EVALUATING THE SEPARABILITY BETWEEN DRY TROPICAL FORESTS AND SAVANNA WOODLANDS IN THE BRAZILIAN SAVANNA USING LANDSAT DENSE IMAGE TIME SERIES AND ARTIFICIAL INTELLIGENCE

H. N. Bendini [1]*, L. M. G. Fonseca [1], B. M. Matosak [1], R. F. Mariano [1], R. F. Haidar [2], D. M. Valeriano[1]

[1] INPE, National Institute for Space Research, 12227-010, Sao Jose dos Campos, Brazil - (hugo.bendini, leila.fonseca, bruno.matosak, dalton.valeriano)@inpe.br, ravimariano@hotmail.com
[2] Dept. of Environmental Engineering, Federal University of Tocantins, Brazil - ricardohaidar@yahoo.com.br

KEY WORDS: Dry Forests, Cerrado, Recurrent Neural Networks, Random Forest, Machine Learning.

ABSTRACT: The Brazilian Savanna is the second largest biogeographical region in Brazil and present different vegetation types, consisting mostly of tropical savannas, grasslands, and forests. The forest types have different tree cover and floristic composition, which is associated to leaf deciduousness. Considering the importance of Cerrado to biodiversity conservation and the maintaining of environmental services, the development of methods to map the different forest types in Cerrado is important for conservation programmes, subsidize restauration plains, and to allow estimations of carbon sink and stock. Mapping heterogeneous tropical areas, such as the Brazilian Savanna, is very complex due to the natural factors and peculiarities of the vegetation types, and it's still particularly challenging to separate between different forest formations. In this study we tested machine learning approaches based on the use of dense image time series, in order to evaluate the separability Dry Tropical Forests and Savanna woodlands. We considered the Brazilian State of Tocantins as the study area, which is located in the Northern region of the country. RF classification of Landsat dense time series showed an overall accuracy of 0.85005, while the LSTM approach presented an overall accuracy of 0.88601, with the highest f1-score for the savanna woodlands class, suggesting the capability of the recurrent neural networks on handling complex long-term dependencies such as the EVI dense time series data. This study showed the potential for the development of a semi-automatic method for discriminating the different types of forest formations in the Brazilian Savanna, based on remote sensing.

1. INTRODUCTION

Considered to be one of the richest and diverse savannas in the world, the Brazilian Savanna, also known as Cerrado, is the second largest biogeographical region in Brazil. It has different vegetation types consisting mostly of tropical savannas, grasslands and forests (Ribeiro and Walter, 2008), and it covers approximately 24% of the country (MMA, 2021).

Furthermore, the Brazilian Savanna is considered a global biodiversity hotspot for conservation (Myers et al. 2000), and houses many endemic and threatened species (Colli et al. 2020). The Cerrado provide important environmental services such as climate regulation and water supply to different regions in Brazil (Oliveira et al. 2015; Strassburg et al. 2017). It also contributes to 43% of surface water outside the Amazon in Brazil (Strassburg et al. 2017).

Despite that, the Cerrado has lost around 88 Mha (46%) of its native vegetation with a projection that 31-34% of the remaining biome is likely to be cleared by 2050 (Strassburg et al., 2017). While much attention has been placed on Amazon, the rate of conversion of native Cerrado vegetation can be up to 2.5 times the conversion observed in the Amazon (Rocha et al., 2011; Strassburg et al. 2017).

Most of the native vegetation conversion tends to occur in areas with dense vegetation that have favourable climate and soil conditions and in flat terrains that are suitable for mechanized farming (Alencar et al., 2020). Therefore, the conversion of natural vegetation into agricultural land (e.g., soybean) and pasture is leading to major carbon emissions (Noojipady et al., 2017) and biodiversity loss (Ratter et al., 1997) through the forest clear-cutting, stressing the importance of frequent mapping approaches that enable monitoring and assessing ongoing change processes.

The Brazil Investment Plan (BIP) under the Forest Investment Program (FIP) seeks to promote sustainable land use and forest management improvement in the Cerrado Biome in order to reduce pressure on remaining forests, reduce greenhouse gas (GHG) emissions and increase carbon dioxide sequestration (Tuchschneider, 2013). As part of BIP, the project “Development of systems to prevent forest fires and monitor vegetation cover in the Brazilian Cerrado” aims to improve Brazil’s capacity to monitor deforestation, prevent the risk of forest fires and improve models for estimating greenhouse gas (GHG) emissions, making tools and data available to environmental agencies[1]. The project will provide the basis for improving the management of water, forest and soil resources in the Brazilian Cerrado, which, together with other projects financed by the FIP in Brazil, should promote the sustainable management of forests. In the context of this project, one of the activities it to modify the existing land cover classification system for the Cerrado developed by IBGE (Brazilian Institute of Geography and Statistics) on the basis of the Food and Agriculture Organization of the United Nations (FAO) Land Cover Classification System framework. Therefore, it will be possible to discriminate forest from non-forest

