Generalization of Convolutional LSTM Models for Crop Area Estimation

Maysa Malfiza Garcia de Macedo, Andrea Britto Mattos, and Dário Augusto Borges Oliveira

***Abstract*—The population growth and consequent global rise in food demand require increasingly efficient agricultural solutions, in what is commonly called digital agriculture. Among promising initiatives, the use of remotely sensed data combined with machine learning algorithms enables handling faster agricultural operations with lower associated cost. One of the most important activities in digital agriculture is crop identification, which is fundamental for managing the inventory of a farm by producers and governmental authorities, and has been addressed by several prior works. In this article, we explore crop identification from a scalability perspective using the premise that data trained at a set of labeled geo-referenced regions in the agricultural pole at central western Brazil may be used for identifying crops at the entire municipality area. We propose to use convolutional long-short term memory networks for identifying crop types using a public labeled training set, and then apply the trained model for estimating crop area in a larger area involving the entire municipality unlabeled data. Our results were evaluated against governmental census data and report evidences that the tested crop identification network is able to successfully estimate crop area from much larger unlabeled data for different crop types.**

***Index Terms*—Agricultural engineering, crops, image classification, machine learning, neural networks, satellites.**

I. Introduction

CROP identification is the process of determining which cultivated areas correspond to each crop type (such as soybean, maize, cotton, etc.). This process is fundamental for the agricultural producer, not only because it represents a key step for crop monitoring, yield forecasting, and risk management, but also because it may be critical for external authorities in charge of area-based subsidy controls and loans.

Due to the huge economic impact associated with agricultural operations and the recent increase in publicly available data, remote sensing applications have demonstrated large potential for solving agriculture monitoring tasks, such as crop identification, yield estimation, and land cover segmentation in an automated manner [1]–[4]. For crop identification, in particular, most works employ supervised approaches based on traditional classifiers, such as support vector machine (SVM) [5], [6] or neural networks (NNs) [7].

The limitation of such data-driven approaches rely on the difficulty in applying trained classifiers for one region to different ones, as small differences in geography and climate usually affect severely the classification results. In this article, we address the crop identification problem focusing on generalization aspects, i.e., by evaluating if a classifier trained on a limited zone may be successfully used for identifying cultivated regions in much larger areas. Our results are further validated considering governmental census data, which are collected with a different methodology and compiled at a much higher scale than the annotations in training set.

This article is structured as follows. In Section II, we list previous works on crop identification and remark the contribution of this work. In Section III, we describe the employed methodology. The experiments and data are reported and further discussed in Section IV, and our concluding remarks are presented in Section V.

II. Related Work

Several works have previously tackled crop identification and proposed a wide range of approaches for handling it. Many authors take as input high-temporal and spatial resolution data, such as synthetic-aperture radar (SAR) images, as in [8], where they evaluated the selection of different features for early crop-type identification and presented the classification results for eight different classes. In [6], polarimetric SAR imagery was used for training a set of classifiers, with SVM producing the best results.

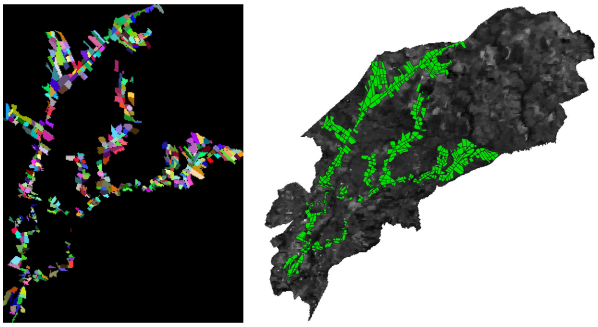
Multisource satellite data, where the authors explore the combination of images with different resolutions, is widely used [9], [10] and Sentinel-2, a popular data source, is employed for crop identification in [11]–[13] with all works exploring SVM and random forest-based classifiers, and Mira *et al.* [14] exploring *K*-nearest neighbor (KNN). Although recent works have shown that Sentinel-2 satellite data can provide accurate crop identification results for regions with a different climate, the classifier performance may be deteriorated in some scenarios [8] affected by cloud cover.

Recent works on crop identification include, still, a bat algorithm clustering approach based on *K*-means [15] and the exploration of stochastic hidden Markov models [16], as alternatives for preliminary works based on multispectral normalized difference vegetation index (NDVI) time series [17]. Deep learning commonly ranks among state-of-the-art approaches, as indicated in [18]. As examples, Castro *et al.* [19] and Cue la Rosa *et al.* [20] compared different methods for crop identification using SAR and optical imagery, based on autoencoders, convolutional NNs (CNNs), random forests, and SVM, reporting a consistent better

Manuscript received November 12, 2019; revised January 20, 2020; accepted February 9, 2020. Date of publication March 6, 2020; date of current version April 8, 2020. *(Corresponding author: Maysa Malfiza Garcia de Macedo.)*

The authors are with IBM Research, Sao Paulo 04007-900, Brazil (e-mail: mmacedo@br.ibm.com; abritto@br.ibm.com; dariobo@br.ibm.com).

Digital Object Identifier 10.1109/JSTARS.2020.2973602



(a) (b)

Fig. 1. Data used in our experiments. (a) Regions of the public dataset used for crop identification (each color denotes a different labeled field—to which we refer to as *polygon*). (b) Overlap of these regions (in green) with the entire municipality area, used for the generalization evaluation.

performance for deep learning-based models. Zhong *et al.* [21] compared the performance of 1-D CNNs and long-short term memory (LSTM) networks with SVM and random-forests for crop identification and reported good results with 1-D CNNs. The work published by Shunping *et al.* proposed a 3-D CNN with a spatiotemporal representation. Kussul *et al.* [22] compared a traditional fully connected multilayer perceptron with random forest and CNN using images from Landsat-8 and Sentinel-1 A RS satellites. Russwurm and Krner [23] evaluated the use of sequential recurrent encoders for crop identification from unfiltered temporal series of top-of-atmosphere reflectance data, achieving state-of-the-art results and coping with cloud coverage without explicit filtering. This present work differs from the literature by evaluating crop identification from a scalability perspective and investigate whether a state-of-the-art approach using convolutional LSTM (CLSTM) networks is able to successfully generalize crop area estimation from a much larger unlabeled area. Our experiments use governmental census data for evaluation, another difference from the literature, and report consistent results with the expected outcome.

