Spatio-Temporal Vegetation Pixel Classification By Using Convolutional Networks

Keiller Nogueira, Jefersson A. dos Santos, Nathalia Menini, Thiago S. F. Silva, Leonor Patricia C. Morellato, and Ricardo da S. Torres

***Abstract*—Plant phenology studies rely on long-term monitoring of life cycles of plants. High-resolution unmanned aerial vehicles (UAVs) and near-surface technologies have been used for plant monitoring, demanding the creation of methods capable of locating and identifying plant species through time and space. However, this is a challenging task given the high volume of data, the constant data missing from temporal dataset, the heterogeneity of temporal profiles, the variety of plant visual patterns, and the unclear definition of individuals' boundaries in plant communities. In this letter, we propose a novel method, suitable for phenological monitoring, based on Convolutional Networks (ConvNets) to perform spatio-temporal vegetation pixel-classification on high resolution images. We conducted a systematic evaluation using high-resolution vegetation image datasets associated with the Brazilian Cerrado biome. Experimental results show that the proposed approach is effective, overcoming other spatio-temporal pixel-classification strategies.**

***Index Terms*—Deep Learning, Pixel Classification, Unmanned Aerial Vehicles, Near-Surface, Phenology.**

I. Introduction

Plant phenology is the science dedicated to the study of the life cycles of plants and their relations with climate conditions. One of the most important challenges faced by *remote* phenology studies refers to the location (identification) of plant individuals of interest within images. Specially, in tropical regions, plant individuals of a large number of species co-exist *spatially* (community level) over time. Also, as species phenological traits for different species often differ from each other, species *temporal* profiles are expected to encode relevant discriminative properties. Properly combining spatial and temporal cues are, therefore, of paramount importance for the creation of effective vegetation pixel-classification methods.

Spatio-temporal vegetation pixel classification, however, is a very challenging task [1], as it requires to handle: (i) high volumes of temporal data; (ii) missing data [2], which is a common aspect of temporal datasets; (iii) heterogeneity of temporal patterns, due to differences between seasons, which impact on, for instance, the leafing and flowering of plants; (iv) variety of plant patterns, given the high intraclass variance and high interclass similarity of species; and (v) unclear definition of individuals' boundaries, as canopies often overlap.

Several studies have been addressing those challenges [3]–[6]. Some of them [3], [7], [8] employ general-purpose handcrafted descriptors to represent image regions. Those methods are data dependent, requiring a full set of experiments to determine the most suitable descriptors for each application [9]. Other initiatives [4]–[6] rely on deep learning approaches [10], such as ConvNets (or CNN), which are capable of learning, at once, data-driven features and classifiers. However, those studies [4]–[6], [8] often focus on spatio-temporal pixel classification of land-use datasets. To the best of our knowledge, none of those deep learning-based initiatives have been exploiting spatio-temporal properties of high-resolution vegetation images in pixel classification problems.

In this paper, we propose a deep-learning-based technique that fills this gap. In other words, we propose a novel ConvNet for vegetation pixel classification in high-resolution temporal images acquired by UAVs and near-surface digital cameras. The method can effectively learn a combined representation of images using distinct spatio-temporal properties. Specifically, the proposed network has initial branches, which are responsible for learning spatial patterns from images of a specific timestamp. All branches are unified into a unique network, which is in charge of combining information from previously learned spatial properties, creating a new spatio-temporal representation. Since the whole process is integrated into a single framework that receives several patches (one for each timestamp) and outputs a single class label, the network can be trained end-to-end using the well-known backpropagation algorithm [10]. Furthermore, since each branch handles a specific timestamp, the proposed approach can be easily adapted to handle missing data [2] by using dropout [11].

