Remote Sensing of Environment 265 (2021) 112679

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Deep machine learning with Sentinel satellite data to map paddy rice production stages across West Java, Indonesia

K.R. Thorp [a,*], D. Drajat [b]

[a] USDA-ARS, U.S. Arid Land Agricultural Research Center, 21881 N Cardon Ln, Maricopa, AZ 85138, United States
[b] Badan Pusat Statistik (Statistics Indonesia), Jakarta, Indonesia

ARTICLE INFO

Edited by xxx

Keywords:
Convolutional neural network
Crop production survey
Google Earth Engine
Long short term memory
Recurrent neural network
Synthetic aperture radar
TensorFlow
Vegetation indices

ABSTRACT

Indonesia recently implemented a novel, technology-driven approach for conducting agricultural production surveys, which involves monthly observations at many thousands of strategic locations and automated data logging via a cellular phone application. Data from these comprehensive field surveys offer immense value for advancing remote sensing technology to map crop production across Indonesia, particularly through the development of machine learning approaches to relate survey data with satellite imagery. The objective of this study was to compare different machine learning scenarios for classifying and mapping the temporal progression of paddy rice production stages across West Java, Indonesia using synthetic aperture radar (SAR) and optical remote sensing data from Sentinel-1 and Sentinel-2 satellites. Monthly paddy rice survey data at 21,696 locations across West Java from November 2018 through April 2019 were used for model training and testing. Five classes related to rice production stage or other field conditions were defined, including rice at tillering, heading, and harvest stages, rice fields with little to no vegetation present, and non-rice areas. A recurrent neural network (RNN) with long short term memory (LSTM) nodes provided optimal performance with classification accuracies of 79.6% and 75.9% for model training and testing, respectively, and reduced computational effort. Other approaches that incorporated a convolutional neural network (CNN) either reduced classification accuracy or increased computational effort. Deep machine learning methods (RNN and CNN) generally outperformed other non-deep classifiers, which achieved up to 63.3% accuracy for model testing. Classification accuracies were optimized by inputting two Sentinel-1 channels (VH and VV polarizations) and ten Sentinel-2 channels. Temporal patterns of paddy rice production stages were consistent between the monthly ground-based agricultural survey data and 10-m, satellite-based rice classification maps obtained by applying the LSTM-based RNN across West Java. The results demonstrated the value of combining modern agricultural survey data, satellite remote sensing, and a recurrent neural network to develop multitemporal maps of paddy rice production stages.

1. Introduction

National crop production statistics are typically computed from estimates of crop yield and harvest area, based on data obtained from agricultural censuses, farmer and land surveys, and remote sensing (Dong and Xiao, 2016; Kuenzer and Knauer, 2013; Mosleh et al., 2015). Since 1973, production statistics for Indonesian paddy rice (*Oryza sativa* L.) have been based on administrative reporting systems, where crop production data are collected by the lowest governmental unit and then aggregated throughout the governmental hierarchy to provide national-level production estimates. While this approach is inexpensive and provides timely data at a fine disaggregation level, it is also prone to large errors due to subjectivity of field observations (Rotairo et al., 2019). In 2018, Indonesia implemented a novel, breakthrough approach for quantifying paddy rice production. An agricultural survey approach, which combined an area sampling frame and extensive satellite-based maps of topography, land cover, and administrative boundaries, was developed to statistically identify more than 200,000 sampling locations for ground-based agricultural surveys. At the end of each month, Indonesia now deploys 4778 crop surveyors and 1608 supervisors to collect data and classify conditions at these sampling locations, utilizing an Android cellular phone application to log georeferenced information and photographs. The new approach has improved estimates of paddy rice production in Indonesia, although labor costs are high.

* Corresponding author.
E-mail address: kelly.thorp@usda.gov (K.R. Thorp).

https://doi.org/10.1016/j.rse.2021.112679
Received 24 February 2021; Received in revised form 17 August 2021; Accepted 25 August 2021
Available online 2 September 2021
0034-4257/Published by Elsevier Inc.

Furthermore, the extensive data sets compiled from these surveys offer great opportunity for novel remote sensing studies that develop machine learning methods for estimating rice production stages from satellite data, mapping rice production nationally, and forecasting harvest events for improved rice supply chain planning. As irrigated paddy rice is Indonesia's predominant cropping system (Panuju et al., 2013), improved paddy rice mapping and monitoring techniques are crucial to support policy formulation for ensuring food security.

Satellite remote sensing has been developed to map paddy rice production for several decades, primarily in Asian countries (Dong et al., 2016; Kuenzer and Knauer, 2013; Manjunath et al., 2015; Nelson et al., 2014; Xiao et al., 2006) but also in the Mediterranean region (Nguyen and Wagner, 2017) and Australia (McCloy et al., 1987). The agronomic practices of field flooding and rice transplantation, followed by the crop development cycle, provides a unique temporal reflectance signal and mechanism for differentiation of rice fields from other land uses (Kuenzer and Knauer, 2013; Xiao et al., 2002). Many rice mapping studies have used optical remote sensing products from the Landsat program or the Moderate Resolution Imaging Spectrometer (MODIS) and have employed diverse methodologies for temporal classifications of spectral reflectance data or spectral indices, such as the Normalized Difference Vegetation Index (NDVI), the Enhanced Vegetation Index (EVI), the Land Surface Water Index (LSWI), and others (Boschetti et al., 2017; Clauss et al., 2016; Dong et al., 2015; Kontgis et al., 2015; Okamoto and Fukuhara, 1996; Sakamoto et al., 2018; Shi and Huang, 2015; Son et al., 2013; Tennakoon et al., 1992; Xiao et al., 2005; Yin et al., 2019; Zhang et al., 2015). In Indonesia, methods using MODIS spectral indices have been successful for assessing the temporal dynamics of paddy rice production on Java Island (Setiawan et al., 2014, 2016) and for providing estimates of paddy rice extent that had strong correlation ($r^2 > 0.92$) with national statistics (Bridhikitti and Overcamp, 2012; Lee et al., 2012; Nuarsa et al., 2012). However, these studies were conducted prior to availability of modern paddy rice survey data in Indonesia, and novel studies that use these data sets are therefore needed. In general, many studies have concluded that satellite optical remote sensing methods can be used to identify the paddy rice production cycle and map production areas at regional scales. A main complicating factor is reduced optical image quality from frequent and persistent cloud cover over many paddy rice production regions (Nelson et al., 2014), made worse by the relatively low temporal frequency (16 days) of Landsat image collections (Yin et al., 2019). Furthermore, paddy rice fields are often too small and fragmented to be adequately resolved using MODIS data, which have a minimum spatial resolution of 250 m. However, with the launch of the two Sentinel-2 satellites by the European Union's Earth observation program (named “Copernicus”) in 2015 and 2017, free and open-access optical remote sensing data (13 channels in the visible, near-infrared, and short-wave infrared wavelengths) are now available globally with 5-day revisit frequency and spatial resolutions of 10, 20, or 60 m (depending on the band). Therefore, additional studies are needed to evaluate this modern satellite system for mapping paddy rice production.

Other studies have used synthetic aperture radar (SAR) satellite data for paddy rice mapping (Choudhury and Chakraborty, 2006; Clauss et al., 2018; Nguyen and Wagner, 2017). In West Java, Indonesia, Ribbes and Le Toan (1999) mapped paddy rice based on temporal changes in SAR data from RADARSAT, and the mapping result was 87% accurate as compared to an available land cover map. Because SAR sensors involve longer wavelengths than optical sensors, the SAR signals are able to penetrate clouds and provide useful data even during overcast conditions (Bazzi et al., 2019; Bouvet and Le Toan, 2011; Nelson et al., 2014; Nguyen et al., 2016). Furthermore, because SAR backscatter is sensitive to surface roughness characteristics, the agronomic practice of flooding paddy rice fields provides a substantially reduced backscatter signal prior to rice transplantation, which then increases with growth of rice biomass and leaf area index (Inoue et al., 2002; Kuenzer and Knauer, 2013; Ribbes and Le Toan, 1999). Several recent paddy rice mapping studies were based on SAR data from the two Sentinel-1 satellites (Bazzi et al., 2019; Clauss et al., 2018; Nguyen et al., 2016; Nguyen and Wagner, 2017), which were launched by the European Space Agency (ESA) in 2014 and 2016 and provide free and open-access C-band SAR data globally with a 6-day revisit frequency and spatial resolutions between 5 and 40 m (depending on collection mode). Among several polarization options offered by Sentinel-1, the VH polarization (cross-polarized) was more sensitive to the rice production cycle as compared to the VV polarization (co-polarized) (Nguyen et al., 2016). Recent studies have also mapped paddy rice by combining SAR data from Sentinel-1, RADARSAT, or PALSAR satellites with optical remote sensing data from Landsat or Sentinel-2 satellites (Park et al., 2018; Torbick et al., 2017; Zhang et al., 2018b). With only a couple exceptions (Nelson et al., 2014; Ribbes and Le Toan, 1999), few studies have tested SAR satellite data for paddy rice mapping in Indonesia, and additional studies that combine Sentinel-1 SAR data and Sentinel-2 optical data for rice mapping are needed.

Algorithms for mapping paddy rice have often focused on the temporal responses in SAR or optical remote sensing data over the rice growing season. In many studies, paddy rice identification was based on development of simple decision rules or thresholds for temporal remote sensing indices (Boschetti et al., 2017; Bridhikitti and Overcamp, 2012; Kontgis et al., 2015; Lee et al., 2012; Ribbes and Le Toan, 1999; Setiawan et al., 2016). Other studies have employed machine learning methods to identify temporal paddy rice profiles, including *k*-means clustering (Setiawan et al., 2014), ISODATA classifiers (Manjunath et al., 2015; Nguyen et al., 2012), random forest classifiers (Bazzi et al., 2019; Torbick et al., 2017; Zhang et al., 2018b), support vector machines (Park et al., 2018; Zhang et al., 2018b), and artificial neural networks (Chen and Mcnairn, 2006; Ndikumana et al., 2018; Shao et al., 2001; Zhang et al., 2018a). Limited studies based on neural networks, also known as deep learning methods, have shown improved performance and great promise. For example, Zhang et al. (2018a) found that a convolutional neural network (CNN) could identify paddy rice in temporal Landsat NDVI with classification accuracy of 97%, which was 6% and 8% greater than results for support vector machine and random forest classifiers. Also, Ndikumana et al. (2018) demonstrated that a recurrent neural network (RNN) could identify paddy rice in Sentinel-1 SAR data with accuracy of 96%, which was 10% greater than that from classical machine learning approaches. Zhu et al. (2021) combined an RNN with rice development characteristics to identify paddy rice areas with accuracy between 89% and 94%. Because an RNN specifically models the temporal dimension of data, it is particularly suited for temporal remote sensing analyses, such as that required for identification, mapping, and forecasting of paddy rice production stages. Current studies typically report classification accuracy for rice versus other crops or non-rice areas, and few report classification accuracy for unique rice production stages, likely due to limitations of reference data sets for model training. Thus, the rice production stage information obtained from the Indonesian crop surveys can facilitate novel mapping and forecasting of rice production stage progression using CNN and RNN modeling.

