Deep Convolutional Neural Network for Classifying Satellite Images with Heterogeneous Spatial Resolutions

Mateus de Souza Miranda[1(✉)], Valdivino Alexandre de Santiago Jr[1(✉)], Thales Sehn Körting[1(✉)], Rodrigo Leonardi[2(✉)], and Moisés Laurence de Freitas Jr[3(✉)]

[1] Instituto Nacional de Pesquisas Espaciais (INPE), Avenida Dos Astronautas, 1758, São José Dos Campos, SP da Granja - 12227-010, Brazil
{mateus.miranda,valdivino.santiago,thales.korting}@inpe.br
[2] Agência Espacial Brasileira, Brasília, DF 70610-200, Brazil
rodrigo.leonardi@aeb.gov.br
[3] Instituto Federal de Educação, Ciência e Tecnologia do Tocantins, Povoado Santa Tereza, Countryside, Araguatins, TO 77950-000, Brazil
moises.junior@ifto.edu.br

Abstract. Deep Learning, and most notable Deep Neural Networks, have largely driven Artificial Intelligence in the area of remote sensing, mainly image classification tasks. In this paper, we present an approach based on Convolutional Neural Networks to classify Earth Observation satellite images as environmental preserved or non-preserved areas. One interesting feature of our approach is the fact that we used sensors with different spatial resolutions to assess the performance of a traditional network. We relied on images from the Tocantins Cerrado obtained by the Wide-Scan Multispectral and Panchromatic Camera of the CBERS-4A satellite, with a spatial resolution of 8m to create the training dataset. For testing, we set up a set of images of the Sentinel satellite, with a spatial resolution of 10m from Goiás Cerrado. Results imply that Convolutional Neural Networks are feasible and are a good alternative for classifying remote sensing areas even when dealing with images from various sensors, and also with different spatial resolutions, where the model used in this study obtained an accuracy of 0.87. This study demonstrates the flexibility of Convolutional Neural Networks concerning the ability to generalize knowledge for classifying remote sensing images.

Keywords: Deep convolutional neural network · Satellite image classification · Remote sensing

1 Introduction

The Brazilian Cerrado has a rich biodiversity, being the second largest biogeographic region in South America. This biome can be found in the States of

© Springer Nature Switzerland AG 2021
O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12955, pp. 519–530, 2021.
https://doi.org/10.1007/978-3-030-87007-2_37

Maranhão, Piauí, Bahia, Goiás, Distrito Federal, Mato Grosso, Minas Gerais, São Paulo, Paraná, Rondônia, and Tocantins, corresponding to 23% of the Brazilian territory [4]. This biome has great floristic diversity, with flora characteristics that since the formation of forests, associated with watercourses; savannas, and fields, both characteristics congruent to climatic and geological conditions [21]. This biome presents large agricultural industries that, despite boosting Brazil in food production, collaborate for annual deforestation increments, as well as illegal exploitation activities in preserved areas [3].

The activities in the Brazilian Cerrado are monitored by space research institutions, such as Brazil's National Institute for Space Research (INPE), which observe the Earth through satellite images or aerial vehicles that carry sensors for imaging and mapping and perform tasks such as classification images. From this, deforestation measures can be monitored annually, checking agricultural and urban areas expansion. For example, in 2020 there was an increase of 7.3 km^2 on the Cerrado, as illustrated in Fig. 1. Compared to the last year, approximately 0.8 km^2 more kilometres were registered, the one with the lowest rate of deforestation since 2004, where more than 29.000 km^2 of original vegetation was deforested [2]. This may mean returning to the intense extraction of raw material or transforming this environmental area into pasture or for planting crops, for example, soy.