III. Methodology

Our methodology consists of two major steps: First, we train a convolutional LSTM (CLSTM) temporal model using a publicly available annotated dataset [24] for crop identification; then, we use the trained model to evaluate its generalization against census crop data [25] at a much larger region. In Fig. 1, we can visualize the difference in scale between the regions from the two experiments described.

The proposed CLSTM model is able to classify four classes: *soybean*, *maize*, *cotton*, and the remaining crops were grouped in the class *others*. The first three classes were chosen because they represent the majority of crops in the considered municipality and, thus, play a very important role in the region's economy. The dataset was balanced based on the undersampling method, with all classes having the same number of samples.

To differentiate between these four classes, satellite images from sensor MODIS-16 were used in both experiments—MOD13Q1 version 6 product was chosen [26]. The acquired images derived six different channels: blue, enhanced vegetation index (EVI), NDVI, red, middle infra-red (MIR), and near infra-red (NIR). Red, NIR, Blue, and MIR are 16-b reflectance bands (units 1, 2, 3, and 7, respectively), and EVI and NDVI are primary vegetation layers. The satellite images were processed for every 16th date between the period of October 2015 to July 2016. MOD13Q1 v006 algorithm chooses the best available pixel value from all the acquisitions from the 16-day period. The pixel selection criteria considers low clouds, low view angle, and the highest-NDVI/EVI value. MODIS data, with a spatial resolution of 250 m, was used for avoiding the cloud cover problem observed at Sentinel-2 higher resolution data, mentioned in Section II, and because our approach targets very large areas. The 16-day composite image did not require any additional processing for cloud removal.

A. Convolutional LSTM Networks

Crop identification methods usually take into account the temporal sequence of remotely sensed images to properly characterize the plant growth cycle and consequently identify the crop type itself. In this context, recurrent NNs (RNNs) have been widely used to tackle sequential data by employing a hidden layer output h_t that is computed at time t using not only current input data x_t but also the previous output h_{t^1}. This allows such networks to embed short-memory capabilities and handle temporal or sequential data, but with a limited memory range.

To handle patterns characterized by longer memory dependencies, such as plant growth cycles, LSTM networks propose to employ cells with a more complex memory structure controlled by gates that allow the modeling of long- and short-term memories. These additional gates control an internal state vector that enables long-term memory learning, and a cascade of LSTM layers with various cells per layer is able to learn temporal patterns from current and all previous n input data observations. Not surprisingly, such networks quickly became very popular in crop identification methods.

B. Crop Identification Using CLSTM Networks

In this article, we used convolutional and LSTM layers (CLSTM) to identify crops, as shown in Fig. 2, and the difference between LSTM models and CLSTM models is precisely the presence of a convolutional block of layers before the LSTM block. This approach follows previous works [27], [28] that demonstrated good classification results; again, we remark that the goal of our article is to assess the scaling generalization aspect, which was not addressed at these works, instead of an architecture evaluation. Our training set is composed of different channel temporal sequences for each observed pixel, and our assumption is that convolutional layers help to identify curve patterns in the input temporal sequences, and LSTM layers help to identify the temporal dependence of these patterns for each crop type.

Each pixel is a training sample represented by the six satellite channels from MODIS chronologically organized according to the acquisition date. The network architecture is represented as a sequence that has 1-D convolutional layers with dropout and ReLU activation, followed by a sequence of LSTM layers

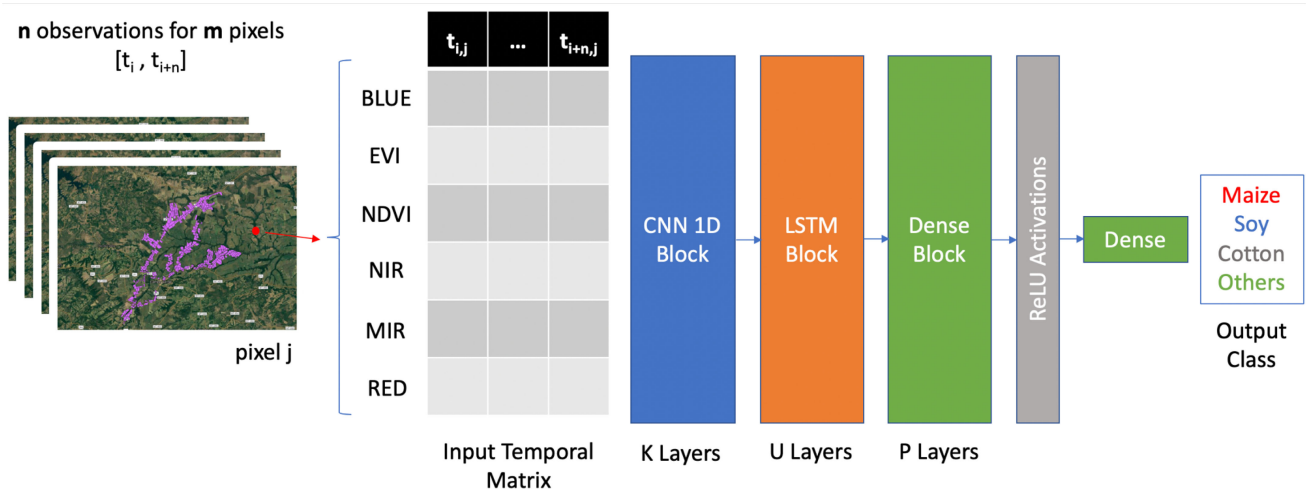


Fig. 2. Proposed NN model. A sequence of 1-D convolutional layers with dropout and ReLU activation, followed by a sequence of LSTM layers with dropout and ReLU activation, and finally by a sequence of dense layers that ends up in a final activation layer for each class using Softmax.

with dropout and ReLU activation, and finally by a sequence of dense layers that ends up in a final activation layer for each class using Softmax. The exact number of layers per block and filters per layer was experimentally defined and is detailed later in Section IV.

