter widths of 3, 5, and 7 were tested, which means that the filter window covers 3, 5, or 7 observations at an 8-day interval. The first convolutional layer has 32 or 64 channels and the channel number increases when going deeper. Pooling layers were fixed as max-pooling with a window size of 2. Dropout is a regularization technique that randomly drops some neurons in a layer during training so that the output of the layer does not rely on only a few neurons (Srivastava et al., 2014). The probability of dropping neurons was set to 20%, 30%, 40%, or 50%. We also designed and tested inception modules by concatenating Conv1D and max-pooling layers with different sizes to process multi-scale features at the same time (Szegedy et al., 2015). Each model contains two fully-connected layers at the output end. The last layer contains 14 neurons corresponding to the probability of the 14 classes. The second last layer collects information from previous layers as a flat array, and the size was determined by considering the dimension of the input to the layer. As a result, there are an extremely large number of potential network architectures and it is impossible to try them all. The selection of hyper-parameters was done step by step based on experience. We started with a relatively simple model with only one convolutional layer. Then we generated new models by changing one or two hyper-parameters, adding a new layer, re-ordering layers, or replacing a part of the network with a more complex component. Among the new models, models with promising classification performance on the validation set were used as the seeds to begin a new round of searching. In this way, the tested model grew in size and complexity until classification results did not improve further. The LSTM-based model was optimized in a similar way. LSTM layers were used jointly with fully-connected layers and dropout. Each LSTM layer might contain 32, 64, 128, or 256 output channels. Settings of dropout and fully-connected layers are the same as Conv1D-based models.

All the three types of ANNs were trained using the Adam optimizer (Kingma and Ba, 2014). Parameters of Adam were fixed as: $\beta_1 = 0.9$, $\beta_2 = 0.999$, ε = 1e-07, and a learning rate decay of 0.001. The learning rate was set to 0.001 and the batch size was set to 4096. As the training set is unbalanced, we used weighted cross-entropy loss function with weights inversely proportional to class abundance. Classification models were built and evaluated using the Keras library (WWW2, n.d) on top of Tensorflow (WWW3, n.d).

2.3.2. *Other classifiers*

For comparison, we also tried three non-deep-learning classifiers: XGBoost, Random Forest (RF), and Support Vector Machine (SVM). These classifiers are renowned for high performance and are often established as baseline models in classification tasks (Fernandez-Delgado et al., 2014). XGBoost is a state-of-art gradient boosting framework of decision trees (T. Chen and Guestrin, 2016), which recently gained popularity in the data science community as the major component of winning solutions of many machine learning competitions. A RF is an ensemble of decision trees based on the bagging technique (Breiman et al., 1984; Liaw and Wiener, 2002). A SVM is a classifier defined by separating hyperplanes and can perform non-linear classification using kernel functions (Cortes and Vapnik, 1995). RF and SVM have been extensively used in remote sensing applications and achieved great success in complex classification tasks (Lawrence et al., 2006, Na et al., 2010, Shi and Yang, 2016, Chapelle et al., 1999, Carrão et al., 2008, J. Zhang et al., 2014).

In this study, classification results by XGBoost, RF and SVM were used as a reference which represents the performance of today's popular non-deep-learning algorithms. For RF and SVM, we used the Python implementation in the package Scikit-learn (WWW4, n.d), and for XGBoost we used the specific Python package (WWW5, n.d). All naming conventions of classifiers and parameters follow the documentation of the corresponding package. To deal with dataset imbalance, training parameters like class weight or sample weight were set as inversely proportional to class abundance so that each class has equal contribution.

Each classifier requires a set of hyper-parameters which need to be

Table 2
Hyper-parameters of three non-deep-learning classifiers and corresponding value ranges for random search.

Classifier	Hyper-parameter	Candidate values	Selected values for input sets		
			Series (S)	S + HANTS	S + TIMESAT
XGBoost	learning_rate	0.01, 0.015, 0.025, 0.05, 0.1	0.1	0.1	0.1
	gamma	0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 1	0.05	0.05	0.05
	max_depth	5, 7, 9, 12, 15, 17, 25	17	17	25
	min_child_weight	1, 3, 5, 7	3	3	5
	subsample	0.6, 0.7, 0.8, 0.9, 1	0.9	0.9	0.7
	colsample_bytree	0.6, 0.7, 0.8, 0.9, 1	0.8	0.8	0.6
	reg_lambda	0.01, 0.1, 1	0.1	0.1	0.1
	reg_alpha	0, 0.1, 0.5, 1	0.1	0.1	0
RF	n_estimators	120, 300, 500, 800, 1200	300	300	300
	max_depth	5, 8, 15, 25, 30, None	30	30	30
	min_samples_split	2, 5, 10, 15, 100	2	2	2
	min_samples_leaf	1, 2, 5, 10	1	1	1
	max_features	'log2', 'sqrt', None	'sqrt'	'sqrt'	'sqrt'
SVM	C	0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1, 3,10,30,100,300,1000	10	3	3
	gamma	0.1, 1, 2, 10, 'auto'	2	2	2

configured during classification model development. Optimal values of hyper-parameters were selected according to the classification accuracy of the validation set. We employed a "random search" strategy to optimize major hyper-parameters of chosen classifiers: the classifier was repeatedly trained many times and each run was based on a random sample from all combinations of hyper-parameter values (Bergstra and Bengio, 2012). Random search greatly improves the efficiency of hyper-parameter optimization when classifier performance is mostly influenced by only a subset of hyper-parameters. Details of hyper-parameter optimization for XGBoost, RF, and SVM are presented in Table 2.

We also tried training XGBoost, RF, and SVM using seasonality metrics extracted from phenology analysis packages. The extraction of seasonality metrics provides additional information that are physically meaningful and useful for classification. Two widely-used algorithms/software were used, HANTS (Verhoef, 1996; Roerink et al., 2000) and TIMESAT (Jönsson and Eklundh, 2004). In HANTS, five components of Fourier Transform were calculated, and five amplitude values and four phase values (the phase of the first component is zero) were used as additional classification inputs (Evans and Geerken, 2006; Geerken, 2009). In TIMESAT we chose the double-logistic fitting method that was used by many studies to model growing cycles (X. Zhang et al., 2003, Fisher et al., 2006, Beck et al., 2006, Fisher and Mustard, 2007, Soudani et al., 2008, Dannenberg et al., 2015, Gonsamo and Chen, 2016). The output of TIMESAT includes 13 seasonality parameters: time for the start of the season, time for the end of the season, length of the season, base VI level, time for the mid of the season, season maximum VI, seasonal amplitude, rate of increase at the beginning of the season, rate of decrease at the end of the season, large seasonal integral, small seasonal integral, VI value for the start of the season, and VI value for the end of the season. For each of the classifier, three input sets were tested: a) EVI time series, b) EVI time series plus seasonality metrics from HANTS, and c) EVI time series plus seasonality metrics from TIMESAT.

2.4. Evaluation

We computed confusion matrices and overall accuracy of the test set to evaluate the performance of all classifiers. Overall accuracy is proportional to the area that is correctly mapped and is suitable for area estimation. In the test set, 1% accuracy corresponds to about 374 ha. We also used the macro-average of F1 score as an indicator of classification capability. For each class, F1 score is the harmonic mean of producer's accuracy and user's accuracy:

$$F1_{class} = \frac{1}{\frac{1}{2}\left(\frac{1}{A_{prod}} + \frac{1}{A_{user}}\right)} = \frac{2\,A_{prod} \cdot A_{user}}{A_{prod} + A_{user}}$$

where $F1_{class}$ is the F1 score of a single class, A_{prod} is the producer's accuracy of the class, and A_{user} is the user's accuracy of the class. The macro-average of F1 score is the simple average of all $F1_{class}$ values, which assigns equal weights to all classes and ignores the actual frequency of occurrence. Because about a half of the study area is occupied by the non-crop class "Other", the importance of this class is reduced in the evaluation based on macro-averaged F1 score compared to that based on overall accuracy. The macro-average of F1 score is of particular interest as it highlights the model capability of identifying the spatial distribution of crops, especially for relatively rare crop types.