The goal of this study was to use deep neural network methods with SAR and optical remote sensing data from Sentinel-1 and Sentinel-2 to map paddy rice production stages across West Java, Indonesia. Model training and testing were based on monthly ground-based crop survey data at 21,696 locations across West Java from November 2018 through April 2019. The specific objectives of the study were to 1) compare different implementations of deep neural networks and other non-deep classifiers to relate multitemporal Sentinel data to monthly observations of rice production stage or other field conditions, 2) contrast the performance of different Sentinel-1 and Sentinel-2 channels for use in the modeling framework, and 3) extend the highest-performing model across West Java to compare rice production statistics obtained from ground-based crop surveys with that from satellite remote sensing data.

2. Materials and methods

2.1. Region of study

The study region was the province of West Java, Indonesia (6.75°S, 107.5°E), which covers an area of 35,378 km^2 and contains 26 administrative subdivisions (Fig. 1). Paddy rice production in West Java occurs in two main areas: an irrigated region in the north and a rainfed region in the south. The main cropping season begins in October or November and ends in March or April. In the irrigated region to the north, there is a second cropping season that begins in March or April and ends in August or September. Field sizes are typically less than 1.0 ha.

2.2. Crop surveys

Ground-based information on paddy rice stage and other land use conditions was collected by Badan Pusat Statistik (BPS or Statistics Indonesia), which is the institution of the Indonesian government that is responsible for conducting statistical surveys for agriculture, economics, and demographics. In 2018, BPS initiated its novel methodology for agricultural surveying, and an area sampling frame was used to statistically identify 23,418 surveying locations across West Java (Fig. 1). In the final week of each month, crop surveyors navigated to each sampling location and classified the condition as belonging to one of nine possible classes related to paddy rice production stage or alternative land use conditions (Table 1). Georeferenced surveying information, other metadata, and photographic evidence were logged with an Android cellular phone application and transferred to a centralized BPS server, which collected surveying results from West Java and from other provinces nationwide. The Android application permitted data entry only when the phone's geoposition was within a 10-m radius of the survey location, and data entries were linked with the surveyor's username for accountability. Supervisors of the crop surveyors verified the accuracy of the survey data using a web-based monitoring system that aggregated data at the district and provincial levels, and they provided feedback on data accuracy and correctness to the surveyors. The present study incorporated six months of surveying results for West Java from November 2018 through April 2019, which was the first complete Indonesian paddy rice growing season surveyed using this new paradigm.

The BPS crop surveyors chose among nine field condition classes (Table 1), each represented by an integer value from 1 to 9. The first three classes pertained to the rice production cycle. Classes 1, 2, and 3, referred to the "first vegetative" or rice transplant stage, the "second vegetative" or rice tillering stage, and the "generative" or rice heading stage, respectively. Class 4 described a field that was recently harvested. Class 5 described a field undergoing tillage or flooding in preparation for rice transplantation. Class 6 described a field that was damaged by pests or another catastrophe. Classes 7, 8, and 9 described fields with crops other than rice, areas with non-agricultural land uses, and fields that were left fallow after rice harvest, respectively. A zero value was used to indicate that no observation was available. The georeferenced survey information for West Java was obtained from BPS in spreadsheet format. Table 1 provides the number of locations (out of 23,418 total locations) assigned to each class on a monthly basis from November 2018 through April 2019.

2.3. Sentinel data

Google Earth Engine (Gorelick et al., 2017) was used to obtain and preprocess Sentinel-1 SAR data and Sentinel-2 optical data for the region of study. Spatial extents for image retrieval were defined by the minimum and maximum geographic coordinates for West Java, Indonesia. Temporal extents were iteratively defined in calendar days as the first half and second half of each month, resulting 12 image retrieval periods over the six-month duration from November 2018 through April 2019 (Table 2). Remote sensing data from the second half of each month represented conditions during the BPS crop surveys. No survey data were available for the first half of each month. Depending on the month, image retrieval periods ranged from 14 to 16 days, which meant at least two satellite revisits were possible during each period.

A script was prepared in the GEE Code Editor to obtain the dual-

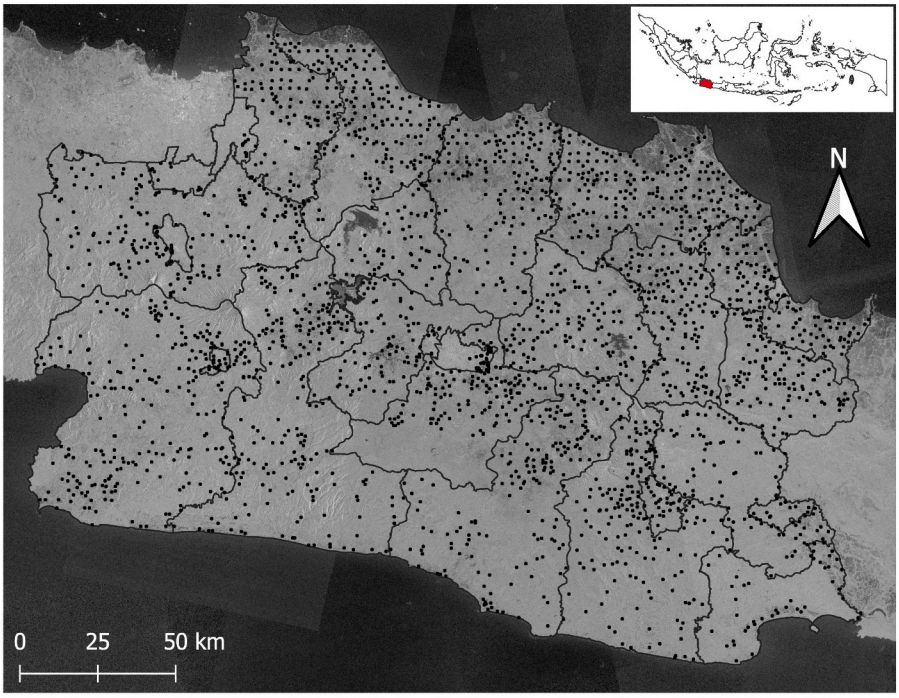

Fig. 1. The study region was the province of West Java, Indonesia (6.75°S, 107.5°E). Beginning in 2018, ground-based agricultural surveys were conducted monthly at 23,418 locations across the province. Each black point represents a pattern of nine survey locations, defined by a 3 × 3, 100-m grid. A composite synthetic aperture radar (SAR) image from Sentinel-1 overpasses during the period from 1 November 2018 through 15 November 2018 is shown.

Table 1
Summary of the field condition classes from monthly field surveys of paddy rice at 23,418 locations across West Java, as recorded by surveyors from Badan Pusat Statistik (BPS or Statistics Indonesia). The number of locations assigned to each class are provided for six months from November 2018 through April 2019. To increase class separability for the data analysis, several classes were merged as indicated.

BPS Class #	Merged With	Class description	Nov 2018	Dec 2018	Jan 2019	Feb 2019	Mar 2019	Apr 2019
0	N/A	No data recorded	1,697	1,647	405	154	33	15
1	5,6,9	First vegetative (transplant)	1,886	5,742	7,427	3,314	2,360	4,496
2	N/A	Second vegetative (tillering)	730	1,055	5,552	5,782	2,376	1,459
3	N/A	Generative (heading)	1,463	932	1,503	6,662	7,370	3,660
4	N/A	Harvest	1,774	1,028	701	793	4,426	4,915
5	1,6,9	Land preparation and flooding	4,769	5,241	1,917	1,180	1,264	2,604
6	1,5,9	Damaged crop	195	109	43	17	19	17
7	8	Crop other than rice	1,665	1,443	1,315	1,253	1,263	1,304
8	7	Non-agricultural land	3,728	3,715	3,859	3,948	3,989	4,018
9	1,5,6	Fallowed after rice harvest	5,511	2,506	696	315	318	930

Table 2
Counts of Sentinel-1 and Sentinel-2 images over West Java from November 2018 through April 2019, accessed through Google Earth Engine. Effects of cloud cover on Sentinel-2 images were quantified in two ways: 1) filtering images with the "CLOUDY_PIXEL_PERCENTAGE" property <20% and 2) computing the percentage of pixels with the QA60 channel indicating cirrus or opaque cloud cover. Ground-based paddy rice survey data were available only for the latter time period of each month.

Time Period	Start Date	End Date	Survey Data?	Sentinel-1 Count	Sentinel-2 Count	<20% Cloud S2 Count	Cloud Pixels
1	1 Nov 2018	15 Nov 2018	No	9	51	5	43.6%
2	16 Nov 2018	30 Nov 2018	Yes	11	51	12	18.7%
3	1 Dec 2018	16 Dec 2018	No	11	43	0	65.7%
4	16 Dec 2018	31 Dec 2018	Yes	15	51	1	36.6%
5	1 Jan 2019	16 Jan 2019	No	12	53	6	15.3%
6	16 Jan 2019	31 Jan 2019	Yes	9	51	0	66.7%
7	1 Feb 2019	14 Feb 2019	No	9	42	6	24.6%
8	15 Feb 2019	28 Feb 2019	Yes	9	44	13	14.0%
9	1 Mar 2019	16 Mar 2019	No	11	51	1	37.3%
10	16 Mar 2019	31 Mar 2019	Yes	9	54	9	9.5%
11	1 Apr 2019	15 Apr 2019	No	11	46	10	21.4%
12	16 Apr 2019	30 Apr 2019	Yes	12	42	9	5.2%

polarized, C-band, ground range detected (GRD) SAR product from both Sentinel-1 satellites in the Interferometric Wide (IW) swath mode, which were preprocessed by GEE for thermal noise removal, radiometric calibration, and terrain correction using algorithms from the Sentinel-1 Toolbox. Data were obtained for both the VV polarization (single co-polarization, vertical transmit and vertical receive) and the VH polarization (dual-band cross-polarization, vertical transmit and horizontal receive). Based on Mladenova et al. (2013), a cosine correction was applied to normalize the SAR data for incidence angle. A refined Lee speckle filter was then used for despeckling. The algorithm was adapted from a shared GEE script by Guido Lemoine, who translated the refined Lee speckle filter from the Sentinel-1 Toolbox to the GEE scripting language. To reduce the SAR data from GEE "Image Collection" format and develop composite images of the study region, the median value of co-located pixels was computed. Based on a histogram analysis, backscatter from composite images was bounded from -40.0 to 5.0 dB, rescaled from 0 to 255, and exported from GEE as 8-bit, unsigned integer images with 10-m spatial resolution. A total of 24 composite SAR images were exported, including one image for each of 12 time periods for both the VV and VH polarizations.