C. Crop Area Prediction From Unlabeled Data

While supervised crop identification is widely tackled in the literature, in this article, we aim at evaluating not only how CLSTM models generalize at a much larger dataset, but also to assess how the results correlate with governmental census data—usually collected using other sources of information [25]. Our goal is to understand if it is possible to use trained models and large unlabeled remotely sensed images to predict census data, which would support the use of automatic methods for land cover census generation using small training sets.

Some adjustments to link these different annotations are necessary, however. Because there are no labels for the municipality test image displayed in Fig. 1(b) and the training data only considered plantations, it was necessary to first remove nonagricultural regions from the broader municipality data (e.g., urban areas, forests, and rivers), so we could have a fair assessment of the method. A robust way to do so would be to build a supervised classifier capable of distinguishing between agricultural and nonagricultural areas. However, due to the lack of annotated data for this purpose, we needed to employ an alternate approach. We assumed that, in comparison to the plantation zones, the nonagricultural regions usually present a smaller variation throughout the considered period of ten months. Therefore, these areas were removed, as depicted in Fig. 3, by employing two setups: we discarded pixels with standard deviation of less than 33% and pixels with a standard deviation less than 45%. More details are provided in Section IV.

Another important aspect of our experiments is that the considered census data provides only the proportions of the cultivated areas per crop type. Therefore, for evaluating our

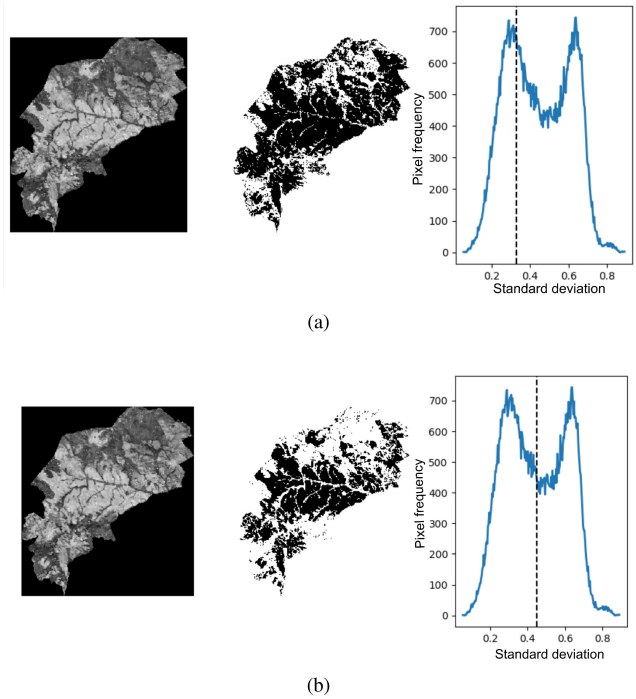


Fig. 3. Visualization of our approach for removing nonagricultural areas, by identifying regions with little or no variation during the considered period of October 2015 and July 2016. The image on the left side represents one of the channels from MODIS-16. The image on the middle represents pixels in black color related to agricultural regions candidates, and the image on the right side shows the distribution of the pixels' standard deviation. (a) Removal of regions with standard deviation less than 33%. (b) Removal of regions with standard deviation less than 45%.

results, instead of considering each individual pixel, we took into account all the pixels within the entire municipality. Clearly, this weak evaluation metric allows mistakes to compensate correct predictions and cannot be used for evaluating individual pixels or polygons, but we expect that with a very large number of samples, the proportions would remain unaffected and reflect the

prediction power of the trained model against the ground-truth proportions from census data.

To alleviate the impossibility of computing commonly used accuracy metrics in our unlabeled data experiments, we performed a bootstrap in the annotated dataset, which offers an overall result estimate and not just for an annotated dataset with a single distribution. We separated the test data from the annotated dataset into 30 subsets with different distributions, and we could get a sense of which would perform well using the employed method. The subsets were randomly formed as an attempt to represent different regions, which may have different types of irrigation and soil. This way, we describe numerically how metrics vary for different subsets inside the same municipality, and indicate how consistent our results are, as further presented in Section IV-B.

IV. Experiments and Discussion

Our experiments were performed from two different perspectives: first, we evaluated the performance of our CLSTM model in the context of crop classification using a publicly available dataset and provide trust in the chosen model; then, we evaluated the generalization of such methodology for estimating crop areas at a much bigger scale, by bootstrapping the outcome and evaluating it against crop area census data.

A. Crop-Type Classification

1) Data: The training data used in our experiments for training the crop identification model was made publicly available in [24] and consists of 513 labeled geo-referenced polygons in Campo Verde, a municipality of Mato Grosso state, in central west Brazil. The labeled dataset used for training consists of 335.880 pixels in total.

Campo Verde is localized in the *Cerrado* (Brazilian Savanna) biome, at a latitude of $15°32'48''$ south and longitude of $55°10'08''$ west, and has an area of 4782.118 km^2, with an altitude of 736 m.

Each region from the dataset was manually classified into the following land-use classes: soybean, maize, cotton, beans, sorghum, NCC-millet, NCC-crotalaria, NCC-brachiaria, NCC-grasses, pasture, turf grass, eucalyptus, Cerrado, and uncultivated soil. However, in Brazil and for this region, soybean, maize, and cotton represents nearly 97% of the cultivated area and, therefore, were selected to be evaluated in our experiments, while we grouped the remaining classes in a single so-called "*others*" class, comprising four classes in total.

2) Experimental Design: The Campo Verde dataset [see Fig. 1(a)] was split into training, validation, and testing, using *K*-means to geographically separate regions and avoid physical overlapping between the three sets. Fig. 4 shows an example of partition into these sets. The networks were trained using 60% of the samples from the Campo Verde dataset, whereas 20% were used for validation and 20% for testing. The training scheme used 100 epochs, batch size of 50, and Adam optimizer with initial learning rate of 0.001 and momentum of 0.999. The loss function was the categorical cross-entropy, commonly used for multiclasses using CNNs.