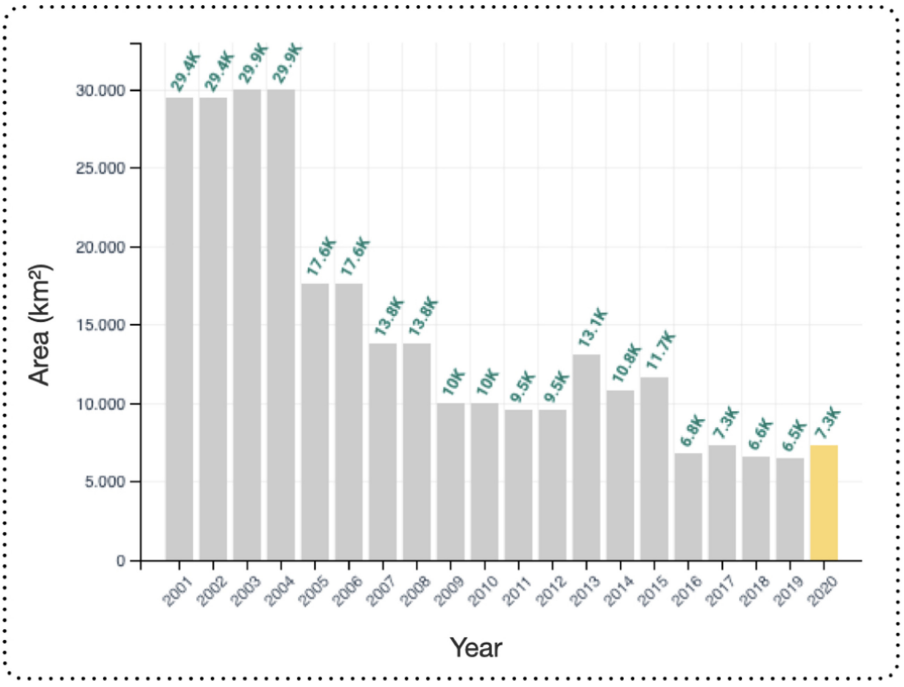

Fig. 1. Annual deforestation in the cerrado. Adapted from TerraBrasilis.

Image classification is a traditional Machine Learning/Pattern Recognition task that has been so long explored by the Remote Sensing community [19]. Classifying Land Use and Land Cover (LULC) using Remote Sensing imagery is not a trivial task, due to several factors, including that Remote Sensing is a source of big data [16], since every day new images are obtained (increasing *data volume*), and previously captured images are also combined as time series (*data velocity*). Moreover, the available imagery is expanding the number of spectral channels (i.e. Landsat 5 and 7 satellites contain 8 bands, and Landsat 8 has 11 bands), which means different ways to observe the interaction between targets and electromagnetic radiation (*data variety*).

Several authors have published in the areas of LULC classification, using different strategies, ranging from pixel-based techniques to state-of-the-art methodologies, including object-based approaches and more recently Deep Learning techniques [18]. Methods based on pixels and objects often rely on the use of data enhancements, which in accordance to [15] include feature extraction, statistical measures, texture, ancillary data, spectral indices, and others. However, the spotlight is now on Deep Learning techniques, which provide a pattern recognition system based more on automatic learning and less on hand-designed heuristics [17], or data enhancements.

Among all Deep Learning techniques, Convolutional Neural Networks (CNNs) [9] have attracted the attention of both academia and industry to solve problems related to computational vision. Despite this great interest and several studies that have already been published, we believe that it is interesting to realize whether CNNs are suitable to correctly classify satellite images considering heterogeneous sets of images, acquired by different sensors and with different spatial resolutions.

In this paper, we present an approach which relies on CNN to classify images as preserved and non-preserved areas. One interesting feature of our solution is the fact that we used sensors with different spatial resolutions to assess the performance of a traditional CNN. We relied on images from the Tocantins Cerrado obtained by the Wide-Scan Multispectral and Pancromatic Camera (WPM) of the CBERS-4A satellite, with a spatial resolution of 8m to create the training dataset. For testing, we set up a set of images of the Sentinel satellite, with spatial resolution of 10 m from Goiás Cerrado.

This paper is organized as follows. Section 2 introduce related works. Section 3 presents Material and Methods. Results and discussion are presented in Sect. 4. Section 5 presents conclusions and future directions.

2 Related Work

Convolutional neural networks (CNNs) are efficient not only because of the dynamics in extracting image characteristics but above all because of the ability to learn to recognize patterns. This artificial neural network provides several tools, including the creation of feature maps, encoder and decoder, for the improvement or development of new architectures with specific tasks, for example, classification of scenes from satellite images.