[1] More information in: http://fip.mma.gov.br/projeto-fm

* Corresponding author: hnbendini@gmail.com

vegetation taking into account the spectrum of structural vegetation complexity in the Cerrado. However, mapping heterogeneous tropical areas, such as the Cerrado, is challenging due to the natural, climatic and topographic factors and the peculiarities of the characteristic physiognomies (Fonseca et al., 2021).

The woody-dominated stratum of Cerrado can be divided in riparian forests, dry tropical forests and savanna woodlands. These forest types have different tree cover and floristic composition, which is associated to leaf deciduousness. The loss of leaf area corresponds to a physiological response to water shortages in dry seasons (Zalamea and González, 2008).

Dry Tropical Forests are characterised by different levels of deciduous trees, depending on climatic conditions and mainly on the depth of the soil (Ribeiro and Walter, 2008). In the Cerrado, the Dry Tropical Forests can be divided according to the tree cover in different seasons. Both Semideciduous Forest and Deciduous Forest have a tree cover ranging from 70 to 95% in the wet season, however the tree cover of Semideciduous Forest ranges from 50 to 60% in the dry season, while the tree cover of Deciduous Forest ranges from 30 to 50%. Furthermore, the canopy height of Semideciduous Forest is about 16 meters, while in Deciduous Forest is about 10 meters (Ribeiro and Walter 2008). In addition, Dry Tropical Forests were considered one of the most threatened tropical ecosystems (Miles et al. 2006). The savanna woodlands are forests with similarities to savannas due to species composition. The tree cover can range from 50 to 90%. Although it may contain evergreen plants, many species show deciduousness (Haidar, 2017).

For these reasons, it's still a challenge to separate between these different forest formations (Ferreira et al., 2004; Sano et al., 2010; Grecchi et al., 2013; Schwieder et al., 2016; Girolamo-Neto et al., 2017; Neves et al., 2021; Bendini et al., 2020). Previous studies highlighted the benefits of dense remote sensing time series, derived land surface phenological metrics (here, also called as phenometrics) and analysed their relationship to the grassland-savanna-forest gradient of the Cerrado (Schwieder et al., 2016; Bendini et al., 2019b). They tried to separate between the different vegetation types using traditional machine learning approaches such as the Support Vector Machine (SVM) and Random Forests (RF), but accurately differentiating between dry forests and savanna woodlands still require more efforts. The Deep Learning approaches, such as the Convolutional Neural Networks (CNN) and Recurrent Neural Networks approaches were already proved to be capable on learning patterns of different land-use and land-covers, for instance on the detection of deforestation on the Brazilian Amazon (Maretto et al., 2020) and Cerrado (Taquary et al., 2021; Matosak et al., 2022), and for vegetation mapping in the Cerrado (Neves et al., 2021; Bendini et al., 2021b).

Considering the importance of Cerrado to biodiversity conservation and the maintaining of environmental services, the development of methods to map the different forest types inserted in Cerrado is important to create bases for conservation programmes, subsidize restauration plains, and to allow estimations of carbon sink and stock. In this study we tested two different machine learning approaches based on the use of dense image time series, RF and LSTM, in order to evaluate if it's possible to separate Dry Tropical Forests and Savanna woodlands.

2. MATERIALS AND METHODS

2.1 Study Area and Reference

We considered the Brazilian State of Tocantins as the study area, which is located in the Northern region of the country and has a large area (277,423.63 Km² according to IBGE 2020), harbouring different ecosystems types. This state is covered by two biogeographic regions, the Cerrado (predominance of savannas) and the Amazon Forest (predominance of forests).

According to Sano et al (2009), which divided the Cerrado into 19 ecoregions that are unique in terms of landscape characteristics, the Cerrado in Tocantins is composed by eight ecoregions: Alto Parnaíba, Araguaia Tocantins, Bananal, Bico do Papagaio, Chapadão do São Francisco, Parnaguá, Planalto Central e Vão do Paranã.

The State of Tocantins is situated in a climate transition area, and the climatic classification of Thornthwaite applied to this State pointed out the presence of three types: i) C1A'w2a' (dry subhumid, megathermal, with large summer water surplus, and a temperature efficiency regime normal to megathermal) in the eastern end of the State; ii) C2A'wa' (moist subhumid, megathermal, with moderate winter water deficiency, and a temperature efficiency regime normal to megathermal) in the central range; and iii) B1A'wa' (humid, megathermal, with moderate winter water deficiency, and a temperature efficiency regime normal to megathermal), in the southwest of Tocantins (Souza et al. 2019).