II. Multi-Temporal ConvNet

The proposed approach, called Multi-Temporal ConvNet, is based on pixelwise classification [9]. In this technique, each pixel of the image is independently processed and classified by the network. Since the pixel itself has not enough information to be extracted from the network, it is usually represented by a context window, which aggregates spatial information to the learning process. Technically, a *context window* is a

K. Nogueira, and J. A. dos Santos are with the Department of Computer Science, Universidade Federal de Minas Gerais (UFMG), Brazil. Emails: {keiller.nogueira, jefersson}@dcc.ufmg.br
Nathalia Menini and Ricardo da S. Torres are with Institute of Computing, University of Campinas, Brazil.
L. P. C. Morellato is with Instituto de Biociências (IB), Universidade Estadual Paulista (Unesp), Brazil.
T. S. F. Silva is with IGCE/Unesp, Brazil and with Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, UK.

The authors thank the Pró-Reitoria de Pesquisa da UFMG; FAPEMIG (APQ-00449-17); São Paulo Research Foundation FAPESP (2013/50155-0, 2013/50169-1, 2009/54208-6, 2016/26170-8, 2018/06918-3); CNPq (424700/2018-2); CNPq Research fellowship to JAS, LPCM, RST, TSFS; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 (88881.145912/2017-01); the Cedro Têxtil, Reserva Vellozia, Parque Nacional da Serra do Cipó, PELD-CRSC-17.

fixed-size patch with the pixel that the method should classify centered on it. Overall, the context windows, generated for each and every pixel of the image, are delivered to the ConvNet, which processes them as standard images, outputting a unique classification which is related to the centered pixel of the currently processed window.

Networks proposed for pixelwise classification aggregate only spatial information. Our vegetation pixel-classification method proposes to incorporate temporal information into pixelwise classification networks in order to learn a spatio-temporal model. Precisely, instead of receiving only one context window, the proposed model receives as input several context windows, all of them centered on the same pixel along the temporal domain. These windows are processed by several branches, each one related to an entry in the temporal domain. Each branch receives a single context window that is directly connected to the *timestamp*, i.e., there is an exact match between the branch and context window (since both should be related to the same timestamp). These branches are directly responsible for learning spatial patterns related to the respective *timestamp*. All these branches are then depth-wise concatenated and processed by further layers of the network, responsible for combining the previously learned information by extracting patterns related to the whole time series and creating a spatio-temporal representation. This final representation is, then, used to classify the pixel centered on the context windows. This whole process is integrated into a unique network that receives several patches and outputs a class label, allowing the ConvNet to be trained end-to-end using backpropagation [10].

A. Network Architecture

The multi-temporal ConvNet architecture, proposed to perform spatio-temporal pixel-classification of vegetation images, is presented in Figure 1. This network receives as input 25×25 context windows. These patches (one for each timestamp) are processed by their respective network branch, which is responsible for capturing the spatial information of that specific timestamp. All of these branches have a convolution (with 64 neurons and kernel of size 4×4) and a max-pooling operation (with both kernel and stride of size 2×2). Then, all branches are depth-wise concatenated and further processed by two other layers, both composed of convolution and max-pooling operations. In these layers, the convolutions have 128 (with 4×4 kernels) and 256 (with 3×3 filters) neurons, respectively. Both max-pooling operations use the same kernel (2×2) but differ in the stride: the first one uses 2×2 while the second employs 1×1. Finally, fully connected layers create the final classification of the pixel centered on the input context windows. Rectified Linear Unit (ReLU) [10] was the processing units used in all layers. Our implementation also includes: (i) batch normalization [10], which is employed after each convolution, and (ii) dropout [11], which is mainly used within the fully connected layers to avoid overfitting.

B. Missing Data Adaptation

Since the proposed method has a branch for each timestamp, it can be easily adapted to handle one of the challenging problems of temporal series: missing data [1]. Very common in spatio-temporal datasets, this issue is defined as the lack of stored information for a timestamp of the time series. This problem can have a significant effect on the conclusions that can be drawn from the data [2]. In order to handle this crucial problem, we propose to use dropout [11].