3. Result

The optimized MLP architecture is relatively shallow with two hidden layers, and each layer has 512 neurons. Deeper MLP models did not improve the validation set accuracy. The training time of MLP models ranged from a few minutes to 2 h on a P4000 GPU. The selected model was trained for half an hour until the validation accuracy reached a plateau. Architectures of the Conv1D-based and the LSTM-based networks with optimal performance are demonstrated in Figs. 2 and 3, respectively. The optimized Conv1D-based model mainly includes three Conv1D layers with a width (kernel size) of 3 and an inception module. The output of the inception module is a concatenation of a width 3 Conv1D layer, a width 5 Conv1D layer, and a max-pooling layer followed by a width 1 Conv1D layer. Dropout is applied after the inception module, the last two Conv1D layers, and the fully connected layer to improve generalization capability. The training process of Conv1D-based networks took about 1 to 10 h to have a stable validation accuracy, and the time of the selected architecture was about 3 h. The architecture of the optimal LSTM-based model mainly consists of four LSTM units followed by dropout. The number of output channels of each LSTM unit is relatively small, only 32, as further increasing the number of channels deteriorated classification accuracy. The training time of the LSTM-based networks was relatively long, about 9 to 50 h, and the selected one took 16 h to finish training.

The two kinds of deep learning models show distinct classification capability (Table 3). The optimized Conv1D-based has the highest test set accuracy (85.54%) among all deep learning and non-deep learning models. The macro-averaged F1 score of Conv1D (0.73) is much higher than other classifiers (XGBoost: 0.69, SVM: 0.68, RF: 0.67, and LSTM:

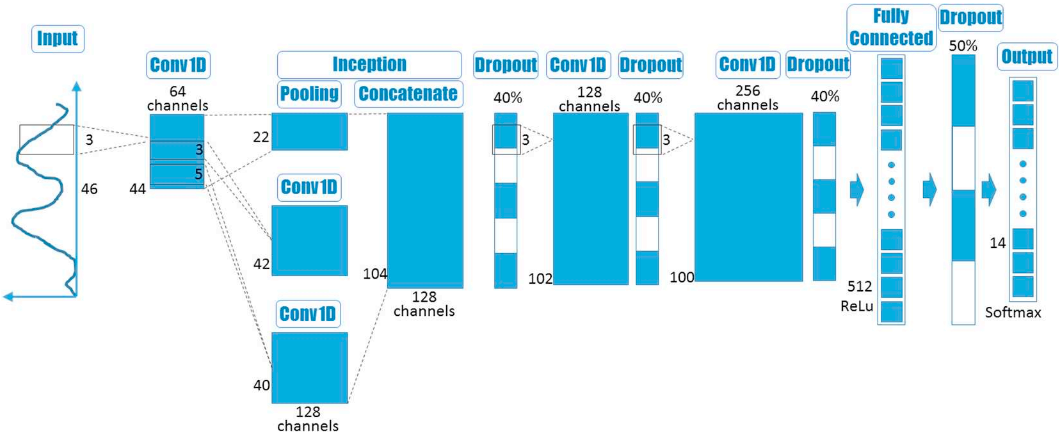


Fig. 2. Architecture of the optimal Conv1D-based model.

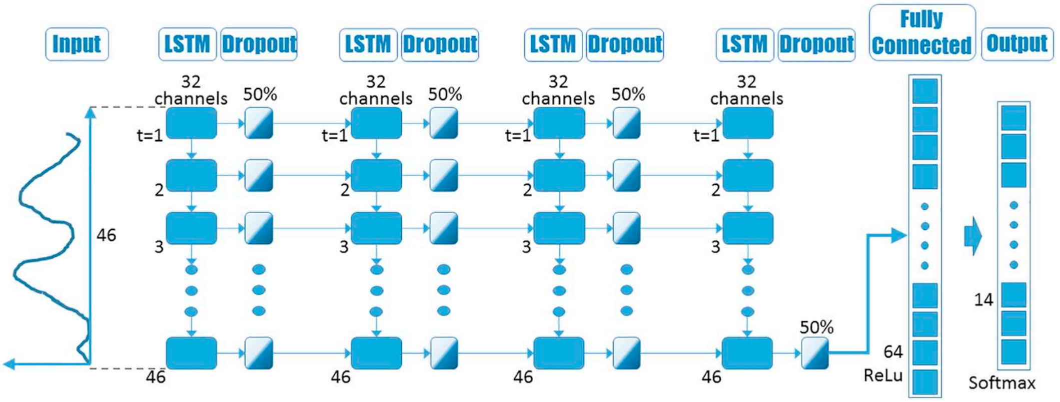


Fig. 3. Architecture of the optimal LSTM-based model.

0.67). By contrast, the best accuracy achieved by LSTM-based models is 82.41%, which is inferior to all non-deep learning models (XGBoost: 84.12%, RF: 83.38%, and SVM: 82.45%). Although the LSTM technique was initially developed for time series related tasks, it does not seem to be suitable for the current classification experiment.

Optimal parameter values found using random search are appended to Table 2. For each non-deep-learning classifier, the three input sets yielded similar results. Adding seasonality metrics to model inputs did not greatly improve classification results, and the accuracy is often lower with additional metrics in the input (Table 3). For TIMESAT, curve-fitting fails for about 20% of the input VI series because some special seasonal patterns in the study area cannot be properly fitted by the double logistic function (further discussed in Section 4.1). Although curve-fitting retrieves useful seasonality metrics from VI series, the method also brings uncertainty that partly offsets the benefits. Another possible explanation is that the number of pixels in our training set is large enough so that derived features cannot further contribute to the classification. Seasonality metrics calculated by double logistic fitting are usually useful when there are no or only a small number of field samples for training (Zhong et al., 2011; Zhong et al., 2012; Zhong et al., 2016a), and/or when the distributions of series are different between training and test sets due to inter-annual or inter-regional variability (Zhong et al., 2014; Zhong et al., 2016a, 2016b). For our current classification task, the advantage of using seasonality metrics derived from fixed-form equations is limited. By contrast, more improvement is seen by using the Conv1D-based classifier, in which a more flexible way of temporal feature extraction is embedded and time series preprocessing for explicit seasonality derivation is unnecessary. Further analysis and discussion will focus on classification experiments with only EVI series as the input.

Table 3
Overall accuracy and macro-averaged F1 score values achieved by various classifiers and inputs on the test set.

Classifier type	Input	Overall accuracy	Macro-averaged F1
MLP	Series only	83.81%	0.69
Conv1D-based	Series only	85.54%	0.73
LSTM-based	Series only	82.41%	0.67
XGBoost	Series only	84.12%	0.69
	Series + HANTS	84.17%	0.69
	Series + TIMESAT	84.09%	0.69
RF	Series only	83.38%	0.67
	Series + HANTS	83.25%	0.67
	Series + TIMESAT	83.26%	0.67
SVM	Series only	82.45%	0.68
	Series + HANTS	83.09%	0.69
	Series + TIMESAT	82.95%	0.68

Tables 4 and 5 show confusion matrices yielded by XGBoost and the Conv1D-based classifier, respectively, which can be considered as the best results by non-deep-learning and deep-learning algorithms for EVI series data, respectively. Table 6 highlights the difference (Table 5 minus Table 4) to demonstrate the relative strength and weakness of the two classifiers on individual classes. Classification errors, represented by Hectare numbers off the diagonal in the confusion matrix, can be generally grouped into several types: i) “inter-crop” errors between the 13 crop types, ii) “crop as non-crop” errors caused by incorrectly classifying crop types as Other (OT) (off-diagonal values in the last

Table 4
Confusion matrix of the test set by the XGBoost classifier using EVI series as the input. Values in the matrix are in Hectare.