A separate GEE script was prepared in the GEE Code Editor to obtain and preprocess Level-1C, top of the atmosphere, optical remote sensing data from both Sentinel-2 satellites. Two methods were tested to filter the data based on cloud cover conditions (Table 2). Filtering based on the "CLOUDY_PIXEL_PERCENTAGE" metadata value of <20% was too restrictive for conditions in West Java. This method removed entire image tiles with cloud cover >20%, and relatively few optical images could be obtained under this condition (Table 2). Instead, a cloud mask algorithm was developed, based on examples from other shared GEE scripts. The algorithm assessed bits 10 and 11 of the Sentinel-2 QA60 band, which provided information on coverage of opaque and cirrus clouds, respectively. Pixels with either of these cloud cover issues were masked from the optical images, which represented 5% to 67% of the area of West Java depending on the time period (Table 2). Composite optical images were developed similarly to the SAR images, by computing the median value of co-located pixels. Reflectance data from composite images were bounded from 0.0 to 1.0, rescaled from 0 to 255, and exported from GEE as 8-bit, unsigned integer images with 10-m spatial resolution. Optical data from 10 Sentinel-2 channels were exported from GEE, representing visible blue, green, and red wavelengths (B2, 492.4 nm; B3, 559.8 nm; and B4, 664.6 nm; respectively), three red edge wavelengths (B5, 704.1 nm; B6, 740.5 nm; and B7, 782.8 nm), two near-infrared wavelengths (B8, 832.8 nm and B8A, 864.7 nm), and two short-wave infrared wavelengths (B11, 1613.7 nm and B12, 2202.4 nm). A total of 120 composite optical images were exported, including one image for each of 12 time periods for each of 10 wavebands.

Following export from GEE, a Python script (https:// www.python.com) was developed using the Geospatial Data Abstraction Library (GDAL, https://www.gdal.org) to extract image data at the location of each georeferenced crop surveying point from BPS. The point-based data was used for input into RNN models. Additionally, data in a 10×10-pixel patch was extracted from each location for use in the CNN models. The RNN methods required time-series data at single points, while the CNN methods required time-series data from patches of image pixels.

2.4. Vegetation indices

Because many past studies have used vegetation indices (e.g., NDVI, EVI, and LSWI) for paddy rice mapping (Clauss et al., 2016; Dong et al., 2015; Kontgis et al., 2015; Son et al., 2013; Xiao et al., 2005; Yin et al., 2019; Zhang et al., 2018b), the performance of these indices were

evaluated herein. Prior to input to machine learning models, the NDVI, EVI, and LSWI were computed as follows:

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{Red}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{Red}}} \quad (1)$$

$$\mathrm{EVI} = 2.5\left(\frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{Red}}}{\rho_{\mathrm{NIR}} + 6\rho_{\mathrm{Red}} - 7.5\rho_{\mathrm{Blue}} + 1}\right) \quad (2)$$

$$\mathrm{LSWI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{SWIR}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{SWIR}}} \quad (3)$$

where ρ_{Blue}, ρ_{Red}, ρ_{NIR}, and ρ_{SWIR} were the reflectance values from the blue (B2), red (B4), near-infrared (B8), and short-wave infrared (B11) channels of Sentinel-2, respectively.

2.5. Machine learning

Because RNN methods are well-suited and designed for time-series modeling, an RNN strategy was deemed appropriate for modeling the temporal progression of rice production stages in West Java. Two RNN strategies were compared in this study, one based on traditional long short term memory (LSTM) nodes (Hochreiter and Schmidhuber, 1997) and another based on the newer gated recurrent unit (GRU) nodes (Cho et al., 2014). The RNN's were structured to relate the multitemporal Sentinel image data to the multitemporal BPS crop survey data from West Java over the six-month timeframe. A Python script was developed, which incorporated Python's "Keras" package as a frontend for the "TensorFlow" machine learning platform (Abadi et al., 2015). The Keras "Sequential" class was used to contain a linear stack of neural network layers. As an example for the LSTM approach (Fig. 2), the LSTM layer was first added to establish the recurrent neural network, which could model the sequence of time-dependent behavior in the data sets. The LSTM layer accepted ($t \times m$)-dimensional input vectors, where t was 12 (i.e., the number of half-month Sentinel imaging time periods) and m was the number of Sentinel image channels included in the model. The LSTM layer was specified to output a ($t \times l$)-dimensional vector, where l was the number of hidden layers within the LSTM-based RNN. The output of the RNN was input to a time-distributed dense neural network layer, which provided a ($t \times n$)-dimensional vector output, where n was the number of BPS survey classes. Finally, because the BPS survey data were available only for the second half of each month, a Lambda layer was added, which called a customized function to subset the results for the latter half of the month, resulting a ($t/2 \times n$)-dimensional output vector. The BPS survey classes, which were originally represented with categorical integer values, were converted to one-hot encoded binary vectors prior to modeling, which is known to improve neural network performance. Furthermore, because the dense layer was specified with "softmax" activation, one-hot encoding enabled the network to output a vector of categorical probabilities for each BPS survey class. The survey class simulated with the greatest probability was used in accuracy assessment calculations. The Keras model was compiled with the "RMSprop" optimizer and a loss function based on categorical cross-entropy among the observed and simulated one-hot encoded BPS class vectors. In this way, an LSTM-based RNN network was established to relate 12 half-monthly time periods of Sentinel imaging data to 6 end-of-month time periods of BPS survey data. The GRU-based RNN network was implemented by changing one line of code to switch the LSTM layer to a GRU layer in the Keras Sequential stack (Table 3).

While the RNN models were appropriate for the temporal nature of the problem, they did not consider image textural information in the area surrounding the BPS survey locations. However, CNN models are well-suited for this task, as they can analyze the spatial content of images using filter convolution. Three implementations of CNN models were tested by modifying the layers in the Keras Sequential stack (Table 3), including 1) a time-distributed CNN layer without an RNN

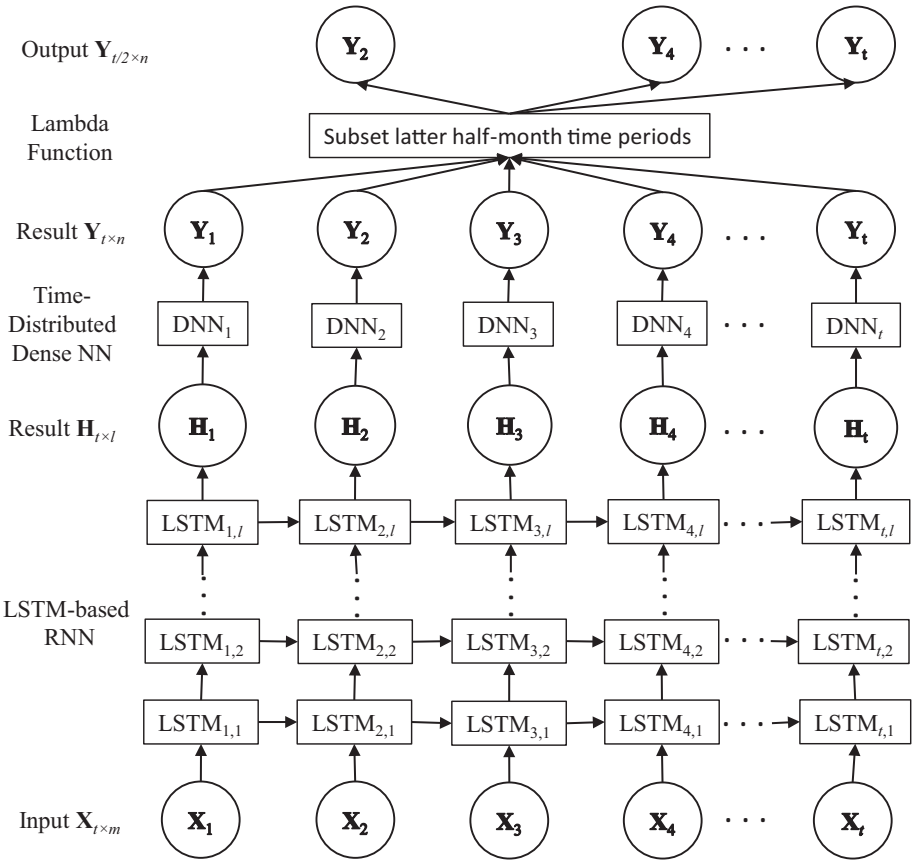

Fig. 2. A recurrent neural network (RNN) with long short term memory (LSTM) nodes received inputs ($\mathbf{X}_{t\times m}$) from m Sentinel image layers over t half-monthly time periods ($t = 12$), resulting intermediate output ($\mathbf{H}_{t\times l}$) from l hidden layers in the RNN. A time-distributed dense neural network (DNN) with "softmax" activation computed probabilities ($\mathbf{Y}_{t\times n}$) for each of n field survey classes related to paddy rice stage or other land uses. A lambda function subsetted results for the latter half of each month for comparison to the observed field survey classes, represented with one-hot encoding.

Table 3
Deep machine learning models based on recurrent and convolutional neural networks (RNN and CNN, respectively) as implemented in the Keras package for Python. Variables include the number of Sentinel image layers ($m = 12$), the number of half-monthly time periods for Sentinel imaging ($t = 12$), the number of RNN hidden layers ($l = 750$), the CNN patch size ($p = 10$), the number of convolutional layers ($c = 32$), and the number of survey classes related to paddy rice production stage or other land uses ($n = 5$). The input shapes were $t \times m$ for the RNN models with long short term memory (LSTM) and gated recurrent unit (GRU) nodes and $t \times p \times p \times m$ for the models with a CNN component (Conv2D, Conv2D LSTM, and ConvLSTM2D).