In order to build a relevant dataset for training and validation, we used reference data provided by the Secretariat of Planning and Budget of the Tocantins state (SEPLAN, 2013), which consists on a detailed map in a scale of 1:100,000, produced with Landsat imagery, supported by visual interpretation of high spatial resolution imagery (CBERS 2B), the SRTM (Shuttle Radar Topography Mission) elevation model geology, terrain and soil maps.

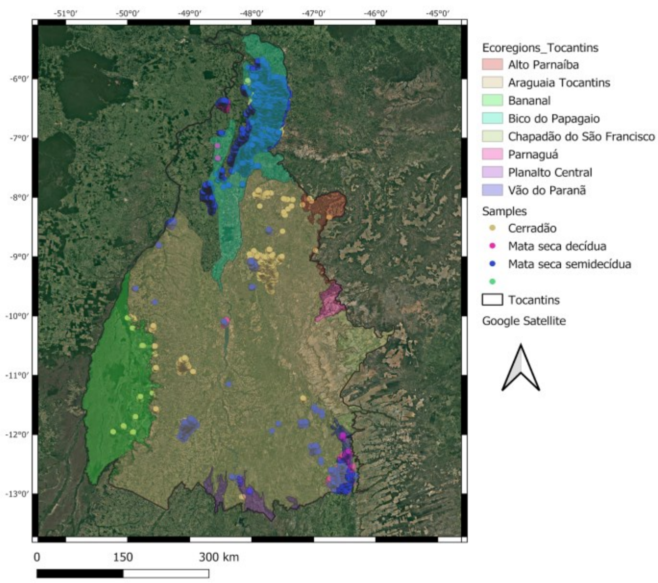

Figure 1. Study area.

2.2 Data and Preprocessing

We extracted a set of 40,041 EVI (Enhanced Vegetation Index) Landsat 7 and 8 point-based time series, using the GEE Time Series Explorer QGIS Plugin (Rufin et al., 2021), during the period of August-2014 to October-2015, involving the 3 classes obtained from the reference database, and considering the Ribeiro & Walter (2008) definition. The classes were: 1) Savanna Woodlands, 2) Deciduous and 3) Semideciduous Dry Forests. We made it compatible over time, removing the areas deforested after 2014 with the PRODES Cerrado deforestation mask (INPE, 2019)[2]. The EVI was chosen because it is known to increase sensitivity for biomass estimation through a de-coupling of the canopy background from the signal and a reduction in atmospheric and soil reflectance influence (Huete et al., 2002).

Then, we applied a weighted ensemble of Radial Basis Function (RBF) convolution filters as a kernel smoother to fill data gaps such as cloud cover and Scan Line Corrector (SLC)-off data (Schwieder et al., 2016, Bendini et al., 2019a). A total of 11 phenometrics were derived using TIMESAT (Jönsson, Eklundh, 2004), extracted for the seasonal cycle observed in the EVI time series. Phenometrics included day-of-the-year (DOY) of start, mid, end, length of crop seasons and phenological proxies like peak, base value, seasonal amplitude or rate of increase, decrease (Jönsson & Eklundh, 2004). and the polar features, which are based on the representation of the time series by projecting the values onto angles in the interval [0,2π] (Körting et al., 2013).

2.3 Random Forest and LSTM classification

The classification was done using the RF algorithm (Breiman, 2001), with different sets of metrics (RBF fits, phenometrics, polar metrics and the combination of them). We randomly selected 70% of the samples to train and 30% for validation. RF is a non-parametric machine learning algorithm that is based on decision trees. As individual decision trees are prone to errors, RF uses an ensemble of many decision trees that were independently trained with random subsets of the input data to overcome this limitation (Breiman, 2001). The RF classifier works based on creating decision trees that are used to predict over data. The class with the majority of votes among all trees is chosen as the final prediction result. RF needs two parameters to be tuned including the number of trees to grow (ntree), and the number of variables randomly sampled as candidates at each split (mtry). Figure 2 shows a generalisation example for how a classification is made using the RF algorithm. The results were evaluated by the confusion matrix, which was used to derive the overall accuracy (Chinchor & Sundheim, 1993) and the class f1-scores (Shapiro, 1999). The “randomForest” package in R was used for our classification tasks (R Core Development Team, 2019).