Given this, several studies in the field of remote sensing point to CNNs as indispensable tools for satellite image processing. Since these images are complex, both due to the variety of elements and information, as well as the type of sensor and spatial resolution. Thus, the research conducted by [1], describes how they implemented the late fusion technique using different deep CNN architectures and data sets, applied to recognize binary patterns, textures and classify remote sensing scenes, and presents relevant results precision in classification regarding the use of the architecture proposed by the authors.

Otherwise, [6] use CNN's techniques of the encoder and decoder for onboard cloud screening, selecting images with less cloud cover. For this, the authors made some changes in six CNNS encoder and decoder architectures, reducing the number of spectral bands, the input size, number of network filters and also making use of shallower networks. Another interesting factor regards the use of images with the same spatial resolution and imaging sensor for training, validation and testing. With that, the results indicate the possibility of employ a CNN in satellite with the mission of classifying images with great cloud cover.

Toward [24], one of the main challenges in the image classification performance is the addition of full layers at the end of models CNN, where in some cases can not be efficient in the assimilation of spatial information present in images, for example. Hence, presents suggestions for the best manager of the information contained in images, the first has a group of neurons as a capsule or a vector that replace the neuron in the traditional neural network and can encode the properties and spatial information of features in an image to achieve equivalence.

Another very important aspect is the use of the real size of the satellite image, although it requires high processing due quite rich in details and has a high spatial resolution [1]. Therefore, many studies use the technique of transfer of learning or pre-trained networks [6], in order to speed up processing and make the model more efficient. In addition, another technique used for pre-processing of data set, mainly when there are little data per class for training step, is Data Augmentation, where transformations are applied to images, such as rotate and flip, for example [23].

3 Materials and Methods

3.1 Study Sites

The regions of interest (ROI) are located in Tocantins (ROI1) and Goiás (ROI2), Brazilian states where the Cerrado is the predominant biome. These states stand out with their landscapes reminiscent of savanna and several species of animals, in addition to two large rivers flow between the Araguaia and Tocantins, from south to north.

In the Tocantins, there are not only large grain producers, but also traditional communities, such as quilombolas, indigenous and riverside ethnic groups, and associations of babaçu coconut breakers and others, which use natural resources

for their own subsistence [5]. While Goiás stands out for its mineral extraction. This activity is essential for the state's economy, as well as collaborating for national and international production and consumption. As a result, Goiás not only stands out for these raw material exports but also for environmental problems that put the fauna and flora of the Cerrado biome at risk [7].

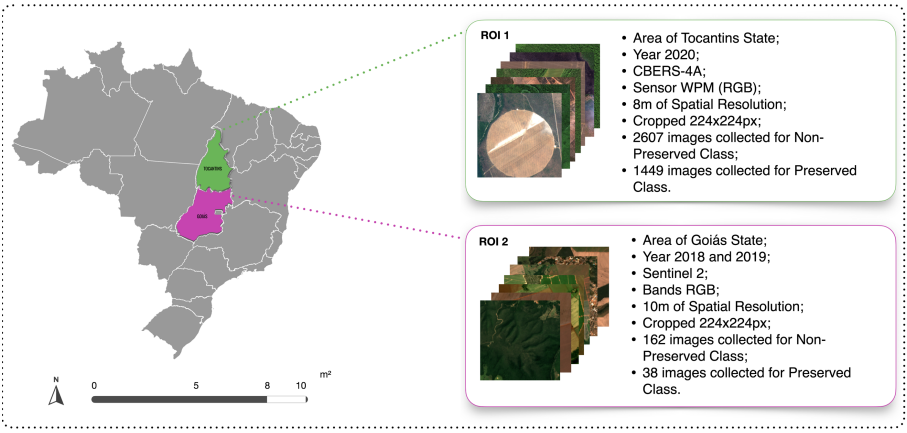

Fig. 2. ROI1 is represented in green and ROI2 in pink.