Model	Keras layers	Output shape
LSTM	LSTM	$t \times l$
	Time-Distributed Dense	$t \times n$
	Lambda	$t/2 \times n$
GRU	GRU	$t \times l$
	Time-Distributed Dense	$t \times n$
	Lambda	$t/2 \times n$
Conv2D	Time-Distributed Conv2D	$t \times (p-2) \times (p-2) \times c$
	Time-Distributed MaxPooling	$t \times (p-2)/2 \times (p-2)/2 \times c$
	Time-Distributed Flatten	$t \times 512$
	Time-Distributed Dense	$t \times n$
	Lambda	$t/2 \times n$
Conv2D LSTM	Time-Distributed Conv2D	$t \times (p-2) \times (p-2) \times c$
	Time-Distributed MaxPooling	$t \times (p-2)/2 \times (p-2)/2 \times c$
	Time-Distributed Flatten	$t \times 512$
	LSTM	$t \times l$
	Time-Distributed Dense	$t \times n$
	Lambda	$t/2 \times n$
ConvLSTM2D	ConvLSTM2D	$t \times (p-2) \times (p-2) \times c$
	Time-Distributed MaxPooling	$t \times (p-2)/2 \times (p-2)/2 \times c$
	Time-Distributed Flatten	$t \times 512$
	Time-Distributed Dense	$t \times n$
	Lambda	$t/2 \times n$

backend (Conv2D), 2) a time-distributed CNN layer with a separate LSTM-based RNN backend (Conv2D LSTM), and 3) the approach of Shi et al. (2015) that nested the CNN spatial filtering within an LSTM-based RNN framework (ConvLSTM2D). All CNN methods were followed by time-distributed max pooling and flattening layers to downsample the filtering output and reduce it to one dimension. The results were fed to a time-distributed dense layer and the Lambda layer, as with the RNN approaches (Fig. 2, Table 3).

Many trials were conducted to establish optimal settings among the parameters related to neural network performance, including the fraction of data used for training versus testing, the number of RNN hidden layers (l), the number of CNN convolutional layers (c), and the batch size and number of epochs used for model training. First, the modeling exercise required splitting the data sets for purposes of model training and independent testing. Trials were conducted to evaluate effects of setting aside every second, fourth, sixth, eighth, and tenth sampling location for model testing purposes. Trials were also conducted to evaluate the impacts of 250, 500, 750, 1000 and 1250 hidden layers in the RNN layer and 8, 16, 32, 64, and 128 convolutional layers in the CNN layer. During model fitting, the batch size established the number of sampling locations evaluated at one time, while the number of epochs specified the number of training iterations. Batch sizes of 4, 8, 16, 32, and 64 were tested, and 5, 10, 20, 40, and 80 epochs were tested. The effects of different parameter settings were evaluated for training accuracy, testing accuracy, and the time duration to both train and test the model. Some parameter configurations could improve accuracy, but with great additional computational expense and processing duration.

The results of the deep machine learning methods (RNN and CNN) were compared with that of other non-deep classifiers, including Gaussian Naive Bayes (GNB), random forest (RF), *k*-nearest neighbors (KNN), multi-layer perceptron (MLP) neural network, and support vector machine (SVM) classifiers. Implementations of these classifiers were obtained from Python's "scikit–learn" package. A main difference between the deep and non-deep learning methods was that the non-deep classifiers were unable to link the data through their temporal dimension. Therefore, remote sensing data from the early half-month period could not be incorporated with the non-deep classifiers, because there was no corresponding ground surveying data. Model training and testing for non-deep classifiers proceeded with half the remote sensing input data compared to the deep learning approaches.

Handling of missing data was important for the analysis. For the BPS survey data, locations with missing observations at any timestep over the six-month period (Table 1) were eliminated from the analysis. This provided a more complete data set for model training, but reduced the number of surveying locations from 23,418 to 21,696 across West Java. Missing data from the Sentinel images was problematic only for the Sentinel-2 optical data and indicated that no cloud-free pixels were available at a given surveying location and timestep (Table 2). These conditions were coded with -1.0 for input to the machine learning models, while available Sentinel data was scaled from 0.0 to 1.0. Thus, the models learned that -1.0 meant no optical data was available at that location and timestep.

Preliminary modeling results suggested that some of the BPS survey classes were more difficult to segregate than others. Many trials were conducted to improve classification accuracy by merging similar classes (Table 1). As discussed later, the efforts reduced the number of BPS classes from nine to five, including a condensed class related to rice production areas with little to no vegetation present (Classes 1, 5, 6, and 9), a condensed class related to areas with vegetation other than rice (Classes 7 and 8), and three unmodified, rice-specific classes for tillering (Class 2), heading (Class 3), and harvest (Class 4) stages. Rice mapping efforts focused on analysis of the the latter three rice production stages.

2.6. Mapping rice production

The LSTM-based RNN was extended to map the five condensed BPS classes from Sentinel image data across the entire area of West Java at 10-m spatial resolution. As with the model development effort (Fig. 2), Sentinel data from 12 half-month timesteps at each pixel location were input to the model, and the probability of membership to each of the five condensed BPS classes at the six latter half-month timesteps was computed for each pixel. Resulting data were plotted as red, green, and blue (RGB) color images where the R, G, and B channels represented the probability of membership to BPS Classes 2, 3, and 4, respectively.

3. Results

3.1. Sentinel data

Tests of individual Sentinel satellite channels revealed their ability to accurately classify among rice production stages or other land use conditions through the modeling framework (Fig. 3(a)). Among the individual Sentinel bands, Sentinel-1 SAR data with the VH polarization performed best with 67.0% accuracy for model training and 67.2% accuracy for model testing. This verified results from other studies showing the value of the VH band for paddy rice mapping (Nguyen et al., 2016). Among the optical channels from Sentinel-2, the two near-infrared bands (B8 at 832.8 nm and B8A at 864.7 nm) and a red edge band (B7 at 782.8 nm) resulted in the greatest classification accuracies, from 63.4% to 65.9% for model training and from 60.6% to 61.5% for model testing. No individual optical channel performed better than the SAR data with VH polarization, perhaps due to missing optical data from cloudy pixels.

Among the three tested vegetation indices, the EVI classified rice production stage and other land use conditions better than any individual Sentinel-2 channel with 71.3% training accuracy and 66.7% testing accuracy, while the NDVI performed similarly with 70.7% training accuracy and 65.7% testing accuracy (Fig. 3(b)). The LSWI resulted in a 63.2% training accuracy and 57.6% testing accuracy. Combining the two Sentinel-1 bands (VV and VH) for input to the model

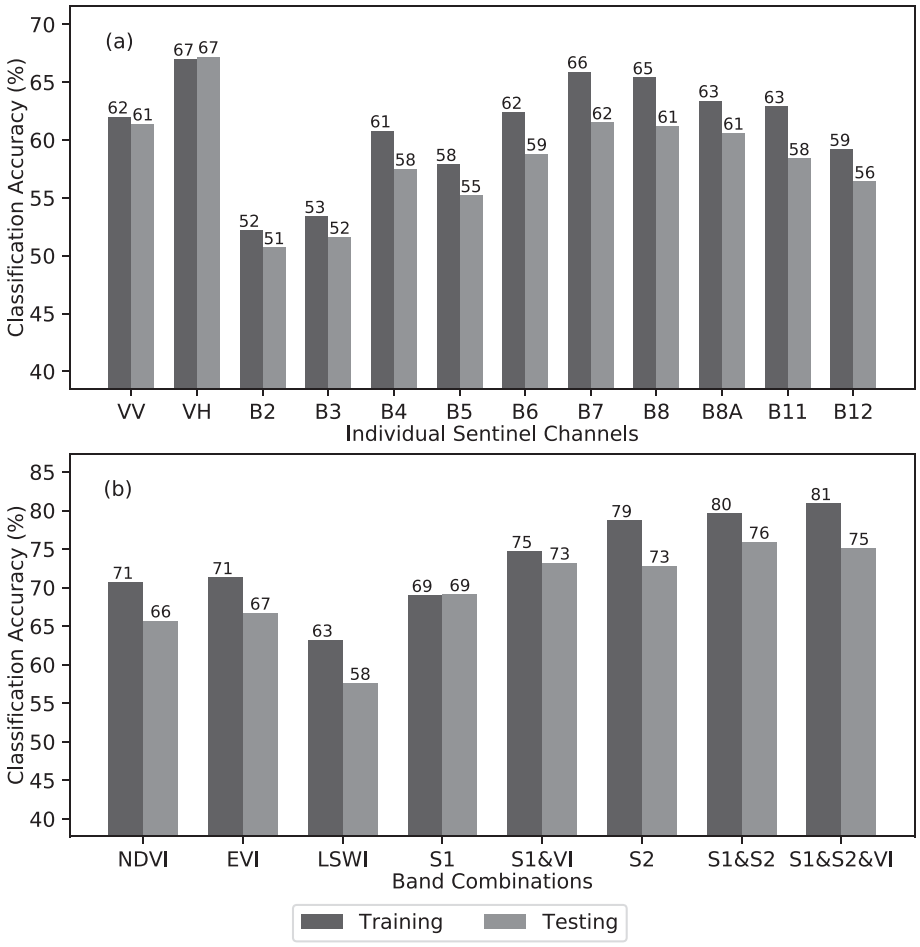

Fig. 3. Accuracy of classifying rice production stage or other land use conditions using (a) individual channels or (b) vegetation indices and various band combinations from Sentinel-1 and Sentinel-2 satellite data using a recurrent neural network based on long short term memory (LSTM) nodes. Sentinel-1 channels (S1) included synthetic aperture radar (SAR) data in the VV and VH polarizations. Sentinel-2 channels (S2) included optical remote sensing data from the following wavelengths: B2 (blue, 492.4 nm), B3 (green, 559.8 nm), B4 (red, 664.6 nm), B5 (red edge, 704.1 nm), B6 (red edge, 740.5 nm), B7 (red edge, 782.8 nm), B8 (near-infrared, 832.8 nm), B8A (narrow near-infrared, 864.7 nm), B11 (short-wave infrared, 1613.7 nm), and B12 (short-wave infrared, 2202.4 nm). Vegetation indices included the Normalized Difference Vegetation Index (NDVI), the Enhanced Vegetation Index (EVI), and the Land Surface Water Index (LSWI).

resulted in training and testing accuracies of 69.0% and 69.1%, which was only slightly improved from tests of the VH channel alone. Combining all ten Sentinel-2 bands (B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12) for input to the model resulted in training and testing accuracies of 78.7% and 72.8%. Thus, while the individual Sentinel-2 channels performed relatively poorly, their combined influence on the model substantially improved classification accuracy in spite of missing data due to cloud cover.