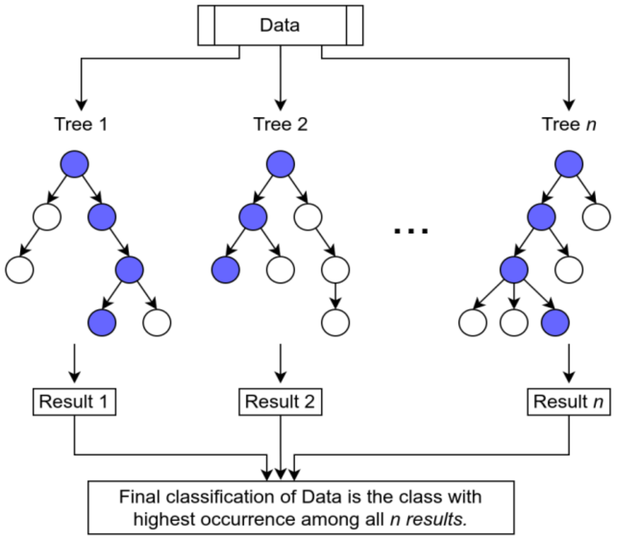

Figure 2. Diagram example of a generalisation for the classification made over Data using the Random Forest algorithm.

The algorithm implementation in R further allows to assess the variable importance of each input variable based on the Gini coefficient (Liaw et al., 2002).

We also tested recurrent neural network models with Long-Short Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) layers, which is a Recurrent Neural Network (RNN). The LSTM was proposed in order to solve problems related with loss of long-term dependency and gradient vanishing present in previous RNN architectures. The LSTM cell is composed of a cell state and input, output, and forget gates, as it is shown in Figure 3.

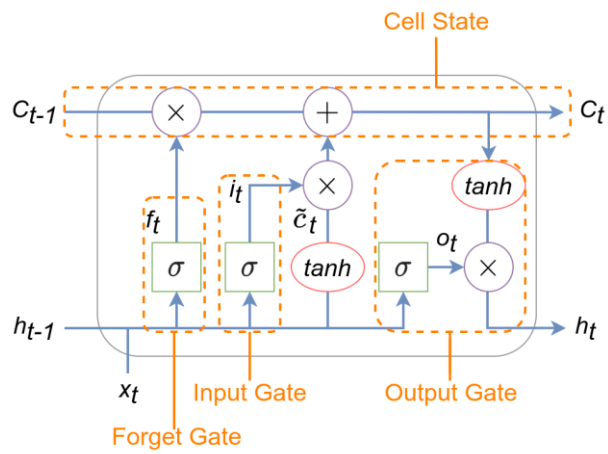

Figure 3. Anatomy of a LSTM cell.

For each input element, ht is computed with Equations 1 to 6:

$$i_t = \sigma\,(W^i x_t + U^i h_{t-1} + b^i) \tag{1}$$

$$f_t = \sigma\,(W^f x_t + U^f h_{t-1} + b^f) \tag{2}$$

[2] Available from http://terrabrasilis.dpi.inpe.br/download/dataset/cerrado-prodes/vector/prodes cerrado 2000 2018 v20190405.zip (download date: 17-Dec-19)

$$\widetilde{C_t} = tanh\,(W^{\tilde{C}} x_t + U^{\tilde{C}} h_{t-1} + b^{\tilde{C}}) \tag{3}$$

$$o_t = \sigma\,(W_o x_t + U^o h^{t-1} + b^o) \tag{4}$$

$$c_t = \tilde{C}_t \cdot i_t + c_{t-1} \cdot f_t \tag{5}$$

$$h_t = o_t \cdot tanh\; c_t \tag{6}$$

where

h_t: is the hidden state at time *t*;
c_t: is the cell state at time *t*;
x_t: is the input at time *t*;
h_{t-1}: is the hidden state at time *t-1* or the initial hidden state;
i_t: is the input gate;
f_t: is the forget gate;
$\tilde{C}_t$: is the cell gate;
o_t: is the output gate;
σ: is the sigmoid function.

The same samples used in the RF classifier were also used with the LSTM. In this case, five distinct sequential models were used. Regarding their composition, model number 5 had 5 LSTM layers, with 96, 48, 24, 12, and 6 units, respectively. Each LSTM was followed by a batch normalisation layer, and at the end there was a fully connected layer (Dense) with 3 units and *SoftMax* activation function. Model number 4 was similar to model number 5, but without the first LSTM and batch normalisation layers, and so on until model 1, which had only one LSTM layer with 6 units, the batch normalisation and fully connected layers. Table 1 shows details about the layers of each model.