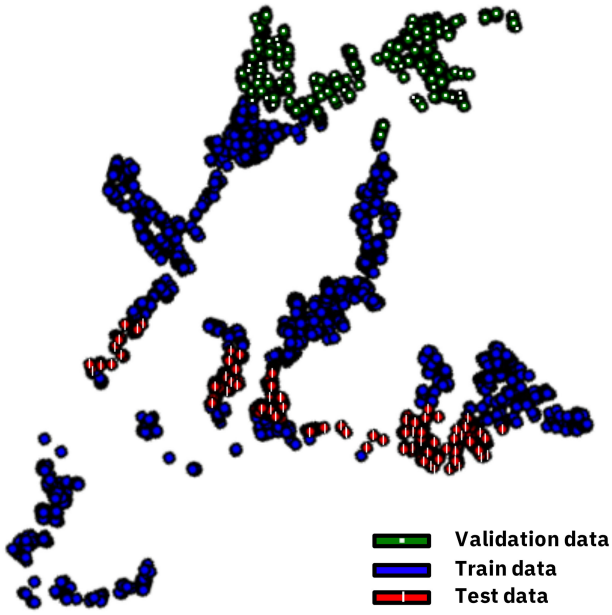


Fig. 4. *K*-means example for geographical region separation. This method can avoid physical overlapping between the three sets: validation, training, and testing.

The best model was experimentally achieved and we provide next a few different model results for roughly providing directions on parameters sensitivity. We varied the number of layers of CNN and LSTM blocks from 1 to 3, and the size of filters from 32 to 512. For the dense layer, we tested 1 and 2 layers with number of nodes varying between 10 and 100. We trained all models until convergence, adopting the early stop criteria considering validation accuracy drop.

We tested all the architectures above using 20% of the Campo Verde dataset (equivalent to 67.176 pixels), in a fivefold cross-validation schema. The criterion for choosing the best architecture was the average overall accuracy, which considers the average of accuracy metric including all dates (October 2015 to July 2016). Another way of evaluation is to compute the accuracy for each crop on the harvest peak (one date). This accuracy is computed from the confusion matrix mounted from each pixel classification. All experiments were performed using python, tensorflow, and keras.

3) Results for Campo Verde Dataset: The five best models for this first test are shown in Table I. The best model, with three CNN layers (128, 256, and 512 filters, respectively), three LSTM layers (64, 64, and 64 filters, respectively), and a single dense layer with 30 nodes, was later used for the test of estimating crop areas at the entire Campo Verde municipality. This model achieved 69.72% of overall accuracy, considering the mean value for all considered dates (between the period of October 2015 to July 2016) and a mean accuracy of 91.95% considering the dates of the harvest peaks for each crop (December 19th, May 8th, June 9th, and October 16th, for *soybean*, *maize*, *cotton*, and *others*, respectively).

Fig. 5 shows the obtained results for each crop type at each considered date. Notice that the first period (October–February) corresponds to the warmer period at the south hemisphere (when

TABLE I
DIFFERENT NN ARCHITECTURES WITH THE BEST RESULTS FOR EACH CROP AND RESULTS FOR HARVEST PEAKS USING THE MODEL WITH THE HIGHEST OVERALL MEAN ACCURACY

CNN	LSTM	Dense	Overall Mean Accuracy(%)	Soybean (%)	Maize (%)	Cotton(%)	Others(%)
[32]	[32]	[40]	64.07	**81.99***	66.83*	60.4*	47.08*
[32]	[32, 32, 32]	[30, 10]	66.95	66.03*	**69.21***	57.54*	75.02*
[128, 128, 128]	[128, 128]	[30]	67.79	63.89*	62.87*	**82.64***	61.79*
[32]	[64, 128, 256]	[30]	68.23	65.12*	59.39*	65.81*	**82.61***
[128, 256, 512]	[64, 64, 64]	[30]	**69.72**	67.94*	67.11*	66.87*	76.95*
Results for harvest peak			91.95	92.11	89.96	89.42	96.29

*The results indicate the mean value related to all computed dates.
The boldface values indicate the best architecture performance for each crop.

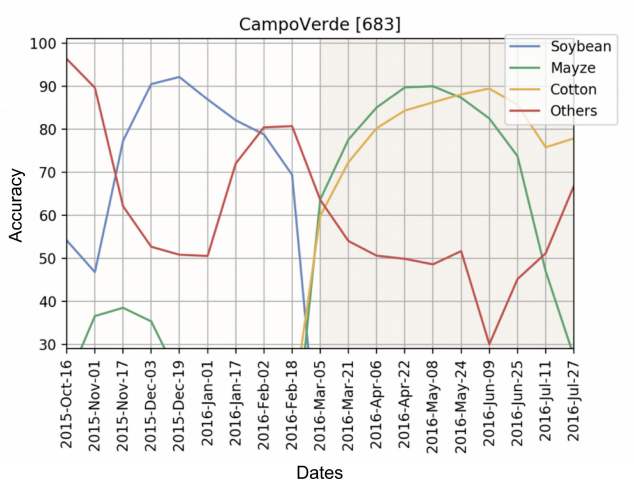


Fig. 5. Test result for Campo Verde dataset showing the performance for each class along a year.

soybean is cultivated), and the second period (March–July) corresponds to the colder period (when *maize* and *cotton* are cultivated). Achieving the best results at the harvest peaks is intuitive, once at the beginning and end of the harvest period, we have a higher amount of exposed (uncultivated) soil, which poses a challenge for the classifier. Because the *others* class comprises several different crops, no seasonal behavior is evident, and the best result for this class coincides with the beginning of the *soybean* harvest.

Indeed, if we analyze the accuracy for each individual date, the model can reach much higher accuracy values during the harvest peak of each crop type—which is quite predictable in Brazil and could be safely used. Fig. 6 shows the best-achieved results, which are 92.11% for *soybean*, 89.96% for *maize*, 89.42% for *cotton* (all values achieved at the harvest peaks), and 96.29% for *others*, an average of 91.95% of accuracy.

B. Crop-Type Area Prediction From Unlabeled Data

1) Data: For evaluating the generalization of the trained models for estimating crop area, we used an unlabeled dataset containing 2503.762 pixels (nearly seven times larger than training set), comprising the whole area of Campo Verde municipality, and the census data describing the proportions of cultivated area in the entire municipality. This census data were acquired by the Brazilian Institute of Geography and Statistics (IBGE), a Brazilian governmental company. This institute's website provides public reports with the crop area estimation for different agricultural regions in Brazil per year [25]. The method for crop area estimation used by the institute includes interviews and field visits.