Combining the two Sentinel-1 bands with all ten Sentinel-2 bands improved the classification accuracy to 79.6% for model training and 75.9% for model testing. This band combination provided the greatest testing accuracy among all the evaluated scenarios. An evaluation of NDVI and EVI in addition to the 12 channels from Sentinel-1 and Sentinel-2 resulted in a 80.9% training accuracy and 75.1% testing accuracy, which was not substantially better than results without the two vegetation indices included. Therefore, all further modeling was conducted using only the 12 channels from Sentinel-1 and Sentinel-2, and use of vegetation indices was abandoned in the analysis. Vegetation indices may offer advantages when simplicity is desired, but the deep machine learning approaches can accept data from multiple Sentinel channels without need to compute vegetation indices.

Because the testing accuracy demonstrated the model performance for independent data, it was the main criteria used to judge the appropriateness of data inclusion in the model. On the other hand, training accuracy can be impacted by overfitting, making it a less robust metric than testing accuracy for selection of variables. No tested subset or superset of variables improved testing accuracy better than simply inputting the two Sentinel-1 and ten Sentinel-2 channels. No individual channel, no individual vegetation index, and neither Sentinel-1 nor Sentinel-2 alone could outperform the combination of 12 Sentinel channels. Enhancing the 12 channels with additional vegetation index data did not improve testing accuracy above that of the 12 channels alone. Pairing NDVI and EVI with the two Sentinel-1 bands achieved a testing accuracy of 73.2%, meaning efforts to consolidate the Sentinel-2 data using vegetation indices led to a reduction in testing accuracy. Finally, testing accuracy when leaving out one of each of the 12 channels was not improved above that for all 12 channels (not shown). Among various tests of model input data, no combination of inputs could improve testing accuracy better than the simple and straight-forward approach of inputting the 12 Sentinel channels.

Further justification for choosing to input the 12 Sentinel channels follows from the modeling framework. Unlike multiple linear regression where multicollinearity can affect the interpretation of regression coefficients, neural network models are not linear models and are not similarly affected by multicollinearity. Neural networks are typically overparameterized to the extent that interpretation of coefficients is difficult and mostly unimportant; the main goal of deep learning is prediction rather than model interpretation. Secondly, there was no substantial reduction in computational effort by reducing inputs. The time duration to fit and test models for results shown in Fig. 3 ranged from 0.83 to 0.89 hours, which shows the minimal impact of number of inputs on computational duration. Finally, the configurations of Sentinel-1 and Sentinel-2 are fixed and operational, and its data products are established and available. Given the goal to accurately predict rice production stages, the sensible strategy was to use any available data that contributed to this goal, as demonstrated by improvements to the testing accuracy. Because the inclusion of all 12 Sentinel channels tended to improve model testing accuracy and there was no clear reason

for reducing the number of input variables, the choice to use 12 Sentinel channels was justified. Meanwhile, refitting the model with fewer inputs demonstrated the loss of accuracy from simplifications to the model inputs, while refitting the model with individual channels demonstrated the relative contribution of each channel to the prediction problem (Fig. 3).

3.2. Cloud impacts

Assessments of the Sentinel-2 imagery revealed substantial cloudiness issues over West Java during the study period (Table 2). Using the "CLOUDY_PIXEL_PERCENTAGE" property to retrieve only the images with cloud cover <20%, the number of Sentinel-2 images was reduced from more than 40 to less than 15 for all of the half-month periods. Also, there were no Sentinel-2 image collections with cloud cover <20% for two of the half-month periods. Cloud pixels were present in amounts ranging from 5.2% to 66.7% of the West Java land area, depending on the half-month period. This means up to two-thirds of the West Java land area had poor quality optical remote sensing data in a given half-month. Throughout the 2018–2019 paddy rice production season, West Java experienced substantial cloudiness issues that likely limited the utility of Sentinel-2 imagery for paddy rice mapping. Nonetheless, the available, cloudless Sentinel-2 data was critically important for the modeling effort. Testing accuracies were 69.1%, 72.8%, and 75.9% for Sentinel-1 alone, Sentinel-2 alone, and Sentinel-1 and Sentinel-2 combined, respectively (Fig. 3(b)). Thus, despite missing up to two-thirds of the Sentinel-2 data in a given half-month, Sentinel-2 alone slightly outperformed Sentinel-1 alone, and the available Sentinel-2 data substantially contributed to improved identification of rice production stages.

3.3. Modeling

Different implementations of deep and non-deep machine learning methods demonstrated variation in classification accuracy and computational duration (Fig. 4). The favored configuration was the LSTM-based RNN (Fig. 2), which demonstrated the greatest accuracy with reduced computational effort among the deep learning methods (Fig. 4 (a)). Furthermore, the LSTM-based RNN lacked a CNN component, which simplified the model implementation. As reported previously, the LSTM-based RNN achieved a classification accuracy of 79.6% for model training and 75.9% for model testing when inputting 12 Sentinel channels to the model. Similarly, the GRU-based RNN lacked a CNN component, but its classification results were slightly worse than the LSTM-based RNN, while its computational duration was 1.4 times longer. The Conv2D model, which included a time-distributed CNN but no RNN method, was the poorest performer among the deep learning methods with classification accuracy less than 60% for both training and testing. On the other hand, the Conv2D LSTM approach, which used a time-distributed CNN followed by an LSTM-based RNN, achieved the greatest training accuracy of 92.0% among 21,696 samples of six monthly observations. However, the testing accuracy was reduced to 75.8% and was no better than the LSTM-based RNN alone (Fig. 4(a)). With the large discrepancy in training and testing accuracy, the Conv2D LSTM was likely overparameterized with too many neural network layers (Table 3), and simpler neural network implementations were justified for this problem. The ConvLSTM2D model, which nested the CNN and RNN components, did not suffer the same overparameterization issues, but the computational duration was 2.3 times longer than the LSTM-based RNN, and there was no improvement to the classification accuracy. The non-deep machine learning classifiers provided results with substantially less computational effort; however, classification accuracy for model testing was not greater than 63.3%

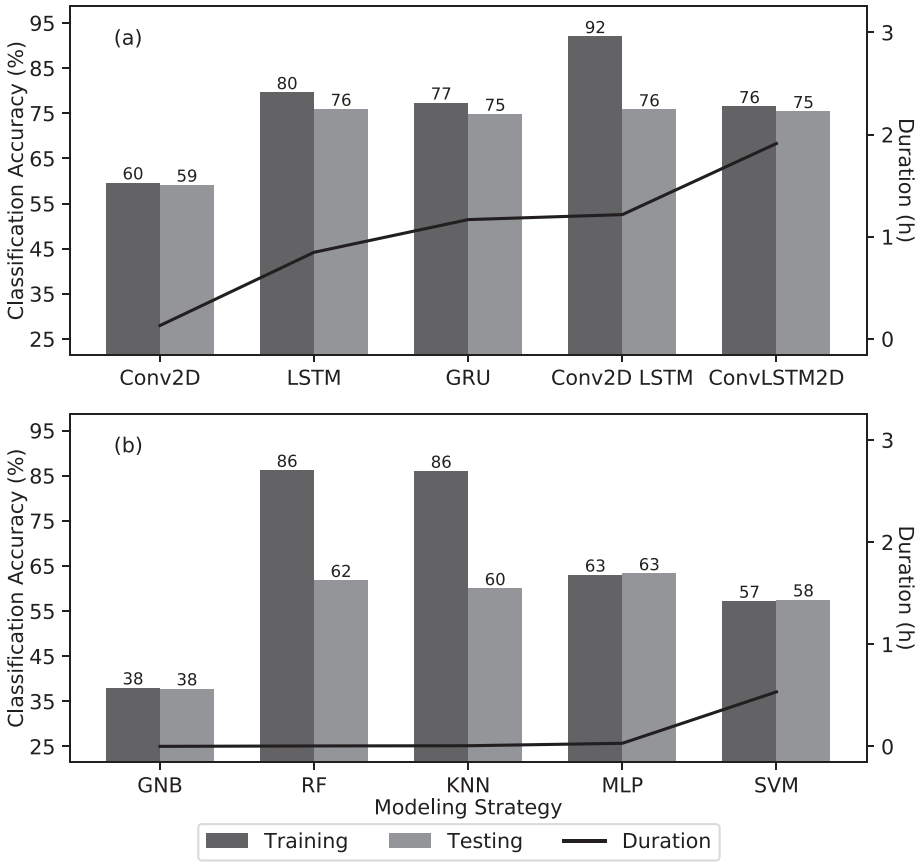

Fig. 4. Accuracy of classifying rice production stage or other land use conditions using (a) deep machine learning methods in the Keras package for Python, including two recurrent neural networks (LSTM and GRU), a convolutional neural network (Conv2D), and two approaches for combining recurrent and convolutional neural networks (Conv2D LSTM and ConvLSTM2D) and (b) other non-deep classifiers in the scikit-learn package for Python, including Gaussian Naive Bayes (GNB), random forest (RF), *k*-nearest neighbors (KNN), multi-layer perceptron (MLP) neural network, and support vector machine (SVM) classifiers.

among the non-deep methods (Fig. 4(b)). Overall, the LSTM-based RNN was a relatively simple, accurate, and efficient method to classify paddy rice production stages in this study. All other results reported herein are based on the LSTM-based RNN method.

Optimal parameter settings for LSTM-based RNN performance were determined while considering classification accuracy for both model training and testing and also the required computational duration. The chosen optimal settings resulted in a classification accuracy of 79.6% for model training and 75.9% for model testing, and computations were completed in 0.86 h (52 min). Both the training accuracy and duration were relatively insensitive to the ratio of training observations to testing observations (Fig. 5(a)). Classification accuracy was greatest for ratios of 8:1 and 10:1. However, because duration increased slightly with increasing ratio, the 8:1 ratio of training observations to testing observations was chosen as the optimal ratio. Computational time increased substantially with reduced batch sizes, while accuracy worsened (Fig. 5 (b)). A batch size of 16 maximized classification accuracy, meaning data were processed in groups of 16 observations. Computational duration also increased with increasing numbers of LSTM hidden layers, without improvements in classification accuracy (Fig. 5(c)). The optimal number of LSTM hidden layers was chosen as 750. Finally, the computational duration increased substantially with the number of epochs (Fig. 5(d)). Also, classification accuracy for model training increased with the number of epochs, but accuracy for model testing was maximized at 20 epochs. The result demonstrated that epochs above 20 likely led to model overfitting, so the optimal setting was chosen as 20 epochs. Based on the result of this sensitivity analysis, all model performance results in this study were reported for (1) an 8:1 ratio of training to testing data, (2) a batch size of 16 for model training, (3) 750 LSTM hidden layers, and (4) 20 epochs for model training.