	Model				
	1	**2**	**3**	**4**	**5**
Layers	LSTM (6) BN Dense (3)	LSTM (12) BN LSTM (6) BN Dense (3)	LSTM (24) BN LSTM (12) BN LSTM (6) BN Dense (3)	LSTM (48) BN LSTM (24) BN LSTM (12) BN LSTM (6) BN Dense (3)	LSTM (96) BN LSTM (48) BN LSTM (24) BN LSTM (12) BN LSTM (6) BN Dense (3)

Table 1. Layers used in each model evaluated in the tenfold cross validation procedure. LSTM (n): Long-short term memory (number of units); BN: Batch normalisation; and Dense (n): Dense fully connected layer (number of output units).

The models were trained with the Categorical Cross Entropy loss function, Adam optimizer with a learning rate of 0.00001, batch size of 256 samples, and 1000 epochs.

3. RESULTS AND DISCUSSIONS

Our best RF model was built with the dataset composed by RBF fits, in which we empirically determined the parameters mtry of 5 and ntrees of 500, and reached an average overall accuracy of 0.85005, while the F1-Scores for the classes Savanna Woodlands, Deciduous Dry Forest, and Semideciduous Dry Forest were 0.68261, 0.70348, and 0.90273, respectively. Figure 4 shows the confusion matrices.

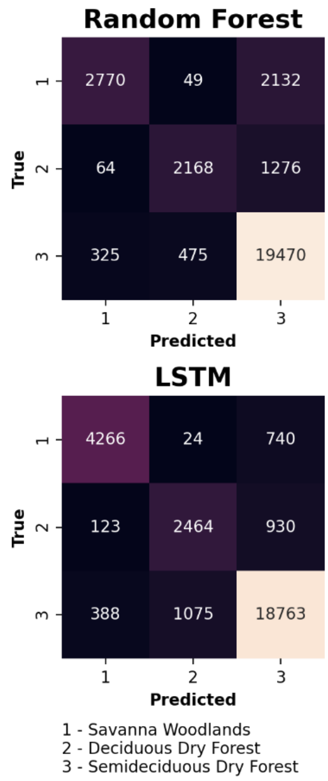

Figure 4. Confusion matrices.

Table 2 present the variable importance analysis results based on the Mean Decrease Gini (MDG). The mean decrease in Gini coefficient is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting RF. The higher the value of mean decrease Gini, the higher the importance of the variable in the model (Han et al., 2016).

Variable	Mean Decrease Gini
Fitted EVI in 2014-09-26	235.72
Fitted EVI in 2014-10-04	197.26
Fitted EVI in 2014-10-12	191.20
Fitted EVI in 2014-08-01	153.99
Fitted EVI in 2015-09-05	157.90

Table 2. Variable importance analysis results based on the MDG.

Figure 5 shows the averaged EVI phenological profiles for each class in the season 2014–2015. We found that the EVI values were higher during the early of wet season (i.e., December and January) and lower during the end of the dry season (September and October). A higher amplitude can be observed in the EVI time series for the Deciduous Dry Forest class, while the Savanna

Woodlands presented lower amplitude with a higher length of season.

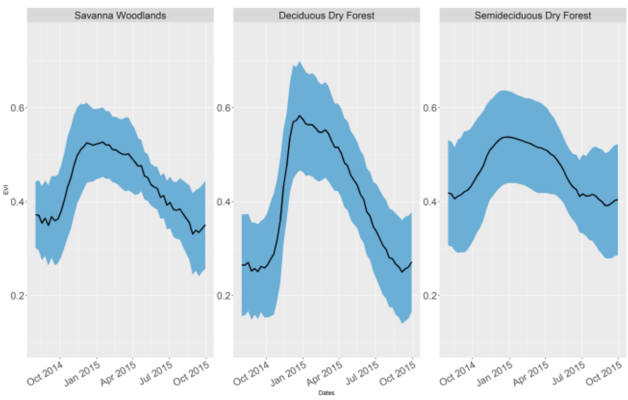

Figure 5. Averaged EVI phenological profiles for each class in the season 2014–2015 (black lines), with their respective standard deviations (blue margins)

Figure 6 shows the boxplots of the two most important variables for each class in the season 2014–2015, based on the variable importance analysis by the MDG, provided by the RF algorithm.

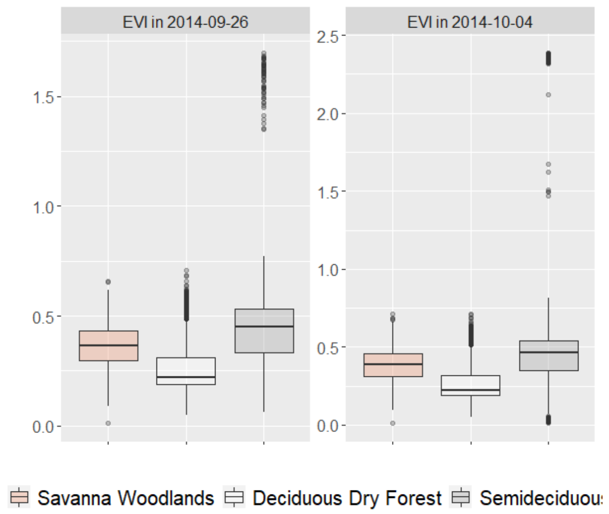

Figure 6. Boxplots of the two most important variables for each class in the season 2014–2015.