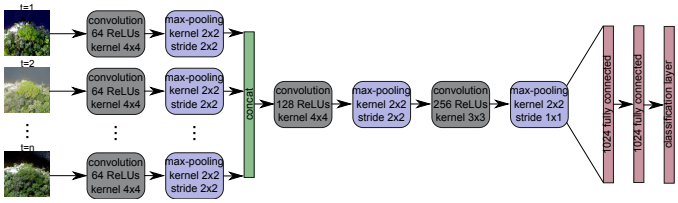


Fig. 1. ConvNet architecture for spatio-temporal pixel classification.

Dropout [11] is a regularization technique that drops some neurons during the **learning** process. Technically, for each learning iteration of the network, dropout method randomly selects (based on a predefined probability, which is usually 50%) a set of neurons that: (i) do not contribute (i.e., propagates zero) during the forward step, and (ii) do not receive updates during the backpropagation phase. Therefore, this set of neurons does not participate in the learning process during an iteration. This allows the creation of an ensemble of models that share parameters, improving the representation learned by the network. During the **inference** step, all neurons are preserved and contribute to the final outcome of the network.

In order to handle missing data, we proposed several small changes in the method and in the dropout technique. One change is the inclusion of a dropout layer just before the concatenation layer, which is responsible for dropping some neurons (or, in a broader view, branches). This new layer acts in its traditional way during the learning phase with the probability (to randomly select the neurons that are preserved) set to $1/t$, where t is the temporal length. This value is defined based on the worst scenario of missing data, i.e., when there is only 1 available timestamp in the whole temporal domain t. Hence, by setting this probability following this idea, we force the network to learn a model that should extract all feasible information from the available data without being hampered by the missing data (even if it is the worst scenario). Although working traditionally during the learning phase, dropout layer works differently during the prediction step. Instead of preserving all neurons, the dropout layer drops the branches (or timestamps, since each branch is correlated to a timestamp) with missing data, which do not contribute to the final result.

III. Experimental Setup

A. Phenological Temporal Datasets

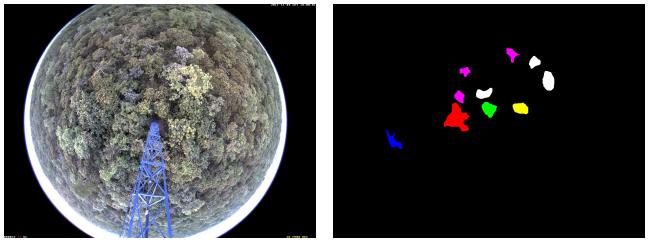
1) Itirapina Dataset: The Itirapina dataset [3] comprises images collected with a near-remote phenological system composed of a camera set up in an 18m tower in a Cerrado sensu stricto, a savanna-like vegetation located at Itirapina, São Paulo State, Brazil. This camera was set up to automatically take hourly photos (at 1280×960 pixels of resolution) from 6:00 to 18:00 h (UTC-3) between August 29th and October

TABLE I
ITIRAPINA DATASET

Class	#pixels
Aspidosperma tomentosum	4,040
Caryocar brasiliensis	6,601
Myrcia guianesis	2,630
Miconia rubiginosa	6,715
Pouteria ramiflora	2,373
Pouteria torta	2,037

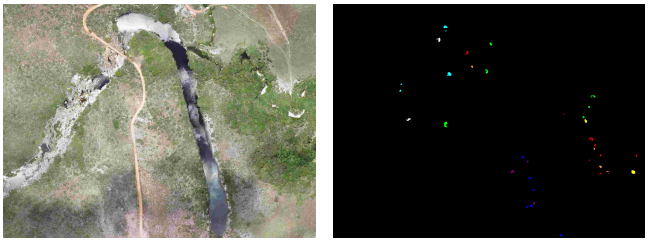
TABLE II
SERRA DO CIPÓ DATASET.