Preliminary neural network modeling aimed to classify the data according to the ten original BPS survey classes (Table 1), which resulted in a training accuracy of 70.1% and testing accuracy of 60.5% (not shown). Improvements in model performance were made by combining and eliminating some classes. First, the “no data” class was eliminated by removing any location with missing data for any of the six observation periods. Errors of omission were above 90% for the “no data” class (not shown), so overall model performance was easily improved by eliminating locations with missing data. Preliminary modeling also often simulated Class 5 (land preparation) as Class 1 (first vegetative), Class 1 and Class 9 (fallowed after rice harvest) as Class 5, and Class 5 as Class 9. Due to the confusion among these classes, all three were combined into one class. Furthermore, because Class 6 (damaged crop) was often confused with these classes and because Class 6 was observed very infrequently, it was also merged with Classes 1, 5, and 9. Because Classes 1, 5, 6 and 9 related to the early-season transplant or land preparation stages, damaged conditions, or fallow periods after rice harvest, respectively, the condense class represented conditions with little to no vegetation. Class 7 (crops other than rice) and Class 8 (non-agricultural land) were also often confused, so these two classes were merged to represent a non-rice class. These efforts reduced the number of classes in the final model to 5. The resulting overall training accuracy was 79.6% (Table 4) among 115,710 observations of paddy rice over the six-month period in West Java, and the testing accuracy was 75.9% (Table 5) among 14,466 paddy rice observations. Depending on the class grouping, errors of omission and comission ranged from 15.1% to 38.4%. Given the small error rates with such a large amount of observations, the model was considered reasonable for development of paddy rice maps across West Java.

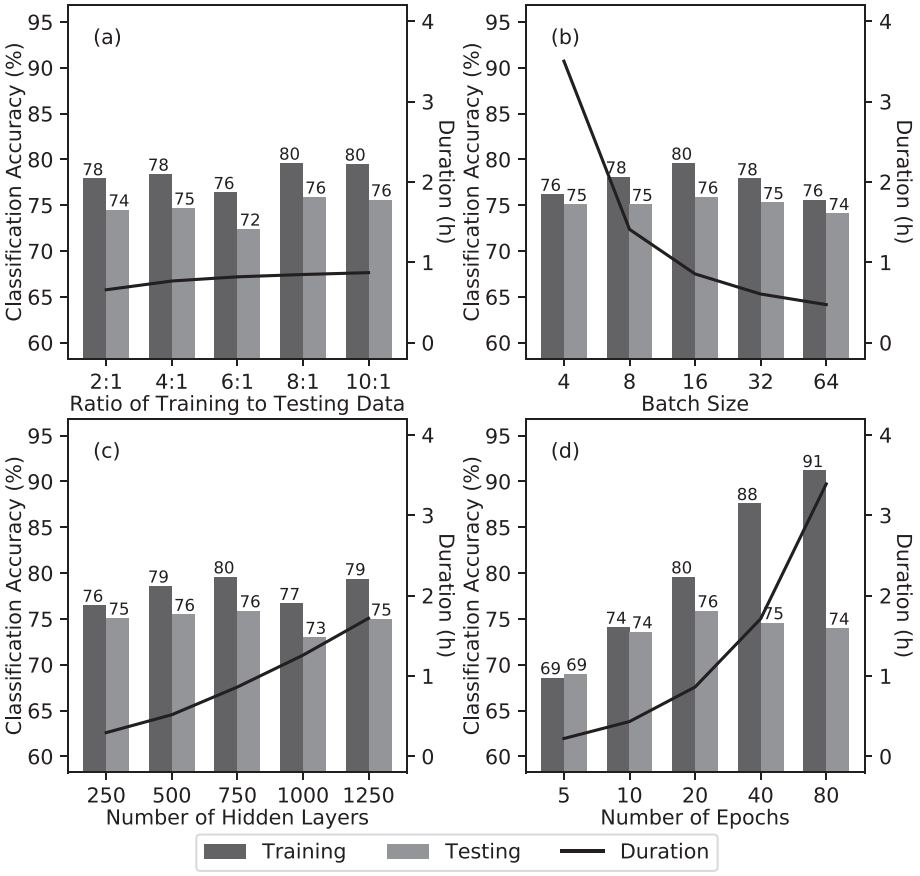

Fig. 5. Accuracy of classifying rice production stage or other land use condition via a recurrent neural network with long short term memory (LSTM) nodes, as affected by (a) the ratio of training observations to testing observations, (b) the batch size used for model training, (c) the number of LSTM hidden layers, and (d) the number of epochs used for model training. The computational duration (h) for both training and testing is also provided for each case.

Table 4
Classification accuracy assessment for training a recurrent neural network with long short term memory (LSTM) nodes. The overall classification accuracy was 79.6% among 115,710 paddy rice observations from 6 months of agricultural surveying at 19,285 locations across West Java, Indonesia. Bold font denotes the diagonal of the confusion matrix.

		Modeled						
	Class #	1,5,6&9	2	3	4	7&8	Total	Omission Error (%)
Observed	1,5,6&9	**38,522**	1,272	232	1,193	4,168	45,387	15.1
	2	1,895	**9,484**	1,556	21	1,124	14,080	32.6
	3	719	791	**14,353**	912	1,259	18,034	20.4
	4	1,628	15	616	**8,461**	742	11,462	26.2
	7&8	4,207	356	525	360	**21,299**	26,747	20.4
	Total	46,971	11,918	17,282	10,947	28,592	**115,710**	
	Comission Error (%)	18.0	20.4	16.9	22.7	25.5		

Table 5
Classification accuracy assessment for testing a recurrent neural network with long short term memory (LSTM) nodes. The overall classification accuracy was 75.9% among 14,466 paddy rice observations from 6 months of agricultural surveying at 2,411 locations across West Java, Indonesia. Bold font denotes the diagonal of the confusion matrix.

		Modeled						
	Class #	1,5,6&9	2	3	4	7&8	Total	Omission Error (%)
Observed	1,5,6&9	**4,651**	190	44	199	574	5,658	17.8
	2	288	**1,097**	228	6	161	1,780	38.4
	3	80	135	**1,671**	147	191	2,224	24.9
	4	229	1	113	**960**	124	1,427	32.7
	7&8	561	52	100	62	**2,602**	3,377	22.9
	Total	5,809	1,475	2,156	1,374	3,652	**14,466**	
	Comission Error (%)	19.9	25.6	22.5	30.1	28.8		

3.4. Paddy rice mapping

By applying the neural network model to map paddy rice across West Java, five-band images were developed to map the probability of pixel membership to each of the five condensed BPS classes. These data were displayed as an RGB color composite image with the R, G, and B channels representing the probability of membership to rice harvest, heading, and tillering classes, respectively (Fig. 6). Darker color indicated reduced probability of membership to the three rice production stages and consequently greater probability of membership to the other two classes, which involved non-rice areas and rice areas with little to no vegetation present. Results from November and December 2018 showed relatively small areas of rice tillering and harvest in northern West Java (Figs. 6(a) and 6(b)), and darker color throughout much of the province indicated smaller areas in rice production at this time. Beginning in January 2019, a sharp increase of rice area in the tillering stage was present (Fig. 6(c)), which is consistent with the typical rice production cycle. Moving into February 2019, many areas transitioned from the tillering to the heading stage, while new areas at the rice tillering stage appeared (Fig. 6(d)). March and April 2019 were primary months for rice harvest, with harvest expanding from south to north in the irrigated paddy rice region of northern West Java (Fig. 6(d) and 6(e)). Progression of rice production from south to north in northern West Java agreed with the reasoning of Ribbes and Le Toan (1999), who described the irrigation process whereby water from the Lake of Jatiluhur is distributed to paddy fields from south to north in mid-November through early January. Overall, the rice classification maps demonstrated progression of rice through a typical production cycle with realistic spatial patterns in areas of West Java commonly associated with rice production.

Similar temporal patterns of rice production were present in the BPS survey data and Sentinel satellite data (Fig. 7). Both data sets demonstrated that January and February were predominant months for rice tillering, February and March were predominant months for rice heading, and March and April were predominant months for rice harvest. The temporal similarities of predominant rice production stages between the two data sets were striking, which means both data sets likely provide meaningful information about the progression of rice production across West Java. However, the percentages of BPS observations attributed to each rice class were always greater than the corresponding percentages of Sentinel pixels attributed to each class. This means the BPS area sampling frame is likely biased to have more observations in known rice production regions, and the Sentinel maps may have greater advantage when used to compute the area of paddy rice harvest.

4. Discussion

While the present rice mapping study achieved classification accuracies of 79.6% and 75.9% for model training and testing, other Indonesian rice mapping studies reported classification accuracies greater than 87% (Nuarsa et al., 2012; Ribbes and Le Toan, 1999). However, the previous studies classified only rice versus non-rice areas, and they did not classify unique rice production stages. Furthermore, only limited rice data from existing land use maps (i.e., not ground observations) were used to compute classification accuracy in previous studies. In the present study, the use of BPS agricultural survey data, which involved monthly ground-based observations at 21,696 locations across West Java over a six-month period, was both novel and unprecedented, and classification accuracy of >75% for unique rice production stages was deemed outstanding. Given the great efforts by BPS to collect comprehensive, statistically valid, and quality-controlled data sets on national rice production, the classification accuracies reported herein may provide a truer classification performance estimate as compared to reports in previous studies. Ultimately, classification results depend on the quality of the ground reference data set, which is difficult to quantify, but greater effort was certainly expended to produce the reference data used in this study as compared to previous studies. The immense value of the BPS survey data for development of machine learning models to classify remote sensing images cannot be understated. In fact, the main limitation for transfer of the methodology to other rice production regions is the availability of ground reference data sets like those now