Reference classes	Classified														Total	Producer's accuracy
	RI	SF	CR	FI	AL	PA	CU	TM	TR	AP	OR	SO	VI	OT		
Rice (RI)	**1818**		3	1				1						103	1925	94%
Safflower (SF)		**265**	74	33	1			14			1			211	599	44%
Corn (CR)	21	8	**236**	13			2	89						50	418	56%
Other field (FI)	50	11	1	**1194**	1	8	73	232	2				2	379	1952	61%
Alfalfa (AL)					**2131**	22		43			1			340	2539	84%
Other pasture (PA)	7			20	63	**571**		16		1	7	1	1	333	1020	56%
Cucurbit (CU)	1		1	14		1	**193**	80	8		1			261	560	35%
Tomato (TM)		5	13	133	4		10	**2648**	1					230	3044	87%
Other truck (TR)			2	26	1	1	19	58	**41**		1		5	180	336	12%
Almond & pista. (AP)					1	5				**1508**	40		5	564	2124	71%
Other orchard (OR)					3	3			2	14	**802**	2	53	434	1314	61%
Subtropical (SO)										1	4	**144**		94	243	59%
Vineyard (VI)						7				5	35	2	**977**	458	1485	66%
Other (OT)	43	13	14	101	143	109	9	147	11	99	100	36	82	**18,908**	19,812	95%
Total	1940	302	344	1535	2349	726	307	3328	66	1628	991	186	1125	22,546	37,373	
User's Accuracy (%)	94%	88%	69%	78%	91%	79%	63%	80%	63%	93%	81%	78%	87%	84%		
Overall accuracy = 84.12%								Macro-averaged F1 score = 0.69								

Numbers in bold on the diagonal represent correct classification.

column of the confusion matrix), and iii) "non-crop as crop" errors caused by classifying OT as crops (off-diagonal values in the last row of the confusion matrix). Compared to the results by XGBoost, the Conv1D-based classifier resulted in 284 Ha less inter-crop errors and 481 Ha less crop as non-crop errors while increasing non-crop as crop errors by 233 Ha. Values on the diagonal of Table 6 are mostly positive (in green) and off-diagonal values are generally negative (in red) except for OT, suggesting the strength of Conv1D to separate individual crop types. In general, the XGBoost classifier tends to omit crops and identify more pixels as OT which is the largest class, and the Conv1D-based classifier is more likely to seek for further separation of the 13 crops. Conv1D strikes a balance between producer's and user's accuracies for crop classes each of which occupies at most 9% of the study area. This explains the larger difference in macro-averaged F1 score (0.73 vs. 0.69) than that in overall accuracy. Therefore, the use of Conv1D especially improves the classification results of crop classes. The Conv1D-based classifier is suitable for depicting the spatial distribution and carrying out crop-specific studies for certain crops. Test set parcel groups classified by the Conv1D-based model are shown in Fig. 4(b) for visual comparison with corresponding parcels in the ground reference map in Fig. 4(a).

4. Discussion

4.1. Interpreting the behavior of the deep network model

The two deep neural networks we designed treat input EVI series in different ways. A LSTM cell models the time series with units changing step by step from time 1 to time *n*. A Conv1D layer has a pattern or shape template in each of its channels and matches the patterns with the input through convolution. It is valuable to look into the components of the Conv1D-based model, the top classifier of the present study, to understand how the model recognizes EVI series, which is often a difficult task due to the complexity of the deep neural network.

We visualized the activation of the input series on various layers using deconvolution and guided back-propagation (Zeiler and Fergus, 2014). Fig. 5 includes the plots of 6 EVI series in the test set and the corresponding activation on selected channels in the first and the last convolutional layers and the final fully connected layer (the "softmax" layer). In the first column, the first plot shows a typical EVI curve of tomato which grows from June to September, or DOY ~150 to ~270. The second plot is the EVI series of an almond field, which belongs to the crop category almond & pistachio (AP). The curve shows rapid

Table 5
Confusion matrix of the test set by the Conv1D-based classifier using EVI series as the input. Values in the matrix are in Hectare.

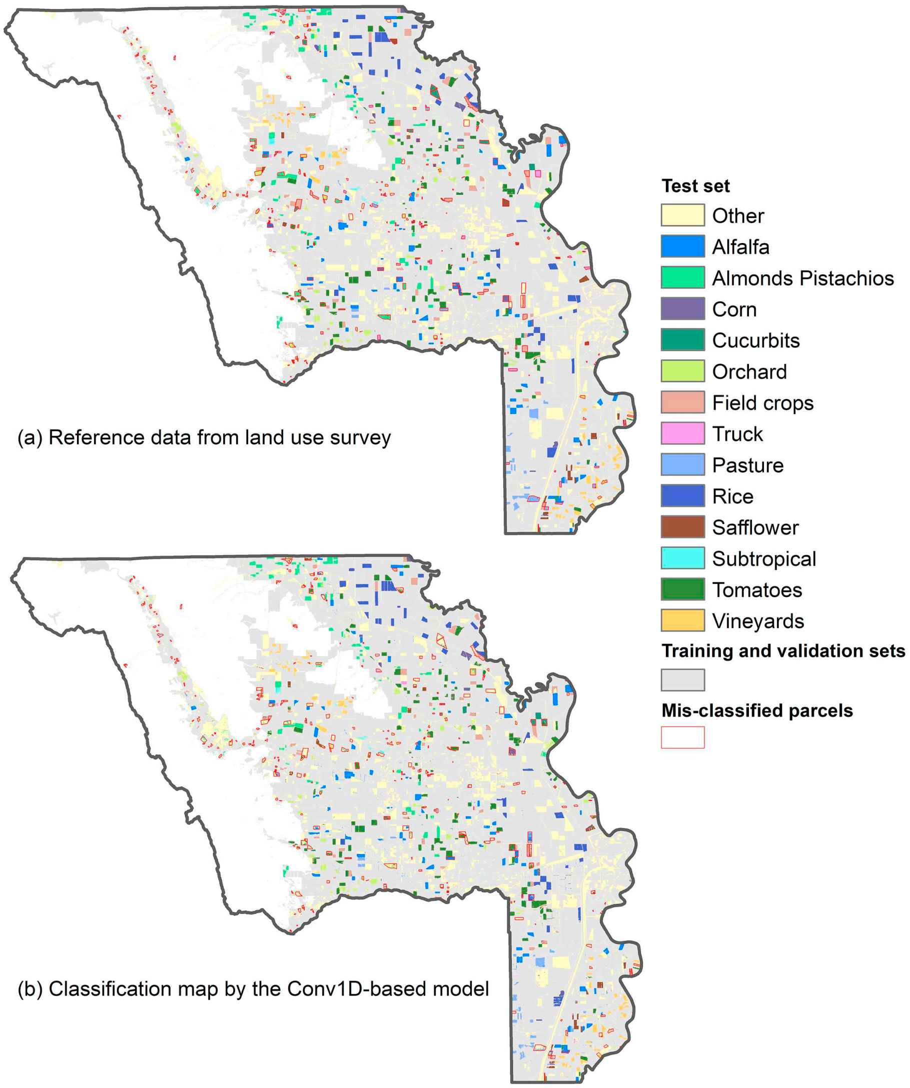
Reference classes	Classified														Total	Producer's accuracy
	RI	SF	CR	FI	AL	PA	CU	TM	TR	AP	OR	SO	VI	OT		
Rice (RI)	**1839**		1											84	1925	96%
Safflower (SF)		**357**	62	10	1			19						150	599	60%
Corn (CR)	10	6	**269**	11				78						42	418	64%
Other field (FI)	39	40	8	**1269**	6	14	55	135	5					381	1952	65%
Alfalfa (AL)					**2133**	20	9	44			1			332	2539	84%
Other pasture (PA)	7			19	56	**617**	1	1		1	11		1	307	1020	60%
Cucurbit (CU)			1	19	1		**194**	55	16		1			271	560	35%
Tomato (TM)		5	11	60	6		9	**2739**	3		1			211	3044	90%
Other truck (TR)	5		7	6	1		34	48	**57**		1		3	172	336	17%
Almond & pista. (AP)						1				**1567**	33		12	512	2124	74%
Other orchard (OR)				3	3	4		1	3	14	**900**	2	29	355	1314	68%
Subtropical (SO)										2	9	**157**		75	243	65%
Vineyard (VI)						6		1		5	11	2	**1194**	265	1485	80%
Other (OT)	53	25	11	137	144	105	9	177	52	126	136	39	123	**18,675**	19,812	94%
Total	1954	433	371	1534	2351	767	311	3298	137	1715	1103	202	1364	21,832	37,373	
Users' Accuracy (%)	94%	82%	73%	83%	91%	80%	62%	83%	42%	91%	82%	78%	88%	86%		
Overall accuracy = 85.54%								Macro-averaged F1 score = 0.73								