Class	#pixels
Vochysia cinnamomea	34,754
Eremanthus erythropappus	31,250
Bowdichia virgilioides	33,137
Set of evergreen species	48,041



(a) 18h - October 3rd, 2011 (b) Ground-Truth

Fig. 2. Image of the Itirapina dataset and the ground truth. Legend – Black: unclassified/background. White: *Miconia rubiginosa*. Red: *Caryocar brasiliensis*. Green: *Myrcia guianesis*. Pink: *Aspidosperma tomentosum*. Blue: *Pouteria torta*. Yellow: *Pouteria ramiflora*.



(a) December 24th, 2016 (b) Ground-Truth

Fig. 3. Image of the Serra do Cipó dataset and the groundtruth. Legend – Black: unclassified/background. Red (train)/Orange (test): *Bowdichia virgilioides*. Green (train)/Yellow (test): Collection of evergreen species. Blue (train)/Purple (test): *Eremanthus erythropappus*. Cyan (train)/White (test): *Vochysia cinnamomea*.

3rd, 2011 (resulting in **36 days or timestamps**). These images were then labeled by experts as belonging to one of the six possible plant species, as presented in Table I. Example of a timestamp and the ground truth are presented in Figure 2.

2) Serra do Cipó Dataset: Images of this dataset were acquired by a Canon SX260 RGB camera aboard a fixed-wing UAV from October 2015 to February 2017, one for each month (except January and December 2016), totaling **15 images (or timestamps)**. The acquired aerial photographs were mosaiced into a single orthoimage, with $6,786 \times 9,069$ pixels, that was then labeled by experts as belonging to one of the four possible plant species, as shown in Table II. The imaged vegetation comprises a *campo rupestre* vegetation of the Brazilian Cerrado biome, including a mixture of grasses, shrubs, and trees growing over sandy or rocky substrate. Specifically, all images were acquired over the Serra do Cipó region, a mountainous and highly biodiverse and heterogeneous landscape in southern-central Brazil. Example of a timestamp and the ground truth are presented in Figure 3.

B. Baselines

Several techniques were considered as baselines for both datasets: (i) recurrence plots (RP) [12]–[14], which are first calculated over the temporal series and then described by hand-crafted descriptors (such as LBP histograms) to be then finally classified by machine learning methods (Linear SVM and Random Forest (with 400 trees), in this case). (ii) MC-DCNN [15], a deep-learning-based method that learns features for each and every single channel (or band) and then, concatenates them to be used in the classifier layer, and (iii) 2-D CNN [5], another deep learning approach that performs pixel-classification using a temporal-wise concatenation of image.

For the Itirapina dataset, the method proposed in [7] was also employed as baseline. In that work, temporal series are described using hand-crafted (color and texture) features, which are provided to a multi-scale classifier (**MSC**) responsible for performing a late-fusion classification of input images.

C. Experimental Protocol

For the Itirapina dataset, the protocol of [7] was employed. Particularly, some instances of *Aspidosperma tomentosum* and of *Miconia rubiginosa* were considered for testing. Pixels of these instances have no influence on the model which is trained using exclusively the remaining samples. Since this dataset comprises 36 days, the proposed network has 36 branches, one for each day, that receive as input a depth-wise concatenation of all hourly images of that day.

For the Serra do Cipó dataset, two protocols were used to validate the effectiveness of the proposed method. The **first one** follows the same concept of the protocol employed in the Itirapina dataset. Specifically, all segments (instances) of this dataset were randomly divided into two independent sets: train and test. The former set is composed of approximately 80% of the annotated instances while the latter has the remaining 20%. Both sets have instances of all classes and the test set has, at least, two segments of each class. Aside from this, the first protocol considers the whole temporal series as input. In this case, the proposed network has 15 input branches, one for each timestamp (or month).

The **second protocol** aims to analyze the robustness of the proposed method to missing data. Moreover, this protocol is based on the phenology study and its cyclic temporal series. Precisely, phenological monitoring is cyclic, as plants have a well-defined life cycle along the months. Therefore, in this protocol, the proposed method is trained and validated using different cycles of the plants (which have missing data). This process aims at learning and understanding the full life cycle of the plants, creating a model resilient to plant changes over the years. Specifically, in this protocol, the network is designed to have the dropout layer (with probability $1/12$) and 12 input branches, one for each month, completing the full cycle of a year. In this case, all images from the largest sequence of time series (i.e., 10 images from 2016, which, compared to a full cycle of the year, has 2 missing data) were used to train the network, while the second largest sequence (i.e., 3 images of 2015, which has 9 missing timestamps) is used for testing.