TABLE II
AREA COMPUTED FROM THE PROPOSED METHOD VERSUS AREA FROM IBGE DATA CONSIDERING REGIONS WITH DIFFERENT NDVI STANDARD DEVIATION

	NDVI std	Soybean	Maize	Cotton
DL method	no threshold	276 K ha	134 K ha	98 K ha
DL method	<0.33	237 K ha	122 K ha	96 K ha
DL method	<0.45	188 K ha	99 K ha	87 K ha
IBGE data	–	210 K ha	88 K ha	79 K ha

2) Experimental Design: Using the best architecture described in Section IV-A2, we tested the entire municipality on every available date between the years of 2015 and 2016. This experiment aims to verify our main goal, which is to assess if CLSTM models trained on a limited annotated region can be used to classify crops at much larger areas.

The virtual impossibility of annotating the whole test region (no geo-referenced labels are available for the entire municipality area, which corresponds to 478 K hectares), led us to use information from a different data source available through governmental census data for cultivated areas.

For evaluation, after predicting the label for each pixel of the municipality, we computed the number of pixels for each crop and estimated the corresponding total cultivated area for each crop type based on the size of the pixel, for instance, 250 m^2. Then, we compared our results with the census data released by IBGE for each crop.

3) Results for the Entire Municipality: Using the best model previously computed, we estimated the cultivated areas for each crop on their respective harvest peak, as shown in Fig. 7. Initially, our method computed 276 K hectares of *soybean* at its harvest peak (middle of December) without filtering forests, rivers, and urban areas, against census data estimation of 210 K hectares of *soybean* from IBGE, for the same region and period. Considering our approach for removal of the nonagricultural regions, the *soybean* planted area was estimated on 188 K hectares (using 45% deviation), thus reducing the difference of the values from IBGE and our method in 44 K hectares. Table II shows the estimated area for the three crop types, with and without data removal, using 33% and 45% deviation, as explained in Section III-C. The estimated areas for each crop were calculated on the dates when the model performed best. Considering the removal of regions

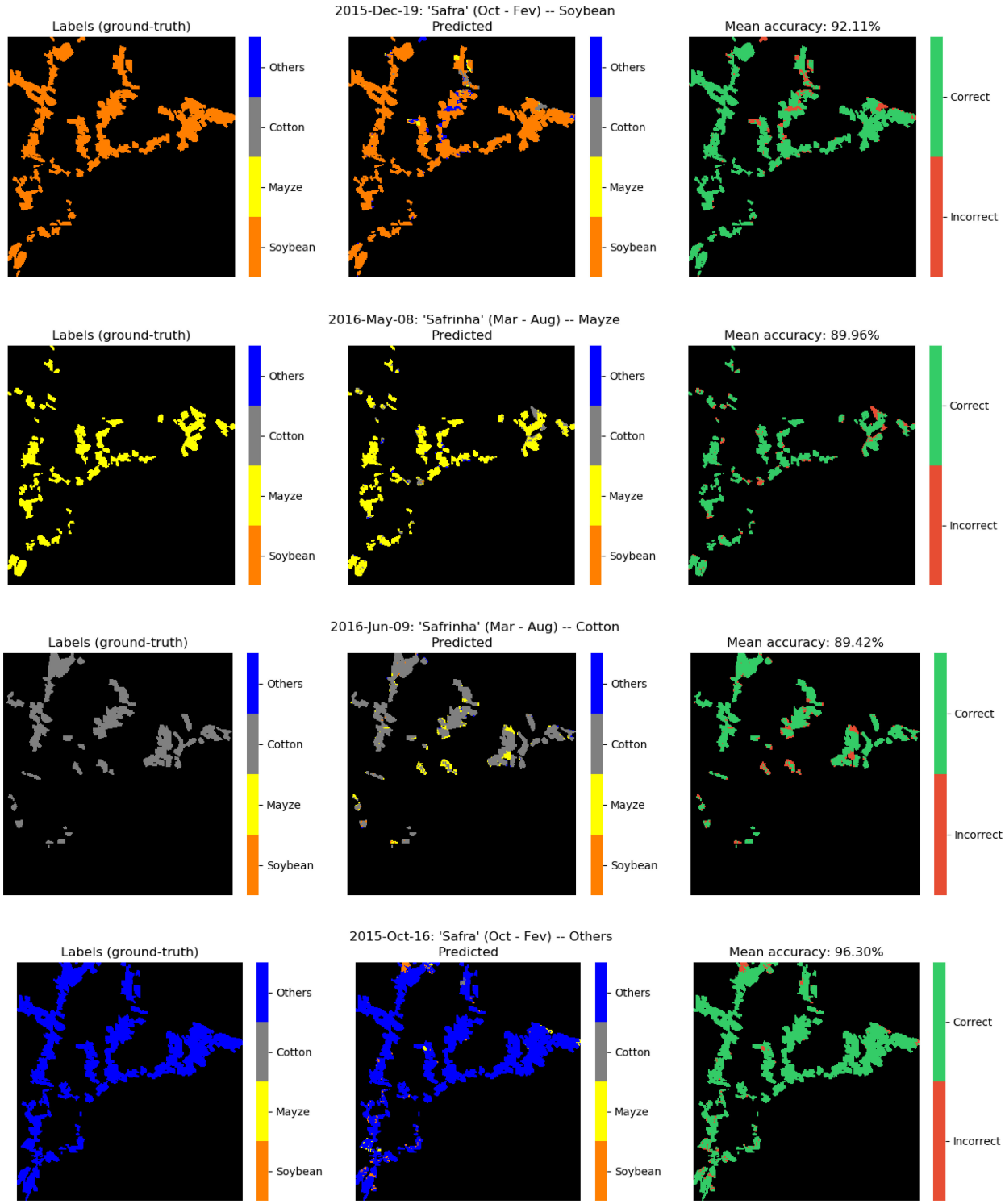


Fig. 6. CLSTM results of the Campo Verde dataset for the dates that achieved the highest accuracy values (which correspond to the harvest peaks for soybean, maize, and cotton). The ground-truth (per date) of each class is displayed on the right side; the predicted results from the automated method are shown in the middle column (different colors correspond to different crop types); and on the left side, we present maps where pixels classified correctly are displayed in green, and pixels assigned to incorrect classes are displayed in red.

with standard deviation less than 45%, we estimated crop areas with errors of 22 K, 11 K, and 8 K hectares for *soybean*, *maize*, and *cotton*, respectively.