Numbers in bold on the diagonal represent correct classification.

Table 6
Difference between confusion matrices of the Conv1D-based and the XGBoost classifiers. Values are calculated as Table 5 minus Table 4.

Reference Classes	Classified													
	RI	SF	CR	FI	AL	PA	CU	TM	TR	AP	OR	SO	VI	OT
Rice (RI)	**21**		−2					−1						−19
Safflower (SF)		**91**	−12	−24				6						−62
Corn (CR)	−11	−2	**34**	−1			−2	−10						−7
Other field (FI)	−11	29	8	**74**	5	7	−18	−97	3				−1	2
Alfalfa (AL)					**1**	−2	8							−8
Other pasture (PA)	−1			−1	−7	**46**	1	−15			5			−27
Cucurbit (CU)	−1			5	1		**1**	−25	9					10
Tomato (TM)			−3	−73	2		−1	**92**	2					−19
Other truck (TR)	5		5	−20		−1	15	−11	**16**				−2	−8
Almond & pista. (AP)					−1	−5				**59**	−7		7	−53
Other orchard (OR)				2		1		1	1		**98**		−24	−79
Subtropical (SO)										1	5	**13**		−19
Vineyard (VI)						−1					−25		**218**	−193
Other (OT)	11	12	−4	36	2	−4		30	42	28	36	3	41	**−233**

Numbers on the diagonal are in bold. Positive values are in green and negative values in red.



(a) Reference data from land use survey

(b) Classification map by the Conv1D-based model

Fig. 4. Reference crop type map from field survey (a) and the crop map produced by the Conv1D-based model which has the highest accuracy (b). Only parcels in the test set are shown in colors, and training and validation sets are in grey. Mis-classified parcels are identified based on the majority type of all pixels within the parcels and delineated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

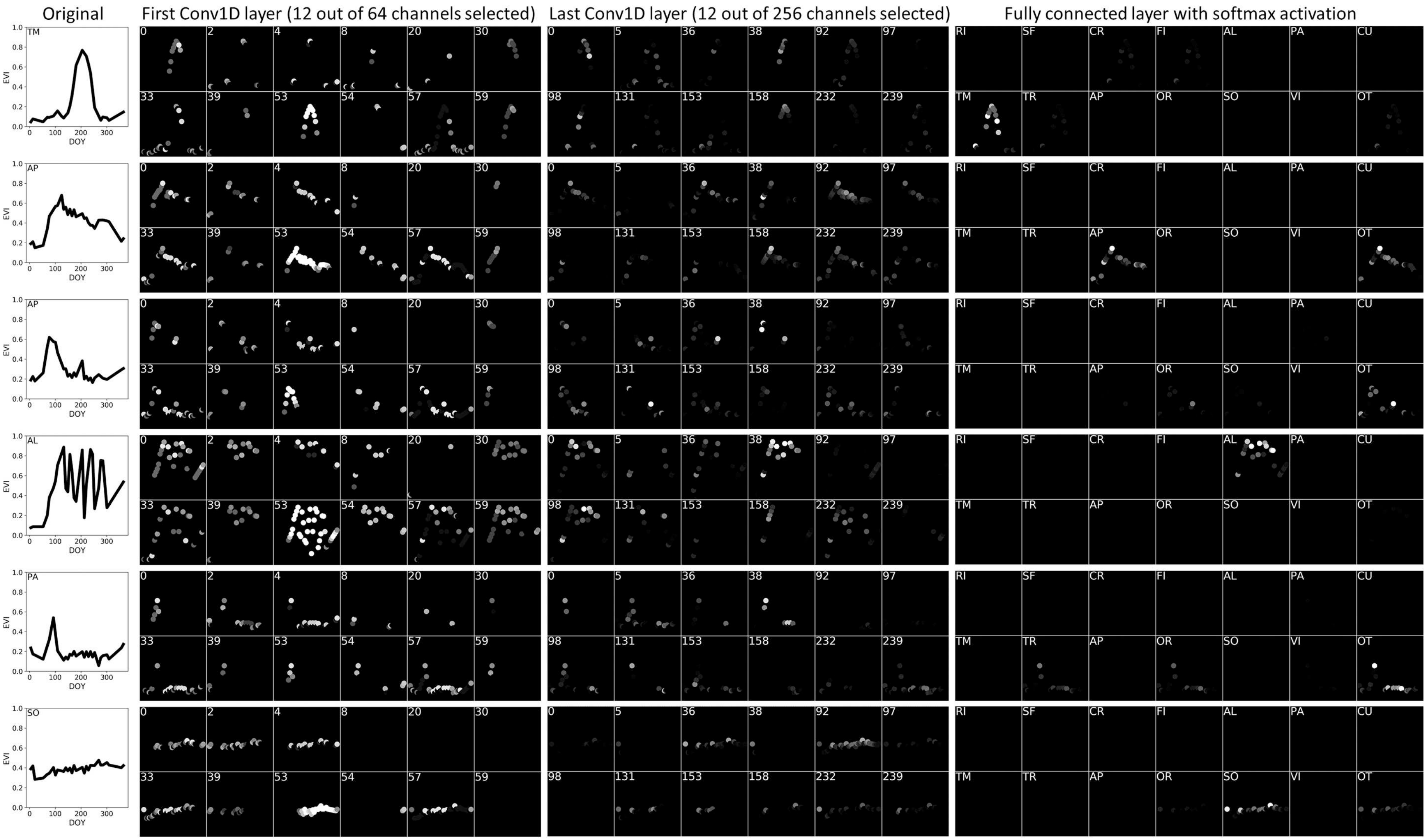


Fig. 5. Examples of EVI series and corresponding activation on the first and the last Conv1D layers and the fully connected layer. The first column includes 6 input EVI series. The crop types are labeled on the upper-left corner: TM - Tomatoes, AP - Almonds and pistachios, AL - Alfalfa, PA - Pasture other than AL, and SO - Subtropical trees. Each column on the right represents a layer, and each numbered box represents a selected channel in the layer. The activation of individual time steps on the three layers was calculated by deconvolution and guided back-propagation. EVI values at different time steps are shown by dots, and the brightness of dots represent the strength of the activation by the current time step.

green-up in early spring and a gradual decrease after peak greenness. The fluctuations in the growing season are possibly caused by irrigation schedule and occasional mowing of cover crops. The third plot is also from a field that was recorded as almond during the field visit, however, the growing pattern of the EVI series does not seem complete. The fourth plot is a typical alfalfa curve, which shows frequent cuts for cattle feed. Alfalfa is usually harvested about 4 to 7 times in a year in the study area. The field corresponding to the fifth plot was seen as miscellaneous pasture with limited irrigation, which may explain the low EVI during most of the summer. The last plot is for orange (in crop category SO), and the seasonal EVI is characterized by a relatively constant level as the nature of subtropical trees.