In Fig. 2, we identify on a map the regions of interest targeted in this study. Within ROI1, we selected images of the RGB bands obtained by the WPM camera of the CBERS-4A satellite, with 8-meter spatial resolution, in the period of 2020. Within ROI2, we also handled images on RGB channels, with a spatial resolution of 10 m, by the Sentinel 2 satellite from 2018 to 2019.

3.2 Datasets

The training dataset is formed by images of the ROI1. Hence, 21 images were considered, with 14.115 × 14.494px. The respective spectral bands of each image were composed and soon after the areas of interest were cut, in the TerraView software[1]. Hence, we got 1,449 images for the preserved and 2,607 images for the non-preserved classes. Given the significant difference in the amount of data between the two classes, we implemented a static data augmentation method using the tools from the imgaug library[2]. In addition, it helps the model to avoid overfitting or specializing the model on training data.

The static data augmentation was organized so that the quantity was the same for both classes, as illustrated in Fig. 3. In this way, 4 transformations were applied to the images of the non-preserved class, with center crop of

[1] http://www.dpi.inpe.br/terralib5/wiki/doku.php.
[2] https://imgaug.readthedocs.io/en/latest/.

224 × 224px: rotate (–50°, 30°); shear (0°, 40°), horizontal and vertical flips. For the preserved class, in addition to the other 4 transformations, rotate and shear was applied for each flip. Consequently, each class has 13,035 images, and hence the training set contains 26,070 images.

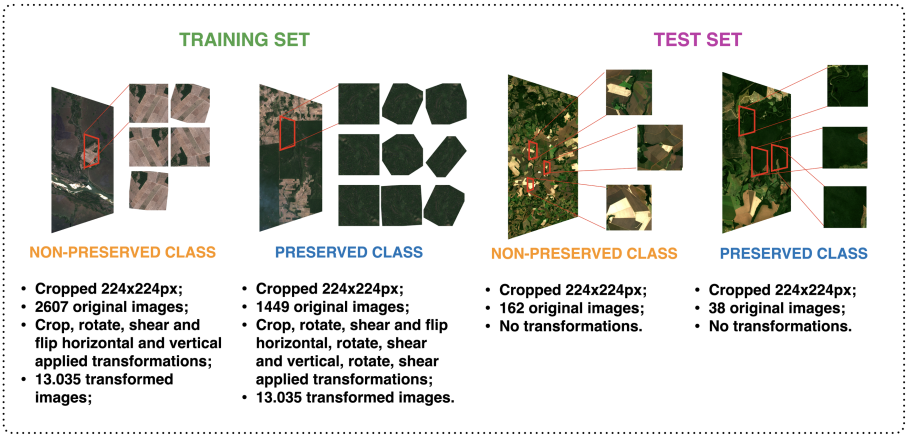

Fig. 3. Sample images of the training and test sets.

Regarding the test dataset (see Fig. 3), we selected 5 images from ROI2, with 16806 × 10986px. The respective spectral bands of each image were composed, using the TerraView software. The areas of interest were clipped, totalling 200 rasters, and standardized for the proportion of 224 × 224px. The clippings were exported to the QGIS software to make adjustments to the colour channels. We arranged them in two preserved and non-preserved sets, each containing 38 and 162 images, respectively.

In general, the preserved class has scenarios with the formation of forests, rivers, grasslands, and some savanna regions; the non-preserved class includes views of pastures, plantations, reforested areas, urban areas, and ore sites. However, we tried to standardize the types of data for each class in order to avoid ambiguities.

3.3 The Model

We designed a traditional CNN that works with satellite images to classify environmental areas as non-preserved or preserved area. This model was based on general recommendations to build a CNN for binary classification by Jones Granatyr [8], but we changed the number of convolutions and fully-connected (dense) layers, as well as we adjusted dropout, kernels in convolutions and neurons in hidden layers, number of epochs, calculation of loss as illustrated in Fig. 4.