All results reported are in terms of average accuracy, computed as the average (per-class) ratio of correctly classified samples. Thus, it is calculated for each class and, therefore, is independent of any bias connected to class size. Note that, for both datasets, only annotated pixels are used, while background ones are ignored. Hence, the results report the average accuracy related only to annotated pixels.

The proposed technique was created using TensorFlow framework[1] and the code has been made publicly available at https://github.com/keillernogueira/spatio-temporal-phenological-segmentation/. All experiments were performed on a 64 bit Intel i7 5820K machine with 3.3GHz of clock, 64GB of RAM memory and a GeForce GTX Titan X with 12GB of memory under a 9.0 CUDA version. Ubuntu version 16.04.4 LTS was used as operating system. During training, aforementioned protocols used the following hyper-parameters: learning rate, weight decay, momentum, and number of iterations are 0.01, 0.0005, 0.9, and 200,000, respectively. After every 50,000 iterations, the learning rate is reduced using the exponential decay[2].

IV. Experimental Results

A. Time Series Classification

For the Itirapina dataset, obtained results are presented in Table III. The worst results were yielded by the methods based on RP-based technique [13]. Such approaches take only a few hours to be trained but are outperformed by all other approaches. This may be justified by the fact that those techniques are based on hand-crafted features which are used to train a single model. This is the same reason why the MSC [7] method, which achieves good results (76.00% and 90.00% of accuracy for *Aspidosperma* and *Rubiginosa* classes, respectively), is outperformed by the ConvNets. Precisely, this method, that requires approximately 8 hours of training to create an ensemble of models, also does not have feature learning but only combines hand-crafted features. Among the deep learning-based methods, MC-DCNN [15] achieved great results (96.74% and 99.17% of accuracy for *Aspidosperma* and *Rubiginosa* classes, respectively) taking only 12 hours to train. However, all these techniques were outperformed by the 2-D CNN [5] and the proposed approach, which achieved the best result for both classes (100.00%), taking 24 and 16 hours to train, respectively.

For the Serra do Cipó dataset, results, based on the first protocol, are presented in the first part of Table IV. Again, the worst results were produced by the methods based on Recurrence Plots technique [13]. Although achieving the worst accuracy such methods are the fastest to train, taking only 2 hours. The MC-DCNN [15] method, which requires 3 hours to train a new model, achieved great results (97.83%). However, the best outcomes were yielded by the 2-D CNN [5] and the proposed method, which achieved almost perfect classification (around 99%) taking approximately 4,5 hours.

[1]http://tensorflow.org/ (As of October 2018).

[2]https://www.tensorflow.org/api_docs/python/tf/train/exponential_decay (As of October 2018).

TABLE III
Results for the Itirapina dataset.

Method	Training Time (h)	Accuracy (%)	
		Aspidosperma	Rubiginosa
Recurrence Plots [13] + SVM	3	64.35	73.15
Recurrence Plots [13] + RF	3	65.14	86.01
MSC [7]	8	76.00	90.00
MC-DCNN [15]	12	96.74	99.17
2-D CNN [5]	24	**100.00**	**100.00**
Multi-temporal ConvNet (ours)	16	**100.00**	**100.00**

TABLE IV
Results for the Serra do Cipó dataset.