As we only have access to the estimated area information provided by IBGE instead of the precise labeled regions of the entire municipality, it is not possible to calculate metrics such as accuracy or false positive rates for this second test. Thus, to at least evaluate the consistency of our results, we performed a bootstrap of the Campo Verde dataset, as detailed in Section III-C. The dataset was split into 30 sets, and for each

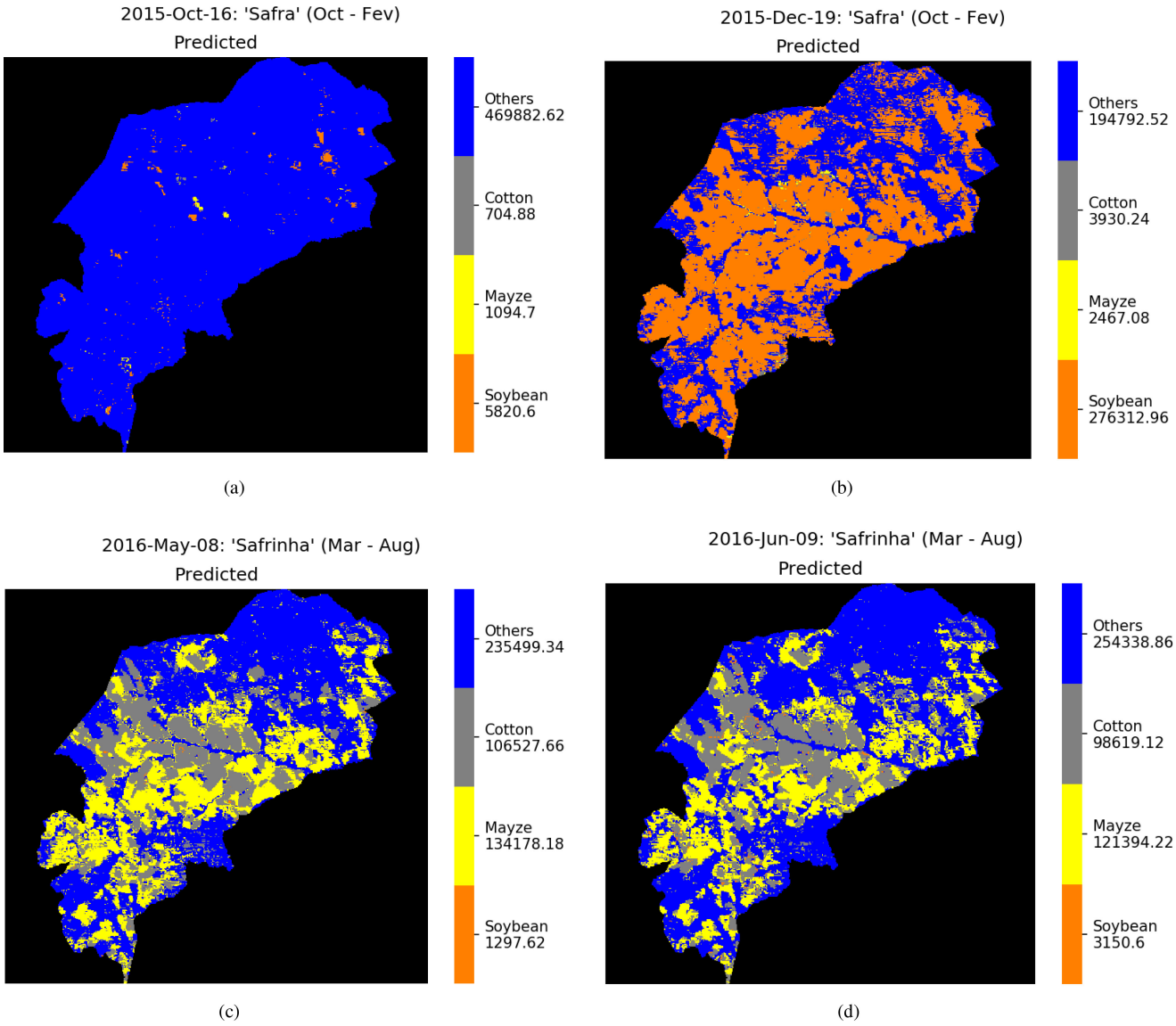


Fig. 7. Crop area estimation results on dates corresponding to the harvest peaks for each crop, computed for the entire Campo Verde municipality. The left color bars show the estimated area for each class. (a) Planting time. (b) Soybean harvest peak. (c) Maize harvest peak. (d) Cotton harvest peak.

TABLE III
BOOTSTRAP RESULTS FOR HARVEST PEAK DATE

Metric	Soybean	Maize	Cotton	Others
	2015-12-19	2016-05-08	2016-06-09	2015-10-16
	[min(%),max(%)]	[min(%),max(%)]	[min(%),max(%)]	[min(%),max(%)]
Sensitivity	[86.1, 94.2]	[63.7, 80.9]	[74.3, 86.5]	[97.8, 100]
Specificity	[67.5, 83.3]	[89.4, 97.0]	[87.5, 95.0]	[50.0, 66.7]
False discovery rate	[2.6, 11.0]	[4.6, 16.4]	[4.2, 14.4]	[1.4, 6.4]

For each metric is shown the maximum variation.

set, some metrics were computed and depicted in Fig. 8, where we can also observe their maximum variation. This allows one to have an idea of the maximum variation of area estimation metrics considering unlabeled pixelwise crop-type annotation. For instance, we can report that *soybean* identification at its harvest peak can achieve a false discovery rate of 11.0% and for *maize*, *cotton*, and *others*, the same metric can achieve 16.4%, 14.4%, and 6.4%, respectively. Table III shows maximum and minimum values computed from the bootstrap for sensitivity, specificity, and false discovery rate.

Our results indicate that CLSTM models can be safely used for estimating crop area in very large regions using relatively small training sets, supporting their use for automatic census data estimation for digital agriculture.