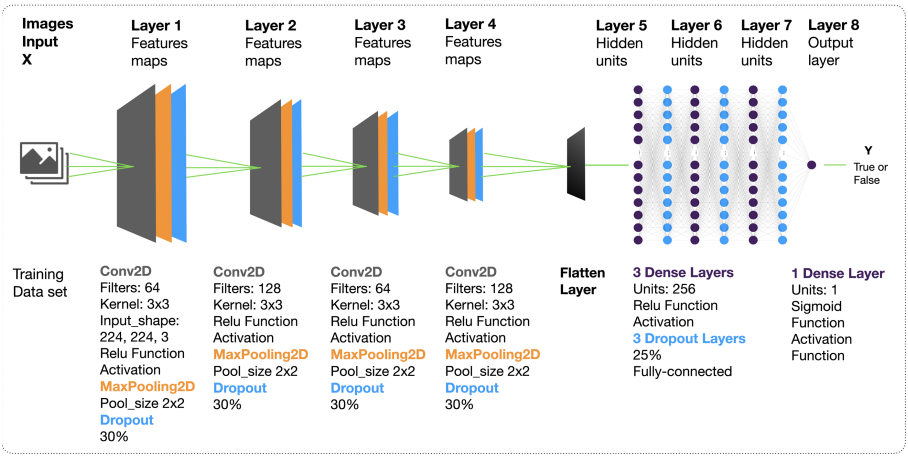

Fig. 4. Architecture of the proposed CNN.

The training images are inserted in the convolutional layers, with input-shape of 224 × 224, 3 × 3 kernel, activated by the ReLU function [12], after each convolution a MaxPooling2D layer [13], pooling 2 was added, 2, and a Dropout layer, with a probability of 30% [11]. After the convolution layers, we have 3 fully-connected layers each one with 256 neurons, activated by the ReLU function, and 3 Dropout layers with a probability of 25%. The output layer has only one neuron, activated by the Sigmoid Function in order to produce probabilistic results on a scale from 0 to 1 [12].

We use Adam's optimizer, adjusted for learning_rate = 0.001, beta_1 = 0.9, beta_2 = 0.999, epsilon = 1e–07, which consists of a descending stochastic gradient method that is based on adaptive estimation of first and second-order moments [10], and the loss function was binary-crossentropy [14]. For training, we defined 80 epochs, saving the models at the end of each epoch.

3.4 Accuracy Assessment

In order to evaluate the performance of our model, we used the following metrics:

1. Accuracy. The number of correct predictions per the total number of predictions;
2. F1-Score. This metric is defined as $F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$. Precision is the ratio $\frac{TP}{TP+FP}$, where TP is the number of true positives and FP is the number of false positives. Recall is the ratio $\frac{TP}{TP+FN}$, where FN is the number of false negatives.

Accuracy and F1-Score metrics were used with each CNN's model saved after each epoch, separately, to label the test set images as preserved or non-preserved.

All these metrics range from 0 to 1, where 0 is the worst performance and 1 is the best one.

Let us consider Fig. 5. We defined 0.98 as the threshold to consider if an image is from the preserved or non-preserved class. Hence, if the prediction is greater than 0.98 the image is classified as preserved, otherwise is non-preserved.

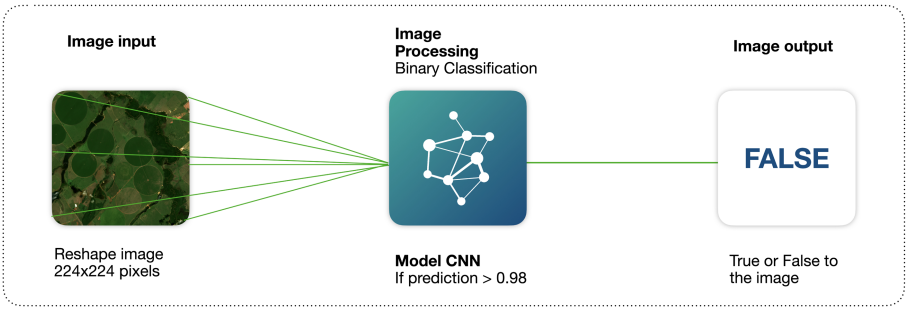

Fig. 5. Decision about the class of an image.

4 Results and Discussion

The first aspect observed in this research is the use of static data augmentation since few images were collected for the training stage. In this way, the number of images per class increased significantly, so that both sets of images had the same amount of data (balanced classes).