To facilitate visualization, the activation on each channel of each layer was back-propagated to the input of the model to demonstrate which part of the input series activates the channel or the neuron most. In Fig. 5, the input EVI series are plotted and the activation is denoted with brightness (bright dots indicate that the part of the input series strongly activates the channel while dark dots indicate weak activation). In general, the input series activate model layers in different ways. In the first Conv1D layer, which is closest to the input, channels are more responsive to small-scale variations or textures. For example, channels #2 and #4 are activated by local peaks. #39 also involves local peaks but it is only activated by peaks that are more apparent. #53 highlights the top half of all peaks. #8 and #59 are mostly activated by the part of the curve that almost reaches the peak. #57 is activated by the decreasing slope of the series. The textures are relatively simple and could be matched by many small segments in the EVI curve. As a result, the activation strength is distributed throughout the whole series. When the input series pass through a few Conv1D, pooling and dropout layers to the last Conv1D layer, the patterns of activation become more complex. For example, channel #232 appears to be sensitive to the growing pattern with a relatively long duration. #97 captures the period with decreasing EVI for some input series when the decreasing period is long and has a flat bottom. These channels in the last Conv1D layer are activated by certain overall properties of the time series and sometimes yield patterns that cannot be defined with simple rules. By forward-propagating through Conv1D layers in the model, simple patterns are aggregated into more complex ones in a hierarchical manner. In general, activations on later Conv1D layers tend to be localized, only occurring at places where the specific complex patterns are matched to the curve.

Another way to understand what the deep convolutional network learns is to generate the input series that maximizes the activation on individual channels of convolution layers. Because activation maximization by back-propagation usually produces unrealistic series with very high magnitude and noisy patterns, we employed the methods of L2-regularization and Gaussian blur for improved visualization (Yosinski et al., 2015). The resulted series give ideas about what patterns will activate certain channels of Conv1D layers most. In Fig. 6, the same channels as Fig. 5 were selected, and the input series that yield maximum activation on these channels are presented. The activations on the first Conv1D layer are mostly maximized by simple repeating patterns, which is typical for a single convolution layer. Only textures at a small scale are needed to yield maximum activations. By contrast, at the last Conv1D layer, the activation is a joint effect of all previous layers, and the maximum activations are produced by more complex input series with certain overall patterns.

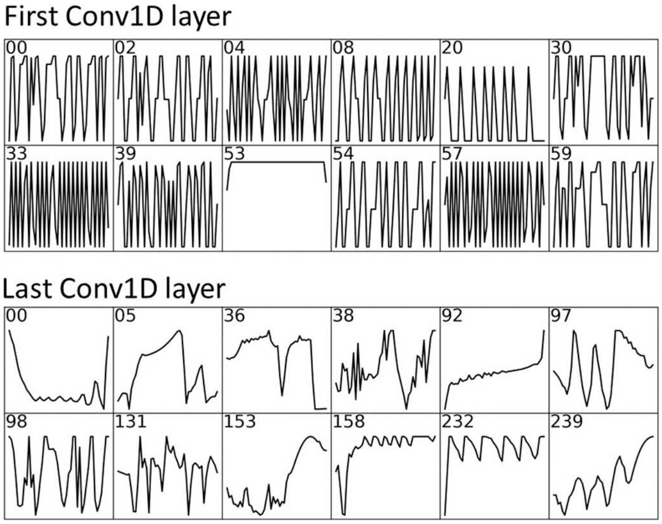


Fig. 6. Input series that maximize the activation on selected channels of the first and the last Conv1D layers. Selected channels are consistent with Fig. 5.

The final fully-connected layer combines high-level features and calculates the likelihood of each class via softmax. In Fig. 5, for the first series (tomato), EVI values in the growing season strongly activate the channel for tomato and the pixel is correctly classified. Some stages in the growing season also slightly activate classes including corn, other field crops, other truck crops, and non-crop due to the similarity in phenology. In reality the planting date of tomato and most field crops and truck crops are relatively flexible and their growing season lengths are close. Multi-spectral data are usually required for the challenging task of classifying these summer field crops and truck crops. Considering that we used only EVI series in the classification for the purpose of visually interpreting classifier behavior, the accuracy is quite high for tomato, which is the largest crop class in the study area. The second series is a typical almond curve. The overall pattern in the growing season activates the crop category almond & pistachio (AP) more, and activates the other (OT) class less, possibly because the curve resembles irrigated non-crop trees or orchards that were missed in field visits. The third series is also almond according to the field record, but the pixel was classified as OT while the series slightly activates other orchard (OR) and subtropical (SO) too. Based on our experience, the seasonal pattern is so different from normal almond that the trees are unlikely to be managed and harvested properly. In 2014 and 2015, California suffered from severe drought and the water supply to many farms was curtailed in the driest months in summer. It is possible that the almond field was observed in the beginning of the growing season and later abandoned as a result of low forecasted yield (old trees) and high water cost. Similarly, the fifth plot corresponds to a field that was observed as pasture but classified as OT. The field visit recorded a special condition that the pasture field was not fully irrigated. Under water shortage, the low-value pasture probably experienced discontinued irrigation for most of the summer. The two examples of misclassification suggest that some classification errors were caused by temporal inconsistency between reference data collection and series-based classification: the former is instantaneous, while the latter uses whole-year input data. As for the Alfalfa and the orange plots, they both have prominent activations in the correct class channel, which indicates that the unique seasonal patterns of the two classes under ideal growth conditions can be clearly recognized by the model.

4.2. Strategies of handling VI time series in classification

The hierarchical feature generation process of the Conv1D-based model provides a flexible way to formulate and identify complex seasonal patterns in the EVI time series. No mathematical equations are needed to pre-define or specify the patterns of interest. Instead, pattern features are automatically extracted during training via back-propagation. Fig. 5 shows examples of low level features (the first Conv1D layer) and high level features (the last Conv1D layer) retrieved from the input series. The 6 EVI series are so different from each other that it is extremely difficult to define a universal mathematical model to represent the time series. For example, the double-logistic function, which is implemented in TIMESAT (Jönsson and Eklundh, 2004) and widely used in phenology studies (X. Zhang et al., 2003, Fisher et al., 2006, Beck et al., 2006, Fisher and Mustard, 2007, Soudani et al., 2008, Dannenberg et al., 2015, Gonsamo and Chen, 2016), is suitable to fit the EVI curve of tomato. For the almond curve, the fitting uncertainty is likely to increase because of the sharp peak, the slow-decreasing segment, and the high fluctuations on the decreasing segment. The rapid increase-and-decrease pattern of alfalfa and the constant pattern of orange apparently cannot be handled by the double-logistic function effectively. As for the two mis-classified curves with special conditions, the double-logistic function may or may not capture the growing season phenology depending on whether the impaired growth still shows a full growing cycle. In California, the diversity of economic crops results in a variety of growing patterns, which increases the difficulty of phenology derivation compared to other major agricultural regions in the US. The Conv1D-based model directly takes the original EVI series as the input and embeds feature extractors in the training process instead of predetermining functions of a specific form to represent crop growing cycles. Therefore, the model is especially appropriate for the study area where no existing curve functions are universally applicable to various patterns of seasonal dynamics.