We found that the EVI values during the end of the dry season were the most informative variables for separating among the different vegetation classes, suggesting that there are significant differences between their seasonal dynamics during this period (Figure 6). This result suggests that the leaf cover is lowest in the end of the dry season. Therefore, it is the best period to acquire satellite images to separate the different forest types of Cerrado.

Haidar (2017) analysed the leaf phenological cycles of the remnants of dry forests, using vegetation indexes time series and phenological metrics. The author determined four phenological groups with trends that vary from deciduous to evergreen canopy species, as well as meaningful differences in the cycles of the remnants. The existence of these different phenological groups can explain the high variance within the semideciduous forests class, which may present a gradient varying between deciduous and evergreen forests. This high variance led to confusions between the semideciduous forests and savanna woodlands, in which pixels of savanna woodlands were misclassified as semideciduous forests. The variation of the phenology is related to the differences in the soil physical and chemical properties and the average temperature, which can vary across the Tocantins state.

Bendini et al. (2020) used Landsat Analysis Ready Data (ARD) in combination with different environmental data for a modelling exercise for classifying the vegetation in the Cerrado in two different hierarchical levels. In the second level, the authors also reported the complexity of classifying in details the forest classes. Bendini et al. (2021a) spatialized the results from the mentioned model producing a map for the whole Cerrado, and reported a f1-score of 0.658 and 0.611, for Savanna Woodlands and Dry Forests, respectively. Neves et al. (2021) also used a Deep Learning technique, based on the adaptation of Convolutional Neural Network architecture, the U-Net (Ronneberger et al., 2015), to process a WorldView-2 image with 2-m spatial resolution to hierarchically classify the physiognomies of a Cerrado protected area (Brasilia National Park), reaching a f1-score of 0.86 for the Savanna Woodlands class, although they didn´t account for the Dry Forests.

The Model 4 of the LSTM approach presented the highest accuracy during training, and therefore was used to predict over the test samples. The test samples overall accuracy was 0.88601, while the F1-Scores for the classes Savanna Woodlands, Deciduous Dry Forest, and Semideciduous Dry Forest were 0.86999, 0.69604, and 0.92294, respectively. LSTM approach presented a higher accuracy for the savanna woodlands class, in comparison to the RF approach. For this reason, the LSTM model reached the highest overall accuracy, showing its potential for automatically separating the different types of forests in the Cerrado. This result suggests the suitability of the recurrent neural networks on handling complex long-term dependencies such as the EVI dense time series of forest vegetations, finding patterns that were not captured by a traditional machine learning approach such as the RF.

3.1 Final Considerations

In this work, we tested two different machine learning approaches, RF and LSTM, based on the use of dense image time series, in order to evaluate if it's possible to separate Dry Tropical Forests and Savanna woodlands. LSTM approach presented a higher accuracy for the savanna woodlands class, in comparison to the RF approach, reaching the highest overall accuracy. In general, it was possible to separate between the different forest classes. Although the results are motivating, there´s still the potential for improving the accuracy with other deep learning approaches, such as the convolutional Long-Short-Term-Memory (ConvLSTM), and also, testing different spectral bands, other spectral indices and data from Synthetic Aperture Radar (SAR), such as the Sentinel 1. The combination of these image time series with auxiliary data, such as terrain and soil type, can also add value to this classification task. This study showed the potential for the development of a semi-automatic method for discriminating the different types of forest formations in the Brazilian Savanna, based on remote sensing.

3.2 Acknowledgements

We thank the project "Development of systems to prevent forest fires and monitor vegetation cover in the Brazilian Cerrado" (World Bank Project #P143185) – Forest Investment Program (FIP). We also thank the Amazon Fund (Fundo Amazônia - BNDES) through the project "Monitoramento Ambiental dos Biomas Brasileiros".

REFERENCES

Alencar, A., Z Shimbo, J., Lenti, F., Balzani Marques, C., Zimbres, B., Rosa, M., Arruda, V., Castro, I., Fernandes Márcico Ribeiro, J. P., Varela, V. et al., 2020. Mapping Three Decades of Changes in the Brazilian Savanna Native Vegetation Using Landsat Data Processed in the Google Earth Engine Platform. *Remote Sensing*, 12(6), 924.