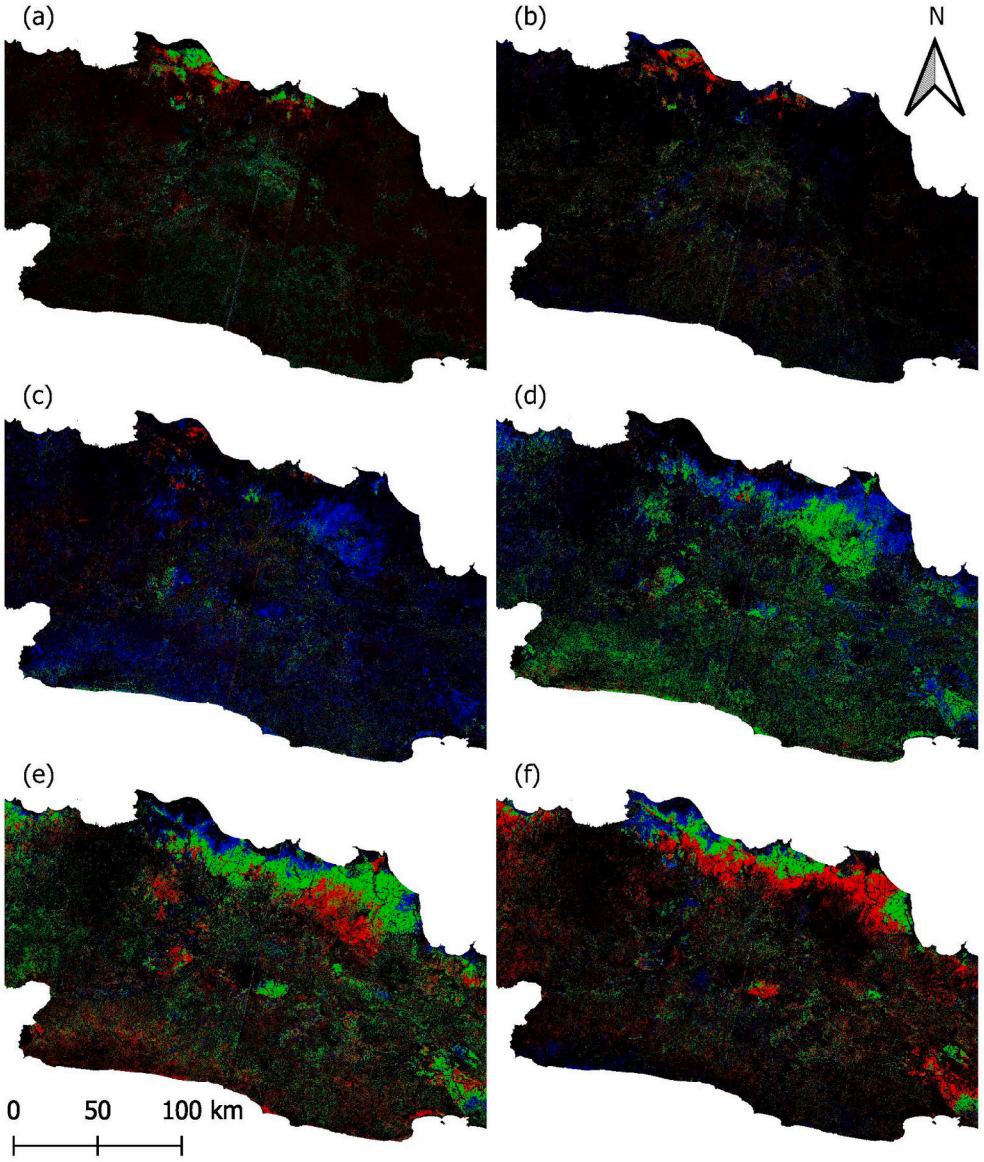

Fig. 6. Maps of rice production stages across West Java, Indonesia during the latter half-months of (a) November 2018, (b) December 2018, (c) January 2019, (d) February 2019, (e) March 2019, and (f) April 2019. Blue, green, and red color represents greater probability of rice at tillering, heading, and harvest stages, respectively. Darker color indicates smaller probability of rice at the three production stages. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

routinely collected by BPS. Nevertheless, improvements to the surveying process could likely improve classification results. For example, the BPS rice surveying strategy could be simplified, because some of the original BPS classifications (Table 1) were difficult to segregate with the Sentinel data. Also, future investigations should identify specific cases where BPS survey data and Sentinel classifications disagree, which could highlight opportunities for correction of surveying data and improvement of surveying methods. In this way, remote sensing image classifications could potentially serve as additional quality control on the ground reference data sets, thereby leading to improved agreement between them.

A primary goal for mapping crop production involves planning and preparation for future handling, redistribution, processing, and often export of agricultural products. Toward this end, the temporal aspect of the recurrent neural network model offers unique capability for anticipating future harvest conditions, because the network learns the temporal progression of the rice production cycle. As shown in Figs. 6(e) and 6(f), not only are the recently harvested rice areas mapped, but also the areas with rice at tillering and heading stages are also mapped. Because a recurrent neural network was used, these maps are linked in time, and there is greater certainty that the reported rice classifications follow the natural cycle of rice production. Thus, it is expected that areas classified at the rice heading stage in a given month should naturally progress to the harvest stage in the near future. In an operational framework, where new data was used to refit the neural network and develop up-to-date classification maps each month, information on anticipated future rice harvest areas will be highly valuable for harvest planning activities at all levels of the supply chain. Because RNN models performed relatively well herein (Fig. 4), the study demonstrated the value of explicitly modeling the temporal progression of the crop production cycle. On the other hand, CNN models did not increase predictive skill, meaning spatial filter convolution was less useful for classifying rice production stage as compared to temporal modeling. The spatial aspect of the Sentinel images was critically important for calculating total rice harvest area from the classification maps but was less important for developing the classification maps themselves.

5. Conclusions

The study combined Indonesia's modern agricultural survey data, open-access Sentinel satellite imagery through Google Earth Engine, and recurrent neural network modeling based on the TensorFlow machine

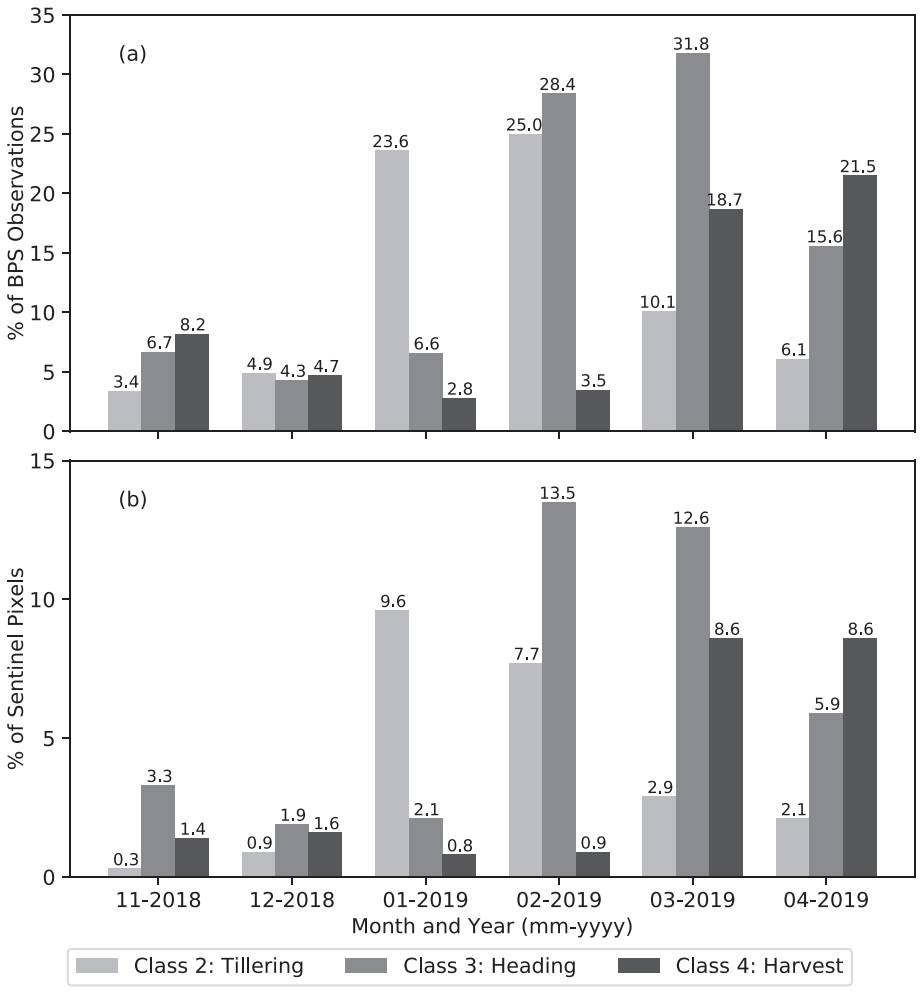

Fig. 7. Presence of rice at three production stages (tillering, heading, and harvest) across West Java, Indonesia during the six latter half-month periods from November 2018 to April 2019. Data are expressed as (a) a percentage of Badan Pusat Statistik (BPS) observations at 21,696 ground surveying locations during each half-month period and (b) a percentage of 376×10^6, 10-m Sentinel pixels classified by a recurrent neural network with long short term memory (LSTM) nodes to maximize probability of the given rice stage during each half-month period.

learning package. Ability to classify the progression of unique paddy rice production stages, including tillering, heading, and harvest, with accuracy >75% was a novel and encouraging finding with implications for better predictions of future rice harvest events and overall estimation of rice harvest area. Future work should expand efforts to evaluate machine learning techniques with ground-based and satellite data sets over longer periods of time and for other crop types. Furthermore, while the value of Indonesia's modern agricultural surveys cannot be understated, the integration of these observations with satellite data through machine learning methods should help identify ways to make corrections and improve efficiencies in how the surveys are conducted.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Kelly Thorp reports financial support was provided by USDA Foreign Agriculture Service.

Acknowledgments

The authors acknowledge the Borlaug Fellowship program of the USDA Foreign Agricultural Service for contributing partial funding for this research.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. URL: http://tensorflow.org/. software available from tensorflow.org.

Bazzi, H., Baghdadi, N., El Hajj, M., Zribi, M., Minh, D.H.T., Ndikumana, E., Courault, D., Belhouchette, H., 2019. Mapping paddy rice using Sentinel-1 SAR time series in Camargue, France. Remote Sens. 11, 887.

Boschetti, M., Busetto, L., Manfron, G., Laborte, A., Asilo, S., Pazhanivelan, S., Nelson, A., 2017. PhenoRice: a method for automatic extraction of spatio-temporal information on rice crops using satellite data time series. Remote Sens. Environ. 194, 347–365.