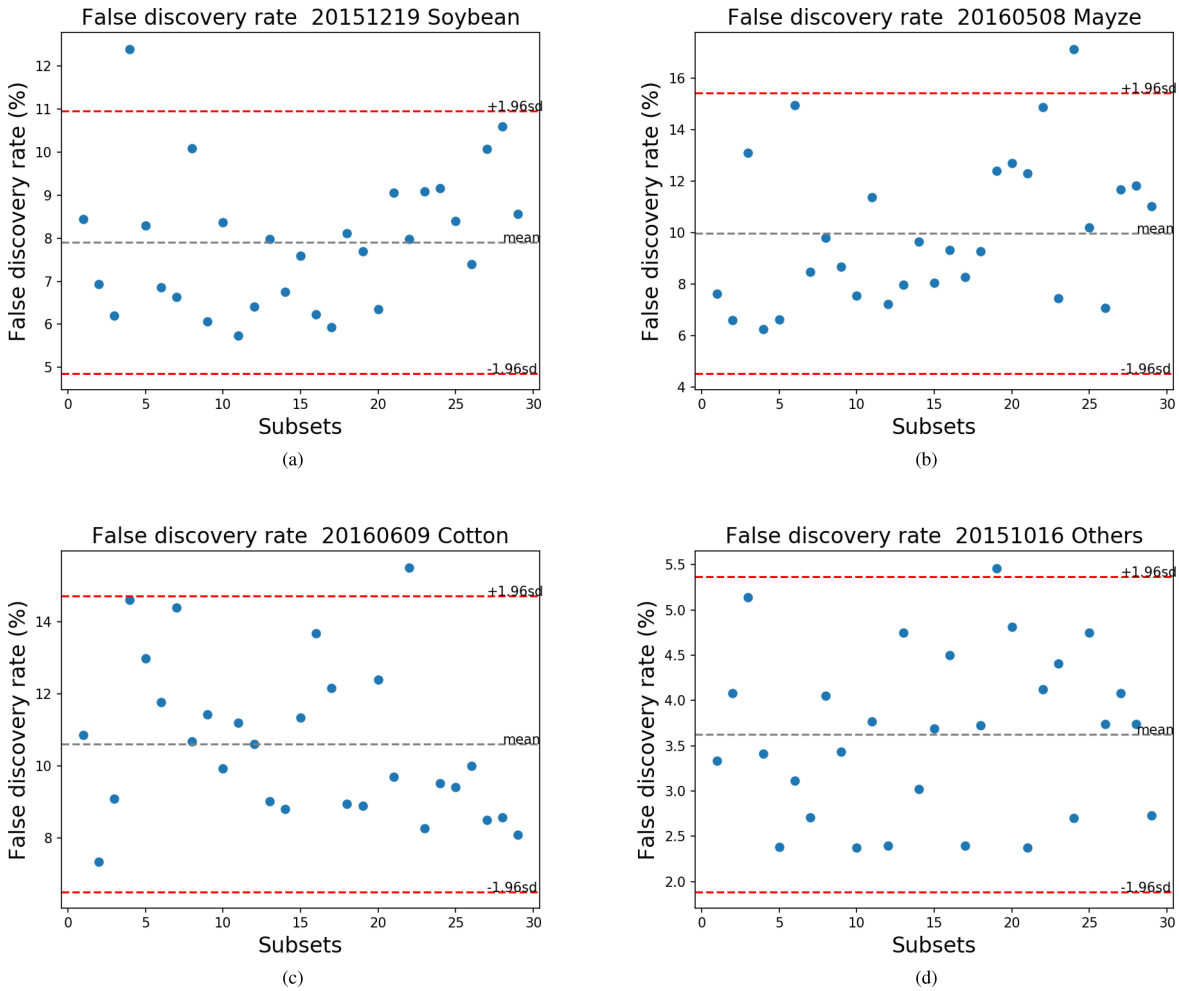


Fig. 8. Bootstrap separating 30 sets of Campo Verde dataset to compute false discovery rate for each crop in their respective harvest peak. (a) Graphic of false discovery rate for soybean crop on the harvest peak, December 19th. (b) Graphic of false discovery rate for mayze crop on the harvest peak, May 8th. (c) Graphic of false discovery rate for cotton crop on the harvest peak, June 9th. (d) Graphic of false discovery rate for other crops on the first day of satellite image capture for the harvest, October 16th.

V. Conclusion

In this article, we evaluated the use of CLSTM networks for crop identification, considering four different classes (*soybean*, *maize*, *cotton*, and *others*), as well as its further capability for generalizing crop area estimation in areas that are much larger than the ones used for training. We also evaluated how this outcome relates to government census data for agriculture in Brazil and reported consistent results with the expected behavior. Our methodology has two phases: training and testing a CLSTM model using a public-labeled dataset; and evaluating its generalization capabilities at the unlabeled entire municipality area with the support of census data for evaluation.

Training and tests of crop identification were performed in two sets with 335 880 pixels for the labeled dataset and 2503.762 pixels for the unlabeled Campo Verde municipality covering 476 K hectares from the MODIS-16 sensor, with the best result reaching an average of 91.95% of accuracy per class considering the harvest peaks of each crop (92%, 89.96%, 89.42%, and 96.29% accuracy for *soybean*, *maize*, *cotton*, and *others*, respectively).

After estimating and removing regions of forests, rivers, and urban areas, we computed crop area estimation using the best trained model and reported errors of 22 K, 11 K, and 8 K hectares for *soybean*, *maize*, and *cotton*, respectively. These results suggest that CLSTM models can be used for estimating crop area in very large regions, with very satisfactory results at harvest peak dates.

While this article focused on evaluating generalization capabilities from a scaling perspective, we recognize that a multiyear study would be fundamental to evaluate additional generalization aspects, such as climate fluctuation. In future work, we would like to assess how the CLSTM behaves when training and testing sets are acquired at different time series.

References

[1] J. Huang, H. Wang, Q. Dai, and D. Han, "Analysis of NDVI data for crop identification and yield estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 11, pp. 4374–4384, Nov. 2014.

[2] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[3] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6712–6722, Nov. 2018.