In Fig. 6, we can notice the values of the metrics obtained during the testing phase. Note that we assessed the CNN model obtained after each epoch during training. Moreover, we decided to provide metrics per class, i.e. Accuracy and F1-Score for classes preserved and non-preserved. With this, we can see that the non-preserved class started to be correctly labelled from model 1 onwards, while the preserved class achieved more accurate results from model 15. However, we noticed better performance between models 47 to 56, where there are low oscillation hits and misses.

Model 49 got the best scores in metrics, which as it correctly measured 137 images for the non-preserved class and 37 for preserved. This represents an Accuracy of 0.8456 and 0.9736 considering the non-preserved and preserved classes, respectively. F1-Scores are 0.9163 and 0.9866 considering the non-preserved and preserved classes, respectively. Hence, the overall Accuracy (acc) of the CNN is:

$$acc = \frac{137 + 37}{162 + 38} = 0.87 \tag{1}$$

Although the results were quite expressive, the model incorrectly labelled some images, as shown in Fig. 7. This happened more frequently with these

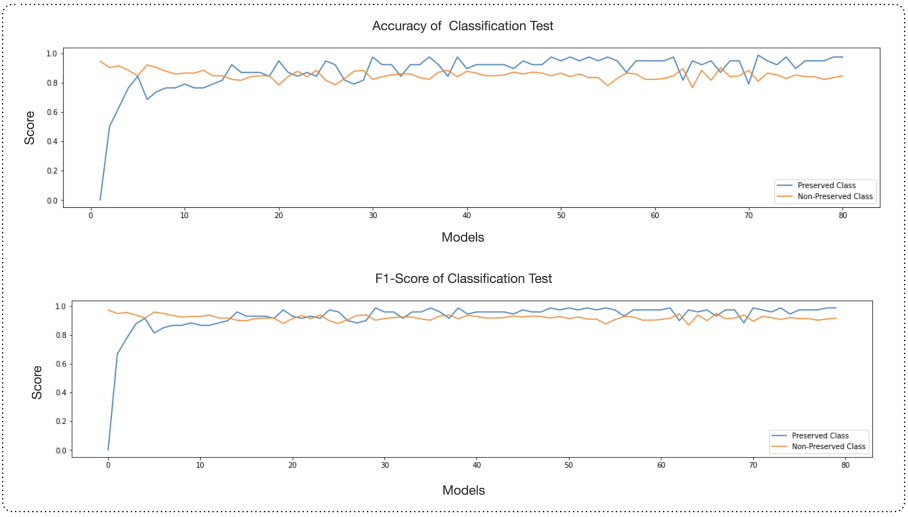

Fig. 6. F1-Score and accuracy metrics of classification test.

images non-preserved being classified as preserved. We explain this fact because of a higher concentration of dark green tones in certain regions of some non-preserved images confusing the CNN.



Fig. 7. Image classification errors that occurred in the tests, where reading preserved would actually be non-preserved.

To give an idea of the characteristics of the images in the test set, Figs. 8 show Ratio Band × Amplitude Band graphs, two spectral features, considering the green and red bands, respectively. These graphs show the separating hyperplane (solid line) and margins (dashed lines) obtained via the Support Vector Machines (SVM) algorithm [20,22] with a linear kernel. The support vectors are highlighted with a black circle around them. Moreover, the images of the preserved class are in pale blue and of the non-preserved one are in pale red.

Note that our idea here is not to use SVM as a classifier, even if we were to do that, we would use the training set to train the algorithm. Our goal is just to