Bendini, H., N., Fonseca, L. M. G., Schwieder, M., Körting, T. S., Rufin, P., Sanches, I. D. A., Leitao, P. & Hostert, P. (2019a). Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series. *International Journal of Applied Earth Observation and Geoinformation*, 82, 101872.

Bendini, H. N. et al., 2019b. Assessing Satellite-Derived Phenological Metrics and Terrain Data as a Proxy for Vegetation Dynamics Along the Brazilian Savanna Corridor. *PECORA - International Symposium on Remote Sensing of Environment*, 38.

Bendini, H. N., Fonseca, L. M. G., Schwieder, M., Rufin, P., Korting, T. S., Koumrouyan, A., & Hostert, P. (2020). Combining environmental and Landsat analysis ready data for vegetation mapping: a case study in the Brazilian savanna biome. I*nternational Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, B3-2020.

Bendini, H. N; et al 2021a: Detailed vegetation maps of the Brazilian Savanna (Cerrado) biome produced with a semi-automatic approach. *PANGAEA*, https://doi.org/10.1594/PANGAEA.932642

Bendini, H. N., Fonseca, L. M., Maretto, R. V., Matosak, B. M., Taquary, E. C., Simões, P. S., ... & Valeriano, D. D. M. (2021b). Exploring a Deep Convolutional Neural Network and Geobia for Automatic Recognition of Brazilian Palm Swamps (Veredas) Using Sentinel-2 Optical Data. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS* (pp. 5401-5404). IEEE.

Breiman, L., 2001: Random forests. *Machine learning*, 45(1), 5–32.

Chinchor, N., & Sundheim, B. M. (1993). MUC-5 evaluation metrics. In *Fifth Message Understanding Conference (MUC-5): Proceedings of a Conference Held in Baltimore, Maryland, August 25-27, 1993*.

Colli, G.R., Vieira, C.R.; Dianese, J.C. 2020: Biodiversity and conservation of the Cerrado: recent advances and old challenges. *Biodiversity and Conservation* 29, 1465–1475. doi.org/10.1007/s10531-020-01967-x

Ferreira, L. G., Huete, A. R., 2004: Assessing the seasonal dynamics of the Brazilian Cerrado vegetation through the use of spectral vegetation indices. *International Journal of Remote Sensing*, 25(10), 1837–1860.

Fonseca, L. M., Körting, T. S., Bendini, H. D. N., Girolamo-Neto, C. D., Neves, A. K., Soares, A. R., ... & Maretto, R. V. (2021). Pattern Recognition and Remote Sensing techniques applied to Land Use and Land Cover mapping in the Brazilian Savannah. *Pattern Recognition Letters*, 148, 54-60.

Körting, Thales Sehn; Fonseca, Leila Maria Garcia; Câmara, Gilberto. GeoDMA—Geographic data mining analyst. *Computers & Geosciences*, v. 57, p. 133-145, 2013.

Haidar, R. F. (2017). Núcleo de florestas tropicais sazonalmente secas do Cerrado: diversidade, fitogeografia, endemismo, fenologia foliar e os controles ambientais e espaciais.

Han, Hong; Guo, Xiaoling; Yu, Hua. Variable selection using mean decrease accuracy and mean decrease gini based on random forest. In: *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*. IEEE, 2016. p. 219-224.

Hochreiter, S., Schmidhuber, J., 1997: Long Short-Term Memory. *Neural Computation*, 9 (8), 1735-1780.

IBGE. Instituto Brasileiro de Geografia e Estatística 2020: Cidades. Sistema agregador de informações do IBGE sobre os municípios e estados do Brasil. Available at https:// cidades.ibge.gov br/. Accessed 10 Dec 2021.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*(3), 18-22.

Matosak, B. M., Fonseca, L. M. G., Taquary, E. C., Maretto, R. V., Bendini, H. D. N., & Adami, M. (2022). Mapping Deforestation in Cerrado Based on Hybrid Deep Learning Architecture and Medium Spatial Resolution Satellite Time Series. Remote Sensing, 14(1), 209.

Maretto, R.V.; Fonseca, L.M.G.; Jacobs, N.; Korting, T.S.; Bendini, H.N.; Parente, L.L. Spatio-Temporal Deep Learning Approach to Map Deforestation in Amazon Rainforest. *IEEE Geosci. Remote Sens. Lett*. 2020, 18, 771–775.

Miles, L., Newton, A.C., DeFries, R. S., Ravilious, C., May, I., Blyth, S., Kapos, V. Gordon, J.E., 2006: A global overview of the conservation status of tropical dry forests. *Journal of Biogeography*, 33(3), 491–505. doi:10.1111/j.1365-2699.2005.01424. x.