In the present study we tested using original EVI series as the input to classifiers. For tree-based classifiers (XGBoost, RF, etc.) and SVM, each time step in the series is treated as an independent dimension and the sequential relationship is not utilized. In this case, the trained classification algorithm may not be robust to variances caused by weather conditions, agricultural practices or missing data. For example, a small growing season shift of only 8 days is likely to result in totally different branches of trees in XGBoost and RF or positions relative to the hyperplane in SVM. To solve this problem, people may either increase the size of the training set to cover all the possible growing season patterns, or derive phenological features and define rules to adjust model sensitivity to certain features. In practice, there is often a dilemma when determining the input data of multi-temporal crop mapping tasks. The use of original time series in classification is simple and straightforward, but the information in the series may not be fully utilized particularly for classifiers that are not designed to handle temporal relationships. Phenology extraction by fitting pre-defined curve functions provides useful features with physical meanings, but the application is limited by the choice of curve function. The Conv1D-based model somehow strikes a balance between the two strategies: the model employs multi-level pattern features to improve the representation of EVI series, but it does not restrict the form of the patterns. The results of this study show that the Conv1D-based framework is a proper strategy to effectively utilize temporal relationships for time series processing and meanwhile reduce user intervention for automated classification. The model behavior to some extent resembles the process that human perceive and identify one-dimensional time series: human beings prefer to plot the series against time and tend to interpret the series based on the shape of the plotted curve, examine patterns at various scales, and compare the shape of the curve to patterns in memory. In the future the Conv1D-based model may provide a possibility to combine human knowledge and experience with patterns learnt by algorithms in a single framework for improved classification.

4.3. Possible extensions of the current study

One of the main objectives of this study is to develop and visualize temporal patterns based on deep neural networks. In the common use of CNNs and other network architectures, there are many widely-implemented strategies that can possibly improve classification accuracy, for example, 2D convolution in the spatial dimension and end-to-end learning of spectral features. Although we tested a few of these possible extensions on our dataset, the relatively conventional options are not presented in the Method and material section and the Result section for a fair comparison between pure temporal representation approaches and clear and concise visualization of temporal patterns. In practice, potential extensions should be tested and added to the classification framework whenever the results can be improved.

4.3.1. Spatial convolution

The original and the most popular use of CNNs in remote sensing is to apply filters to the spatial dimensions to generate feature maps (Zhu et al., 2017). To compare with the temporal convolution approach proposed in this study, we tested network architectures with spatial convolution. We first tried models with 2D spatial convolution (Conv2D) and kept the temporal dimension fully-connected. The best Conv2D-based architecture achieved 84.69% accuracy and 0.70 F1 score on the test set. The result was slightly inferior to that by 1D temporal convolution (85.54% accuracy and 0.73 F1 score), suggesting that temporal convolution is more useful than spatial convolution for this specific dataset and classification task. In crop mapping, crop seasonality from relatively dense EVI series (8-day or longer) is likely to be more important than spatial patterns derived from 30-m medium-resolution images. Then we attempted to capture the hierarchical patterns in the spatial and the temporal dimensions simultaneously by using 3D convolution (Conv3D). The selected Conv3D-based network resulted in 86.31% accuracy and 0.73 F1 score, which were the highest among all types of classifiers. It is expected because the Conv3D-based model utilized patterns in both the temporal and the spatial dimension. As spatial convolution has been widely used in existing literatures, in this paper we focus on 1D time series classification for a fair comparison among various approaches of temporal representation. To yield a high accuracy in future applications, model architectures based on 3D spatiotemporal convolution should be considered.

4.3.2. Feature extraction in the spectral dimension

In our experiments we also tested using Landsat surface reflectance directly as classifier inputs, which is a common approach to employ spectral information to separate crop types. Given sufficient data, deep models are supposed to be able to learn the most appropriate band combination for a specific task, possibly eliminating the need of using vegetation indices. We used the multi-temporal series of all Landsat surface reflectance bands as the input to the Conv1D-based models instead of the EVI, and other settings were the same as the Conv1D experiment. The result (83.35% accuracy and 0.64 F1) was poorer than using EVI (85.54% accuracy and 0.73 F1). The difference suggests that the spectral information from individual bands is not as useful as the EVI for the classification task. Widely-used indices like the EVI were developed based on the experience and knowledge gained from numerous remote sensing studies and have been tested extensively in past decades. The EVI series can reflect the seasonal dynamics of vegetation vigor and reduce the noise in the reflectance series (Huete et al., 2002). Our study area is unlikely to be large enough to generalize a band combination with comparable performance to the EVI. In addition, in our past analysis on spectral separability between land use types in the Central Valley, California, crops were not separable in the wavelength ranges of the Landsat bands. As a result, using Landsat surface reflectance did not provide additional spectral information for crop mapping in the study area. In this study, the spectral dimension shrank to one variable (the EVI) to utilize the advantages of the index, reduce the noises in the time series, and simplify the visualization. For other applications, it is recommended to test the option of directly using multi-spectral bands to see if the spectral features learnt from the specific dataset could outperform vegetation indices that were developed for general purposes.

5. Conclusion

In this study we developed deep neural networks with various architectures to classify summer crops using EVI time series. The most satisfactory classification result was achieved by a deep network model built with one-dimensional convolutional (Conv1D) layers and an inception module. According to the overall accuracy and the capability of identifying individual crop types, the optimal model is superior to popular classifiers such as XGBoost, Random Forest, and Support Vector Machine as well as deep neural network models based on recurrent layers. The advantage of the optimal model shows the importance of a properly designed architecture in time series classification. We employed visualization techniques to examine what the deep network model learns and how it understands the input EVI series. A Conv1D layer is able to capture certain temporal patterns, and a stack of Conv1D layers aggregates simple patterns into complex shapes to represent real-world time series. The hierarchical architecture of the Conv1D-based model uses time series as classification inputs and automatically extracts features of crop growth dynamics during model training. The model does not require pre-determined curve functions or mathematical assumptions for crop seasonality in specific areas, which provides a flexible and highly automated way for crop mapping applications. The study suggests that among numerous alternative methods of time series representation in classification tasks, a deep neural network architecture built with Conv1D layers is a viable option.

References

Audebert, N., Le Saux, B., Lefèvre, S., 2018. Beyond RGB: very high resolution urban remote sensing with multimodal deep networks. ISPRS J. Photogramm. Remote Sens. 140, 20–32.

Badhwar, G.B., 1984. Automatic corn-soybean classification using Landsat MSS data. II. Early season crop proportion estimation. Remote Sens. Environ. 14, 31–37.

Beck, P.S.A., Atzberger, C., Høgda, K.A., Johansen, B., Skidmore, A.K., 2006. Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. Remote Sens. Environ. 100, 321–334.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305.

Bradley, B.A., Mustard, J.F., 2008. Comparison of phenology trends by land cover class: a case study in the Great Basin, USA. Glob. Chang. Biol. 14, 334–346.

Bradley, B.A., Jacob, R.W., Hermance, J.F., Mustard, J.F., 2007. A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. Remote Sens. Environ. 106, 137–145.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. CRC press.

Brown, M.E., de Beurs, K., Vrieling, A., 2010. The response of African land surface phenology to large scale climate oscillations. Remote Sens. Environ. 114, 2286–2296.

Carrão, H., Gonçalves, P., Caetano, M., 2008. Contribution of multispectral and multi-temporal information from MODIS images to land cover classification. Remote Sens. Environ. 112, 986–997.

CDWR, 2009. California water plan, update 2009. In: California Department of Water Resources, Bulletin, pp. 160–169.

Chapelle, O., Haffner, P., Vapnik, V.N., 1999. Support vector machines for histogram-based image classification. IEEE Trans. Neural Netw. 10, 1055–1064.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Chen, J., Jonsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. Remote Sens. Environ. 91, 332–344.

Chen, X., Xiang, S., Liu, C.-., Pan, C.-., 2014. Vehicle detection in satellite images by hybrid deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. 11, 1797–1801.

Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y., 2014. Deep learning-based classification of hyperspectral data. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 7, 2094–2107.