Protocol	Method	Training Time	Accuracy (%)
	Recurrence Plots [13] + SVM	2	57.12
	Recurrence Plots [13] + RF	2	60.96
Protocol 1	**MC-DCNN [15]**	3	97.83
	2-D CNN [5]	4	99.33
	Multi-temporal ConvNet (ours)	5	**99.78**
	Recurrence Plots [13] + SVM		26.39
	Recurrence Plots [13] + RF		27.96
Protocol 2	**MC-DCNN [15]**		34.11
	2-D CNN [5]		35.24
	Multi-temporal ConvNet (ours)		**50.70**

Two interesting facts can be pointed out based on the obtained results: (i) 2-D CNN [5] and the proposed approach achieved very similar results and should be better analyzed, and (ii) such methods produced (almost) perfect classifications, which clearly motivates an ablation study to analyze if all timestamps are necessary. Both issues are better analyzed next.

B. Ablation Study

An ablation study was conducted to assess the robustness of methods that produced perfect classification. It can be divided into three steps: (i) for each analyzed method, a model is trained for each timestamp, (ii) a correlation analysis (based on [16]) is performed for all trained models, and (iii) the most suitable images (timestamps) (with the lowest correlation) are then combined and used to train new models. Note that the combination of the most suitable images is performed from the smallest possible time series (with only two images) to the largest (the whole time series, as in the previous section).

For the Itirapina dataset, as presented in Table V, the best result of the 2-D CNN [5] was achieved using the image of September 10, 2011. However, the proposed method achieved a better result using the same image for the *Aspidosperma* class, whereas reproducing the same perfect classification for the *Rubiginosa* class. The same outcome is generated analyzing the models trained with the proposed approach. In this case, the method achieved perfect classification for both classes using only a timestamp (August 30, 2011). The 2-D CNN [5] trained using the same image achieved perfect classification for the *Rubiginosa* class but worse for the *Aspidosperma* one. This shows the effectiveness of the proposed method even when using only one timestamp.

As presented in Table VI, for the Serra do Cipó dataset, the 2-D CNN [5] and the proposed approach achieved almost

TABLE V
RESULTS FOR THE ABLATION STUDY OVER THE ITIRAPINA DATASET.

Time Series	Method	Accuracy (%)	
		Aspidosperma	Rubiginosa
Sept, 10 2011	**2-D CNN [5]**	94.72	**100.00**
	Multi-temporal ConvNet (ours)	**100.00**	**100.00**
Aug, 30 2011	**2-D CNN [5]**	59.55	**100.00**
	Multi-temporal ConvNet (ours)	**100.00**	**100.00**

TABLE VI
RESULTS FOR THE ABLATION STUDY OVER THE SERRA DO CIPÓ DATASET.

Time Series	Method	Accuracy (%)
Oct, 2015; Feb, 2016; Jun, 2016; Aug, 2016; Oct, 2016; Jan, 2017	**2-D CNN [5]**	99.26
	Multi-temporal ConvNet (ours)	**99.55**
Oct, 2015; Feb, 2016; Oct, 2016; Jan, 2017	**2-D CNN [5]**	97.41
	Multi-temporal ConvNet (ours)	**99.51**

perfect segmentation using 6 and 4 timestamps, respectively. Moreover, the results of the 2-D CNN [5] (99.26%) is slightly outperformed by the proposed approach (99.55%) when using 6 timestamps. On the other hand, the 2-D CNN [5], using only 4 timestamps, is not able to produce satisfactory results (97.41%), like the ones produced by the proposed technique (99.51%). Such results validate previous conclusions that the proposed method can learn a discriminative representation.

C. Missing Data

To analyze the robustness of the proposed method to missing data, we conducted experiments employing the second protocol, proposed in Section III-C, which based on the life cycle of the plants. To allow the evaluation of the baselines, we proposed an adaptation, based on dropout method [11], just as in the proposed method. All approaches were trained considering a year cycle aiming at learning and understanding the full life cycle of the plants. Therefore, all 10 images captured in 2016 were used to train the network, while the 3 images from 2015 were employed for testing. Results are presented in the second part of Table IV. Once again, the worst result was yielded by the Recurrence Plots technique [13] (around 27%). Baselines based on ConvNets outperformed such technique achieving very similar performance (around 35%). However, all baselines were outperformed by the proposed method, that yielded 50.70% of average accuracy. This result shows the ability of the proposed method in: (i) learning a spatio-temporal representation to perform vegetation pixel-classification while dealing with missing data, and (ii) capturing the cycles of plants while being resilient to their changes over the year.