[4] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4823–4833, Jul. 2019.
[5] D. Xie, P. Sun, J. Zhang, X. Zhu, W. Wang, and Z. Yuan, "Autumn crop identification using high-spatial-temporal resolution time series data generated by MODIS and Landsat remote sensing images," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 2118–2121.
[6] J. Guo, P.-L. Wei, J. Liu, B. Jin, B.-F. Su, and Z.-S. Zhou, "Crop classification based on differential characteristics of H/α scattering parameters for multitemporal quad-and dual-polarization SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6111–6123, Oct. 2018.
[7] L. E. Cu La Rosa, P. N. Happ, and R. Q. Feitosa, "Dense fully convolutional networks for crop recognition from multitemporal SAR image sequences," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 7460–7463.
[8] J. Inglada, A. Vincent, M. Arias, and C. Marais-Sicre, "Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series," *Remote Sens.*, vol. 8, no. 5, 2016, Art. no. 362.
[9] Y. Yang *et al.*, "Geo-parcel based crop identification by integrating high spatial-temporal resolution imagery from multi-source satellite data," *Remote Sens.*, vol. 9, no. 12, 2017, Art. no. 1298.
[10] M. W. Liu, M. Ozdogan, and X. Zhu, "Crop type classification by simultaneous use of satellite images of different resolutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3637–3649, Jun. 2013.
[11] V. Sitokonstantinou, I. Papoutsis, C. Kontoes, A. L. Arnal, A. A. Andrés, and J. G. Zurbano, "Scalable parcel-based crop identification scheme using sentinel-2 data time-series for the monitoring of the common agricultural policy," *Remote Sens.*, vol. 10, no. 6, 2018, Art. no. 911.
[12] S. Feng, J. Zhao, T. Liu, H. Zhang, Z. Zhang, and X. Guo, "Crop type identification and mapping using machine learning algorithms and sentinel-2 time series data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3295–3306, Sep. 2019.
[13] J. Inglada *et al.*, "Benchmarking of algorithms for crop type land-cover maps using sentinel-2 image time series," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2015, pp. 3993–3996.
[14] N. C. Mira, J. Catalao, and G. Nico, "Multi-temporal crop classification with machine learning techniques," in *Proc. SPIE*, vol. 11149, pp. 213–224, 2019.
[15] J. Senthilnath, S. Kulkarni, J. A. Benediktsson, and X.-S. Yang, "A novel approach for multispectral satellite image classification based on the bat algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 4, pp. 599–603, Apr. 2016.
[16] S. Siachalou, G. Mallinis, and M. Tsakiri-Strati, "Analysis of time-series spectral index data to enhance crop identification over a Mediterranean rural landscape," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1508–1512, Sep. 2017.
[17] S. Valero, P. Ceccato, W. E. Baethgen, and J. Chanussot, "Identification of agricultural crops in early stages using remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2013, pp. 4229–4232.
[18] A. Kamilaris and F. X. Prenafeta-Boldu, "Deep learning in agriculture: A survey," *Comput. Elect. Agriculture*, vol. 147, pp. 70–90, 2018.
[19] J. D. B. Castro, R. Q. Feitoza, L. C. L. Rosa, P. M. A. Diaz, and I. D. A. Sanches, "A comparative analysis of deep learning techniques for subtropical crop types recognition from multitemporal optical/SAR image sequences," in *Proc. SIBGRAPI Conf. Graph., Patterns, Images*, Oct. 2017, pp. 382–389.
[20] L. E. Cue La Rosa, R. Q. Feitosa, P. N. Happ, I. Sanches, and G. A. O. P. da Costa, "Combining deep learning and prior knowledge for crop mapping in tropical regions from multitemporal SAR image sequences," *Remote Sens.*, vol. 11, no. 17, 2019, Art. no. 2029. [Online]. Available: https://www.mdpi.com/2072-4292/11/17/2029
[21] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sens. Environ.*, vol. 221, pp. 430–443, Feb. 2019.
[22] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.
[23] M. Russwurm and M. Krner, "Multi-temporal land cover classification with sequential recurrent encoders," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 4, Mar. 2018, Art. no. 129.
[24] I. Del'Arco Sanches *et al.*, "Campo verde database: Seeking to improve agricultural remote sensing of tropical areas," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 369–373, Mar. 2018.
[25] IBGE, "SIDRA—Agricultural production system," 2017. [Online]. Available: https://sidra.ibge.gov.br/Tabela/1612, Accessed: Jan. 8, 2020.
[26] NASA, "MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. 2015, distributed by NASA," 2015. [Online]. Available: https://lpdaac.usgs.gov/products/mod13q1v006/, Accessed: Jan. 8, 2020.
[27] Q. Liu, F. Zhou, R. Hang, and X. Yuan, "Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification," *Remote Sens.*, vol. 9, pp. 17–28, Mar. 2017.
[28] S. A. Rahman and D. A. Adjeroh, "Deep learning using convolutional LSTM estimates biological age from physical activity," *Scientific Rep.*, vol. 9, no. 1, 2019, Art. no. 11425. [Online]. Available: https://doi.org/10.1038/s41598-019-46850-0

Maysa Malfiza Garcia de Macedo received the B.Sc. degree in computer science from the State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil, in 2001, the M.Sc. degree in computer science from Computer Science Institute, Federal Fluminense University (UFF), Rio de Janeiro, in 2005, and the Ph.D. degree in computer science from the University of Sao Paulo, Sao Paulo, Brazil, in 2012.

She is currently a Research Scientist with IBM Research, Sao Paulo. Her research focuses on image processing for multidimensional biomedical imaging, computer vision, and machine learning.

Andrea Britto Mattos received the B.Sc. and M.Sc. degrees in computer science from the Institute of Mathematics and Statistics, University of Sao Paulo (IME/USP), São Paulo, Brazil, in 2008 and 2011, respectively, working on three-dimensional registration and structural graph matching.

She is currently a Researcher with the Visual Analytics & Insights group, IBM Research, Brazil, working on projects in the areas of computer vision and machine learning.

Dário Augusto Borges Oliveira received the B.Sc. degree in electrical engineering from the Rio de Janeiro State University, Rio de Janeiro, Brazil, in 2007, and the M.Sc. and Ph.D. degrees in electrical engineering from the Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, in 2009 and 2013, repectively.

He is currently a Researcher with the Visual Analytics & Insights group, IBM Research Brazil, working on projects in the areas of computer vision and machine learning.