Clancey, W.J., 1983. The epistemology of a rule-based expert system—a framework for explanation. Artif. Intell. 20, 215–251.

Connor, J.T., Martin, R.D., Atlas, L.E., 1994. Recurrent neural networks and robust time series prediction. IEEE Trans. Neural Netw. 5, 240–254.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Dannenberg, M.P., Song, C., Hwang, T., Wise, E.K., 2015. Empirical evidence of El Niño–Southern Oscillation influence on land surface phenology and productivity in the western United States. Remote Sens. Environ. 159, 167–180.

Dong, J., Xiao, X., Kou, W., Qin, Y., Zhang, G., Li, L., et al., 2015. Tracking the dynamics of paddy rice planting area in 1986–2010 through time series Landsat images and phenology-based algorithms. Remote Sens. Environ. 160, 99–113.

Evans, J.P., Geerken, R., 2006. Classifying rangeland vegetation type and coverage using a Fourier component based similarity measure. Remote Sens. Environ. 105, 1–8.

Fernandez-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15, 3133–3181.

Fisher, J.I., Mustard, J.F., 2007. Cross-scalar satellite phenology from ground, Landsat, and MODIS data. Remote Sens. Environ. 109, 261–273.

Fisher, J.I., Mustard, J.F., Vadeboncoeur, M.A., 2006. Green leaf phenology at Landsat resolution: scaling from the field to the satellite. Remote Sens. Environ. 100, 265–279.

Friedl, M.A., Brodley, C.E., Strahler, A.H., 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. IEEE Trans. Geosci. Remote Sens. 37, 969–977.

Funk, C., Budde, M.E., 2009. Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. Remote Sens. Environ. 113, 115–125.

Galford, G.L., Mustard, J.F., Melillo, J., Gendrin, A., Cerri, C.C., Cerri, C.E.P., 2008. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. Remote Sens. Environ. 112, 576–587.

Geerken, R.A., 2009. An algorithm to classify and monitor seasonal variations in vegetation phenologies and their inter-annual change. ISPRS J. Photogramm. Remote Sens. 64, 422–431.

Gonsamo, A., Chen, J.M., 2016. Circumpolar vegetation dynamics product for global change study. Remote Sens. Environ. 182, 13–26.

Guidici, D., Clark, M.L., 2017. One-dimensional convolutional neural network land-cover classification of multi-seasonal hyperspectral imagery in the San Francisco Bay Area, California. Remote Sens. 9, 629.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.

Hu, W., Huang, Y., Wei, L., Zhang, F., Li, H., 2015. Deep convolutional neural networks for hyperspectral image classification. J. Sens. 2015.

Hu, F., Xia, G., Hu, J., Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sens. 7, 14680–14707.

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.

Jia, X., Khandelwal, A., Nayak, G., Gerber, J., Carlson, K., West, P., et al., 2017. Incremental dual-memory LSTM in land cover prediction. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 867–876.

Jönsson, P., Eklundh, L., 2004. TIMESAT—a program for analyzing time-series of satellite sensor data. Comput. Geosci. 30, 833–845.

Kampffmeyer, M., Salberg, A., Jenssen, R., 2016. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional. Neural Netw. 1–9.

Kandasamy, S., Baret, F., Verger, A., Neveux, P., Weiss, M., 2013. A comparison of methods for smoothing and gap filling time series of remote sensing observations–application to MODIS LAI products. Biogeosciences 10, 4055–4071.

Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv preprint. arXiv:1412.6980.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Proces. Syst. 2, 1097–1105.

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, 778–782.

Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). Remote Sens. Environ. 100, 356–362.

LeCun, Y., Bengio, Y., 1995. Convolutional networks for images, speech, and time-series. In: Anonymous (Ed.), The Handbook of Brain Theory and Neural Networks.

Li, W., Fu, H., Yu, L., Gong, P., Feng, D., Li, C., et al., 2016. Stacked autoencoder-based deep learning for remote-sensing image classification: a case study of African land-cover mapping. Int. J. Remote Sens. 37, 5632–5646.

Li, Y., Zhang, H., Shen, Q., 2017. Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. Remote Sens. 9, 67.

Li, W., Fu, H., Yu, L., Cracknell, A., 2017. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. Remote Sens. 9.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.

Lloyd, D., 1990. A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery. Int. J. Remote Sens. 11, 2269–2279.

Lyu, H., Lu, H., Mou, L., 2016. Learning a transferable change rule from a recurrent neural network for land cover change detection. Remote Sens. 8, 506.

Lyu, H., Lu, H., Mou, L., Li, W., Wright, J., Li, X., et al., 2018. Long-term annual mapping of four cities on different continents by applying a deep information learning method to Landsat data. Remote Sens. 10, 471.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. High-resolution aerial image labeling with convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 55, 7092–7103.

Marcos, D., Volpi, M., Kellenberger, B., Tuia, D., 2018. Land cover mapping at very high resolution with rotation equivariant CNNs: towards small yet accurate models. ISPRS J. Photogramm. Remote Sens. 145, 96–107.

Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U., 2018. Classification with an edge: improving semantic image segmentation with boundary detection. ISPRS J. Photogramm. Remote Sens. 135, 158–172.

Mou, L., Zhu, X.X., 2018. RiFCN: recurrent network in fully convolutional network for semantic segmentation of high resolution remote sensing images. arXiv preprint. arXiv:1805.02091.

Mou, L., Ghamisi, P., Zhu, X.X., 2018a. Unsupervised spectral-spatial feature learning via deep residual conv-deconv network for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens.

Mou, L., Bruzzone, L., Zhu, X.X., 2018b. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. arXiv preprint. arXiv:1803.02642.

Na, X., Zhang, S., Li, X., Yu, H., Liu, C., 2010. Improved land cover mapping using random forests combined with Landsat thematic mapper imagery and ancillary geographic data. Photogramm. Eng. Remote. Sens. 76, 833–840.

Olsson, L., Eklundh, L., 1994. Fourier-series for analysis of temporal sequences of satellite sensor imagery. Int. J. Remote Sens. 15, 3735–3741.

Penatti, O.A.B., Nogueira, K., Dos Santos, J.A., 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2015-October, pp. 44–51.

Reed, B.C., Brown, J.F., Vanderzee, D., Loveland, T.R., Merchant, J.W., Ohlen, D.O., 1994. Measuring phenological variability from satellite imagery. J. Veg. Sci. 5, 703–714.

Ripley, B.D., 2007. Pattern Recognition and Neural Networks. Cambridge university press.

Roerink, G.J., Menenti, M., Verhoef, W., 2000. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. Int. J. Remote Sens. 21, 1911–1917.

Rogan, J., Franklin, J., Roberts, D.A., 2002. A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. Remote Sens. Environ. 80, 143–156.

Rußwurm, M., Körner, M., 2017. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

Rußwurm, M., Körner, M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. ISPRS Int. J. Geoinf. 7, 129.

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., Ohno, H., 2005. A crop phenology detection method using time-series MODIS data. Remote Sens. Environ. 96, 366–374.

Sakamoto, T., Van Nguyen, N., Ohno, H., Ishitsuka, N., Yokozawa, M., 2006. Spatio–temporal distribution of rice phenology and cropping systems in the Mekong Delta with special reference to the seasonal water flow of the Mekong and Bassac rivers. Remote Sens. Environ. 100, 1–16.

Sakamoto, T., Wardlow, B.D., Gitelson, A.A., Verma, S.B., Suyker, A.E., Arkebauer, T.J., 2010. A two-step filtering approach for detecting maize and soybean phenology with time-series MODIS data. Remote Sens. Environ. 114, 2146–2159.

Sakamoto, T., Wardlow, B.D., Gitelson, A.A., 2011. Detecting spatiotemporal changes of corn developmental stages in the U.S. Corn Belt using MODIS WDRVI data. IEEE Trans. Geosci. Remote Sens. 49, 1926–1936.

Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2013. MODIS-based corn grain yield estimation model incorporating crop phenology information. Remote Sens. Environ. 131, 215–231.

Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2014. Near real-time prediction of U.S. corn yields based on time-series MODIS data. Remote Sens. Environ. 147, 219–231.

Shao, Y., Lunetta, R.S., Wheeler, B., Iiames, J.S., Campbell, J.B., 2016. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. Remote Sens. Environ. 174, 258–265.

Sherrah, J., 2016. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. arXiv preprint. arXiv:1606.02585.

Shi, D., Yang, X., 2016. An assessment of algorithmic parameters affecting image classification accuracy by random forests. Photogramm. Eng. Remote. Sens. 82, 407–417.

Siachalou, S., Mallinis, G., Tsakiri-Strati, M., 2015. A hidden Markov models approach for crop classification: linking crop phenology to time series of multi-sensor remote sensing data. Remote Sens. 7.

Simonneaux, V., Francois, P., 2003. Identifying main crop classes in an irrigated area using high resolution image time series. In: Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International. vol. 1. pp. 252–254.

Simonneaux, V., Duchemin, B., Helson, D., Er-Raki, S., Olioso, A., Chehbouni, A.G., 2008. The use of high-resolution image time series for crop classification and evapotranspiration estimate over an irrigated area in central Morocco. Int. J. Remote Sens. 29, 95–116.

Soudani, K., le Maire, G., Dufrene, E., Francois, C., Delpierre, N., Ulrich, E., et al., 2008. Evaluation of the onset of green-up in temperate deciduous broadleaf forests derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. Remote Sens. Environ. 112, 2643–2655.

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

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al., 2015. Going deeper with convolutions. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 1–9.

Verhoef, W., 1996. Application of harmonic analysis of NDVI time series (HANTS). In: Fourier Analysis of Temporal NDVI in the Southern African and American Continents. 108. pp. 19–24.

Verhoef, W., Menenti, M., Azzali, S., 1996. A colour composite of NOAA-AVHRR-NDVI based on time series analysis (1981–1992). Int. J. Remote Sens. 17, 231–235.

Vicente-Guijalba, F., Martinez-Marin, T., Lopez-Sanchez, J.M., 2014. Crop phenology estimation using a multitemporal model and a Kalman filtering strategy. IEEE Geosci. Remote Sens. Lett. 11, 1081–1085.

Vina, A., Gitelson, A.A., Rundquist, D.C., Keydan, G., Leavitt, B., Schepers, J., 2004. Remote sensing - monitoring maize (*Zea mays* L.) phenology with remote sensing. Agron. J. 96, 1139–1147.

Volpi, M., Tuia, D., 2017. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 55, 881–893.

Walker, J.J., de Beurs, K.M., Wynne, R.H., 2014. Dryland vegetation phenology across an elevation gradient in Arizona, USA, investigated with fused MODIS and Landsat data. Remote Sens. Environ. 144, 85–97.

Walker, J.J., de Beurs, K.M., Henebry, G.M., 2015. Land surface phenology along urban to rural gradients in the U.S. Great Plains. Remote Sens. Environ. 165, 42–52.

Wan, X., Zhao, C., Wang, Y., Liu, W., 2017. Stacked sparse autoencoder in hyperspectral data classification using spectral-spatial, higher order statistics and multifractal spectrum features. Infrared Phys. Technol. 86, 77–89.

Wang, H., Chen, J., Wu, Z., Lin, H., 2012. Rice heading date retrieval based on multi-temporal MODIS data and polynomial fitting. Int. J. Remote Sens. 33, 1905–1916.

Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the US Central Great Plains. Remote Sens. Environ. 112, 1096–1116.

WWW1 U.S. Geological Survey Science (USGS) Land Satellites Data System (LSDS) research and development. https://espa.cr.usgs.gov/, Accessed date: 22 October 2018.

WWW2 Keras: the Python deep learning library. https://keras.io/, Accessed date: 22 October 2018.

WWW3 Tensorflow: an open source software library for high performance numerical computation. https://www.tensorflow.org, Accessed date: 22 October 2018.

WWW4 Scikit-learn: machine learning in Python. http://scikit-learn.org/, Accessed date: 22 October 2018.

WWW5 XGBoost documentation, Python API reference. https://xgboost.readthedocs.io/en/latest/python/python_api.html, Accessed date: 22 October 2018.

Xiao, X., Boles, S., Liu, J., Zhuang, D., Frolking, S., Li, C., et al., 2005. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. Remote Sens. Environ. 95, 480–492.

Xiao, X., Boles, S., Frolking, S., Li, C., Babu, J.Y., Salas, W., et al., 2006. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. Remote Sens. Environ. 100, 95–113.

Xie, Y., Sha, Z., Yu, M., 2008. Remote sensing imagery in vegetation mapping: a review. J. Plant Ecol. 1, 9–23.

Xin, Q., Broich, M., Zhu, P., Gong, P., 2015. Modeling grassland spring onset across the Western United States using climate variables and MODIS-derived phenology metrics. Remote Sens. Environ. 161, 63–77.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., Lipson, H., 2015. Understanding neural networks through deep visualization. arXiv preprint. arXiv:1506.06579.

Yue, J., Zhao, W., Mao, S., Liu, H., 2015. Spectral-spatial classification of hyperspectral images using deep convolutional neural networks. Remote Sens. Lett. 6, 468–477.

Zaremba, W., Sutskever, I., Vinyals, O., 2014. Recurrent neural network regularization. arXiv preprint. arXiv:1409.2329.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: European Conference on Computer Vision, pp. 818–833.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., et al., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84, 471–475.

Zhang, M., Zhou, Q., Chen, Z., Liu, J., Zhou, Y., Cai, C., 2008. Crop discrimination in Northern China with double cropping systems using Fourier analysis of time-series MODIS data. Int. J. Appl. Earth Obs. Geoinf. 10, 476–485.

Zhang, J., Feng, L., Yao, F., 2014. Improved maize cultivated area estimation over a large scale combining MODIS–EVI time series data and crop phenological information. ISPRS J. Photogramm. Remote Sens. 94, 102–113.

Zhang, G., Xiao, X., Dong, J., Kou, W., Jin, C., Qin, Y., et al., 2015. Mapping paddy rice planting areas through time series analysis of MODIS land surface temperature and vegetation index data. ISPRS J. Photogramm. Remote Sens. 106, 157–171.

Zhao, W., Du, S., 2016. Spectral–spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach. IEEE Trans. Geosci. Remote Sens. 54, 4544–4554.

Zhong, L., Hawkins, T., Biging, G., Gong, P., 2011. A phenology-based approach to map crop types in the San Joaquin Valley, California. Int. J. Remote Sens. 32, 7777–7804.

Zhong, L., Gong, P., Biging, G.S., 2012. Phenology-based crop classification algorithm and its implications on agricultural water use assessments in California's central valley. Photogramm. Eng. Remote. Sens. 78, 799–813.

Zhong, L., Gong, P., Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: a multi-year experiment using Landsat imagery. Remote Sens. Environ. 140, 1–13.

Zhong, L., Hu, L., Yu, L., Gong, P., Biging, G.S., 2016a. Automated mapping of soybean and corn using phenology. ISPRS J. Photogramm. Remote Sens. 119, 151–164.

Zhong, L., Yu, L., Li, X., Hu, L., Gong, P., 2016b. Rapid corn and soybean mapping in US Corn Belt and neighboring areas. Sci. Rep. 6, 36240.

Zhu, X.X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., et al., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5, 8–36.

Zou, Q., Ni, L., Zhang, T., Wang, Q., 2015. Deep learning based feature selection for remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. 12, 2321–2325.