Bouvet, A., Le Toan, T., 2011. Use of ENVISAT/ASAR wide-swath data for timely rice fields mapping in the Mekong River Delta. Remote Sens. Environ. 115, 1090–1101.

Bridhikitti, A., Overcamp, T.J., 2012. Estimation of Southeast Asian rice paddy areas with different ecosystems from moderate-resolution satellite imagery. Agric. Ecosyst. Environ. 146, 113–120.

Chen, C., Mcnairn, H., 2006. A neural network integrated approach for rice crop monitoring. Int. J. Remote Sens. 27, 1367–1393.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Stroudsburg, Pennsylvania, USA, pp. 1724–1734.

Choudhury, I., Chakraborty, M., 2006. SAR signature investigation of rice crop using RADARSAT data. Int. J. Remote Sens. 27, 519–534.

Clauss, K., Ottinger, M., Kuenzer, C., 2018. Mapping rice areas with Sentinel-1 time series and superpixel segmentation. Int. J. Remote Sens. 39, 1399–1420.

Clauss, K., Yan, H., Kuenzer, C., 2016. Mapping paddy rice in China in 2002, 2005, 2010 and 2014 with MODIS time series. Remote Sens. 8, 434.

Dong, J., Xiao, X., 2016. Evolution of regional to global paddy rice mapping methods: a review. ISPRS J. Photogramm. Remote Sens. 119, 214–227.

Dong, J., Xiao, X., Kou, W., Qin, Y., Zhang, G., Li, L., Jin, C., Zhou, Y., Wang, J., Biradar, C., Liu, J., Moore III, B., 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.
Dong, J., Xiao, X., Menarguez, M.A., Zhang, G., Qin, Y., Thau, D., Biradar, C., Moore III, B., 2016. Mapping paddy rice planting area in Northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. Remote Sens. Environ. 185, 142–154.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.
Inoue, Y., Kurosu, T., Maeno, H., Uratsuka, S., Kozu, T., Dabrowska-Zielinska, K., Qi, J., 2002. Season-long daily measurements of multifrequency (Ka, Ku, X, C, and L) and full-polarization backscatter signatures over paddy rice field and their relationship with biological variables. Remote Sens. Environ. 81, 194–204.
Kontgis, C., Schneider, A., Ozdogan, M., 2015. Mapping rice paddy extent and intensification in the Vietnamese Mekong River Delta with dense time stacks of Landsat data. Remote Sens. Environ. 169, 255–269.
Kuenzer, C., Knauer, K., 2013. Remote sensing of rice crop areas. Int. J. Remote Sens. 34, 2101–2139.
Lee, N., Monica, A., Daratista, I., 2012. Mapping Indonesian paddy fields using multiple-temporal satellite imagery. Afr. J. Agric. Res. 7, 4038–4044.
Manjunath, K.R., More, R.S., Jain, N.K., Panigrahy, S., Parihar, J.S., 2015. Mapping of rice-cropping pattern and cultural type using remote-sensing and ancillary data: a case study for South and Southeast Asian countries. Int. J. Remote Sens. 36, 6008–6030.
McCloy, K.R., Smith, F.R., Robinson, M.R., 1987. Monitoring rice areas using LANDSAT MSS data. Int. J. Remote Sens. 8, 741–749.
Mladenova, I.E., Jackson, T.J., Bindlish, R., Hensley, S., 2013. Incidence angle normalization of radar backscatter data. IEEE Trans. Geosci. Remote Sens. 51, 1791–1804.
Mosleh, M.K., Hassan, Q.K., Chowdhury, E.H., 2015. Application of remote sensors in mapping rice area and forecasting its production: a review. Sensors (Switzerland) 15, 769–791.
Ndikumana, E., Minh, D.H.T., Baghdadi, N., Courault, D., Hossard, L., 2018. Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. Remote Sens. 10, 1217.
Nelson, A., Setiyono, T., Rala, A.B., Quicho, E.D., Raviz, J.V., Abonete, P.J., Maunahan, A.A., Garcia, C.A., Bhatti, H.Z.M., Villano, L.S., Thongbai, P., Holecz, F., Barbieri, M., Collivignarelli, F., Gatti, L., Quilang, E.J.P., Mabalay, M.R.O., Mabalot, P.E., Barroga, M.I., Bacong, A.P., Detoito, N.T., Berja, G.B., Varquez, F., Wahyunto, Kuntjoro, D., Murdiyati, S.R., Pazhanivelan, S., Kannan, P., Nirmala Mary, P.C., Subramanian, E., Rakwatin, P., Intrman, A., Setapayak, T., Lertna, S., Minh, V.Q., Tuan, V.Q., Duong, T.H., Quyen, N.H., Van Kham, D., Hin, S., Veasna, T., Yadav, M., Chin, C., Ninh, N.H., 2014. Towards an operational SAR-based rice monitoring system in Asia: examples from 13 demonstration sites across Asia in the RIICE project. Remote Sens. 6, 10773–10812.
Nguyen, D.B., Gruber, A., Wagner, W., 2016. Mapping rice extent and cropping scheme in the Mekong Delta using Sentinel-1A data. Remote Sens. Letters 7, 1209–1218.
Nguyen, D.B., Wagner, W., 2017. European rice cropland mapping with Sentinel-1 data: the Mediterranean region case study. Water 9, 392.
Nguyen, T.T.H., De Bie, C.A.J.M., Ali, A., Smaling, E.M.A., Chu, T.H., 2012. Mapping the irrigated rice cropping patterns of the Mekong delta, Vietnam, through hyper-temporal spot NDVI image analysis. Int. J. Remote Sens. 33, 415–434.
Nuarsa, I.W., Nishio, F., Hongo, C., Mahardika, I.G., 2012. Using variance analysis of multitemporal MODIS images for rice field mapping in Bali Province, Indonesia. Int. J. Remote Sens. 33, 5402–5417.
Okamoto, K., Fukuhara, M., 1996. Estimation of paddy field area using the area ratio of categories in each mixel of Landsat TM. Int. J. Remote Sens. 17, 1735–1749.
Panuju, D.R., Mizuno, K., Trisasongko, B.H., 2013. The dynamics of rice production in Indonesia 1961-2009. J. Saudi Soc. Agric. Sci. 12, 27–37.
Park, S., Im, J., Park, S., Yoo, C., Han, H., Rhee, J., 2018. Classification and mapping of paddy rice by combining Landsat and SAR time series data. Remote Sens. 10, 447.
Ribbes, F., Le Toan, T., 1999. Rice field mapping and monitoring with RADARSAT data. Int. J. Remote Sens. 20, 745–765.
Rotairo, L., Durante, A.C., Lapitan, P., Rao, L.N., 2019. Use of Remote Sensing to Estimate Paddy Area and Production: A Handbook. Asian Development Bank, Manila, Philippines.
Sakamoto, T., Sprague, D.S., Okamoto, K., Ishitsuka, N., 2018. Semi-automatic classification method for mapping the rice-planted areas of Japan using multi-temporal Landsat images. Remote Sens. Appl.: Soc. Environ. 10, 7–17.
Setiawan, Y., Liyantono, Fatikhunnada, A., Permatasaria, P.A., Aulia, R., 2016. Dynamics pattern analysis of paddy fields in Indonesia for developing a near real-time monitoring system using MODIS satellite images. Procedia Environ. Sci. 33, 108–116.
Setiawan, Y., Rustiadi, E., Yoshino, K., Liyantono, Effendi, H., 2014. Assessing the seasonal dynamics of the Java's paddy field using MODIS satellite images. ISPRS Int. J. Geo-Inf. 3, 110–129.
Shao, Y., Fan, X., Liu, H., Xiao, J., Ross, S., Brisco, B., Brown, R., Staples, G., 2001. Rice monitoring and production estimation using multitemporal RADARSAT. Remote Sens. Environ. 76, 310–325.
Shi, J., Huang, J., 2015. Monitoring spatio-temporal distribution of rice planting area in the Yangtze River Delta region using MODIS images. Remote Sens. 7, 8883–8905.
Shi, X., Chen, Z., Wang, H., Yeung, D., Wong, W., Woo, W., 2015. Convolutional LSTM Network: A machine learning approach for precipitation nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems. Neural Information Processing Systems Foundation, La Jolla, California, USA, pp. 802–810.
Son, N., Chen, C., Chen, C., Duc, H., Chang, L., 2013. A phenology-based classification of time-series MODIS data for rice crop monitoring in Mekong Delta, Vietnam. Remote Sens. 6, 135–156.
Tennakoon, S.B., Murty, V.V., Eiumnoh, A., 1992. Estimation of cropped area and grain yield of rice using remote sensing data. Int. J. Remote Sens. 13, 427–439.
Torbick, N., Chowdhury, D., Salas, W., Qi, J., 2017. Monitoring rice agriculture across Myanmar using time series Sentinel-1 assisted by Landsat-8 and PALSAR-2. Remote Sens. 9, 119.
Xiao, X., Boles, S., Frolking, S., Li, C., Babu, J.Y., Salas, W., Moore III, B., 2006. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. Remote Sens. Environ. 100, 95–113.
Xiao, X., Boles, S., Frolking, S., Salas, W., Moore, B., Li, C., He, L., Zhao, R., 2002. Observation of flooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data. Int. J. Remote Sens. 23, 3009–3022.
Xiao, X., Boles, S., Liu, J., Zhuang, D., Frolking, S., Li, C., Salas, W., Moore III, B., 2005. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. Remote Sens. Environ. 95, 480–492.
Yin, Q., Liu, M., Cheng, J., Ke, Y., Chen, X., 2019. Mapping paddy rice planting area in northeastern China using spatiotemporal data fusion and phenology-based method. Remote Sens. 11, 1699.
Zhang, G., Xiao, X., Dong, J., Kou, W., Jin, C., Qin, Y., Zhou, Y., Wang, J., Menarguez, M. A., Biradar, C., 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.
Zhang, M., Lin, H., Wang, G., Sun, H., Fu, J., 2018a. Mapping paddy rice using a convolutional neural network (CNN) with Landsat 8 datasets in the Dongting Lake Area, China. Remote Sens. 10, 1840.
Zhang, X., Wu, B., Ponce-Campos, G.E., Zhang, M., Chang, S., Tian, F., 2018b. Mapping up-to-date paddy rice extent at 10M resolution in China through the integration of optical and synthetic aperture radar images. Remote Sens. 10, 1200.
Zhu, A., Zhao, F., Pan, H., Liu, J., 2021. Mapping rice paddy distribution using remote sensing by coupling deep learning with phenological characteristics. Remote Sens. 13, 1360.