V. CONCLUSION

We proposed a novel convolutional network to perform spatio-temporal pixel-classification over high-resolution vegetation images. Specifically, the network has two parts: the first one is composed of branches (in which images of each timestamp are independently processed) responsible for extracting spatial information from each entry in the time series, while the second one receives and combines all previous information extracted by the branches generating a final spatio-temporal representation. Experimental results showed the effectiveness of the proposed method to perform spatio-temporal pixel-classification. In fact, an ablation study showed that the proposed technique achieved state-of-the-art performance, in terms of average accuracy, in two temporal datasets of plant species outperforming traditional and deep learning-based baselines. Furthermore, experiments showed that the proposed approach has support to time series with missing data, a common aspect in temporal datasets. As future work, we intend to better evaluate the robustness of the proposed method to deal with missing data, and analyze the proposed method in other applications.

REFERENCES

[1] K. Anderson and K. J. Gaston, "Lightweight unmanned aerial vehicles will revolutionize spatial ecology," *Frontiers in Ecology and the Environment*, vol. 11, no. 3, pp. 138–146, apr 2013.

[2] P. J. García-Laencina, J.-L. Sancho-Gómez, and A. R. Figueiras-Vidal, "Pattern classification with missing data: a review," *Neural Computing and Applications*, vol. 19, no. 2, pp. 263–282, Mar 2010.

[3] J. Almeida, J. A. dos Santos, B. Alberton, L. P. C. Morellato, and R. da S. Torres, "Phenological visual rhythms: Compact representations for fine-grained plant species identification," *Pattern Recognition Letters*, vol. 81, pp. 90–100, 2016.

[4] N. Di Mauro, A. Vergari, T. M. Basile, F. G. Ventola, and F. Esposito, "End-to-end learning of deep spatio-temporal representations for satellite image time series classification," *Proceedings of the ECML/PKDD Discovery Challenges*, 2017.

[5] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 778–782, May 2017.

[6] H. Song, Q. Liu, G. Wang, R. Hang, and B. Huang, "Spatiotemporal satellite image fusion using deep convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 821–829, Mar 2018.

[7] J. Almeida, J. A. dos Santos, B. Alberton, R. S. Torres, and L. P. Morellato, "Applying machine learning based on multiscale classifiers to detect remote phenology patterns in cerrado savanna trees," *Ecological Informatics*, vol. 23, no. 0, pp. 49–61, 2014.

[8] P. Schäfer, D. Pflugmacher, P. Hostert, and U. Leser, "Classifying land cover from satellite images using time series analytics," in *EDBT/ICDT Workshops*, 2018, pp. 10–15.

[9] K. Nogueira, M. Dalla Mura, J. Chanussot, W. R. Schwartz, and J. A. dos Santos, "Learning to semantically segment high-resolution remote sensing images," in *International Conference on Pattern Recognition*, Dec 2016, pp. 3566–3571.

[10] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[11] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[12] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhysics Letters*, vol. 4, no. 9, p. 973, 1987.

[13] F. A. Faria, J. Almeida, B. Alberton, L. P. C. Morellato, and R. da Silva Torres, "Fusion of time series representations for plant recognition in phenology studies," *Pattern Recognition Letters*, vol. 83, pp. 205–214, 2016.

[14] V. M. A. de Souza, D. F. Silva, and G. E. A. P. A. Batista, "Extracting texture features for time series classification," in *International Conference on Pattern Recognition*, Aug 2014, pp. 1425–1430.

[15] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, "Time series classification using multi-channels deep convolutional neural networks," in *International Conference on Web-Age Information Management*. Springer, 2014, pp. 298–310.

[16] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine learning*, vol. 51, no. 2, pp. 181–207, 2003.