Ministry of Environment, 2021: *Ministry of Environment.* Available online: http://www.mma.gov.br/ (accessed on October 2021).

Myers N., Mittermeier R.A., Mittermeier C.G., Fonseca G.A.B., Kent J. 2000: Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858. doi.org/10.1038/35002501

Neves, Alana K. et al. Hierarchical mapping of Brazilian Savanna (Cerrado) physiognomies based on deep learning. *Journal of Applied Remote Sensing*, v. 15, n. 4, p. 044504, 2021.

Noojipady, P., et al: Forest carbon emissions from cropland expansion in the Brazilian Cerrado biome. *Environmental Research Letters*, 12(2), 025004.

Oliveira P.T.S., Nearing M.A., Wendland E. 2015: Orders of magnitude increase in soil erosion associated with land use change from native to cultivated vegetation in a Brazilian Savannah. Environment, *Earth Surface Processes and Landforms* 40(11), 1524–32.doi.org/10.1002/esp.3738

Ratter, J. A. et al., 1997: The Brazilian Cerrado vegetation and threats to its biodiversity. *Annals of botany*, 80(3), 223–230.

Ribeiro, J.F., Walter, B.M.T. As Principais Fitofisionomias do Bioma Cerrado. In: Sano, S; M Almeida, S.P., Ribeiro, J.F. (Eds) *Cerrado: Ecologia e Flora* vol. 2. Brasília: EMBRAPA CERRADOS, 2008.

Rufin, P., et al., 2021: GEE Timeseries Explorer for QGIS – Instant Access to Petabytes of Earth Observation Data, *ISPRS Journal of Photogrammetry and Remote Sensing*, XLVI-4/W2-2021,155–158.

Sano, E. E., Rosa, R., Brito, J. L., Ferreira, L. G., 2010: Land cover mapping of the tropical savanna region in Brazil. *Environmental monitoring and assessment*, 166(1-4), 113–124.

Sano, E.E., Rodrigues, A.A., Martins, E.S., Bettiol, G.M., Bustamante, M.M., Bezerra, A.S., Couto Jr., A.F., Vasconcelos V., Schuller J., Bolfe, E.L. 2019: Cerrado ecoregions: A spatial framework to assess and prioritize Brazilian savanna environmental diversity for conservation. *Journal of environmental management*, 232, 818-828. doi.org/10.1016/j.jenvman.2018.11.108

de Souza, F.H.M., Viola, M.R., Avanzi, J.C., Giongo, M., Vieira Filho, M. 2019: Thornthwaite's Climate Regionalization for the State of Tocantins, Brazil. *Floresta*, 49(4), 783-792. doi.org/10.5380/rf.v49i4.59188

Schwieder et al., 2016: Mapping Brazilian savanna vegetation gradients with Landsat time series. *International Journal of Applied Earth Observation and Geoinformatics*, 52, 361–370.

Secretariat for Planning and Modernization of Public Management (SEPLAN), 2013: Geographic Database of Phytoecological Regions Mapping and Inventory Forestry of the State of Tocantins. Palmas, SEPLAN/DZE, 2013. DVD-ROM.

Shapiro, D. E. (1999). The interpretation of diagnostic tests. *Statistical methods in medical research*, *8*(2), 113-134.

Strassburg B., Brooks T., Feltran-Barbieri R., Iribarrem A., Crouzeilles R., Loyola R, Latawiec A.E., Oliveira Filho F. J.B., Scaramuzza C.A.M., Scarano F.R., Soares-Filho B., Balmford A. 2017: Moment of truth for the Cerrado hotspot. Nature Ecology and Evolution 1(4), 1–3. doi.org/10.1038/s41559-017-0099.

Taquary, E. C., Fonseca, L. G., Maretto, R. V., Bendini, H. N., Matosak, B. M., Sant'Anna, S. J., & Mura, J. C. (2021). Detecting Clearcut Deforestation Employing Deep Learning Methods and SAR Time Series. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (pp. 4520-4523). IEEE.

Team, R. C. D., 2019. R: A language and environment for statistical computing (version 3.5. 2, R foundation for statistical computing, Vienna, Austria, 2018)

Tuchschneider, D., 2013. Brazil - development of systems to prevent forest fires and monitor vegetation project (English). *Technical report*, World Bank Group, Washington, D.C.

Zalamea, M., González, G., 2008: Leaf fall phenology in a subtropical wet forest in Puerto Rico: from species to community patterns. *Biotropica*, 40(3), 295-304. doi.org/10.1111/j.1744-7429.2007.00389.x