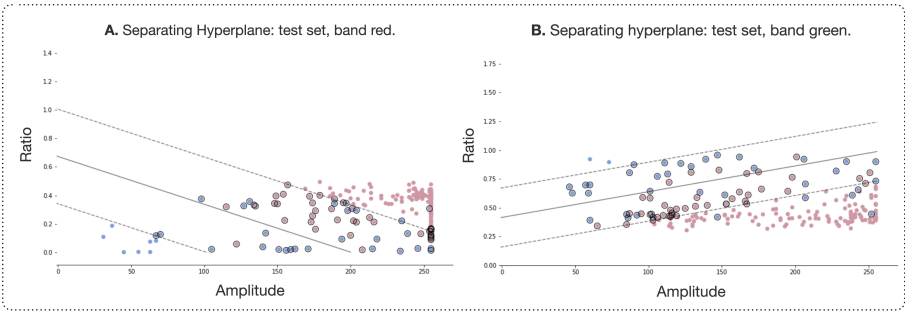

Fig. 8. Ratio × amplitude: separating hyperplane of the test set, green and red bands. (Color figure online)

give an overview of the test set showing its main characteristics. It is evident that looking at each band on its own, the test set is not linearly separable. We have 38.5% of support vectors in the green band and 33.5% in the red band. If our idea was to use this test set as, indeed, a training set for classification via SVM, we must remark that a large number of support vectors is often an indication of overfitting. Although the percentage of support vectors is not very high, we may say that it is not very low either. The whole point is to emphasize that we have used a non-trivial test set and the performance of our traditional CNN was very satisfactory, showing its feasibility for remote sensing classification when dealing with images from diverse sensors, and also with different spatial resolutions.

5 Conclusion

This experiment consisted of analyzing the performance of 80 CNN models, saved at the end of each epoch in the training stage, in the task of classifying preserved and non-preserved scenarios. The results are optimistic, as the models were trained and tested with different composite data sets, with respect to imaging sensor type, spatial resolution, satellite and study area. Where the best model achieved an Accuracy of 0.87 of image classification.

In this perspective, it is intended to improve this model, adding more classes, such as urban, agricultural, preserved, mining and deforestation, to the training and validation data sets, and tests, in order to increase the model's capacity in the task of classifying different scenarios. As well as studying more sophisticated methods to pre-process data; organization and arrangement of the number of filters by convolutional layers; the number of neurons per dense layers, so that you avoid overfitting the weights.

In addition, other evaluation methods will be researched and conducted, either by comparison with other architectures traditionally used for image classification obtained from satellite, either by testing performance in a supervised environment. As well as, using images with different spatial resolutions and sensors both in the training and testing stages, in order to analyze how well the model can do with low, medium or high-resolution images.

Certainly, use the Convolutional Neural Network and other techniques of Deep Learning for Earth observation means investing in a more sustainable and ecological society, since these digital tools boost other research on urban, industrial and environmental growth. With this, it is possible to anticipate possible climatic events, monitor land use and cover, and countless other applications and contributions to our society.

References

1. Anwer, R.M., Khan, F.S., de Weijer, J., Molinier, M., Laaksonen, J.: Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification. ISPRS J. Photogrammetry Remote Sens. **138**, 74–85 (2018). https://doi.org/10.1016/j.isprsjprs.2018.01.023
2. Assis, L., et al.: Terrabrasilis: a spatial data analytics infrastructure for large-scale thematic mapping. ISPRS Int. J. Geo-Information **8**, 513 (2019). https://doi.org/10.3390/ijgi8110513, http://terrabrasilis.dpi.inpe.br
3. Assis, T., Amaral, S.: Landscape and population in deforestation areas of the brazilian cerrado: The case of cerrado baiano. In: Proceedings of the XIX Brazilian Symposium on Remote Sensing (2019). https://bityli.com/3TdqJ
4. Brazil: An area of native vegetation suppressed in the cerrado biome in 2019 was 6,484 km^2. National Institute for Space Research (2019). https://bityli.com/d1Jtd
5. Corrêa, R.: Cerrado: the wealth of tocantins. Portal Tocantins (2020). https://portal.to.gov.br/noticia/2020/9/11/cerrado-a-riqueza-do-tocantins/
6. Ghassemi, S., Magli, E.: Convolutional neural networks for on-board cloud screening. Remote Sens. 11 (2019). https://doi.org/10.3390/rs11121417
7. Gonçalves, R.J.: Mining and territorial fracture of the cerrado in goiás. Élisée - Revista De Geografia Da UEG 9 (2020). https://www.revista.ueg.br/index.php/elisee/article/view/10852
8. Granatyr, J.: Deep learning com python de a a z - o curso completo. Udemy (2019). https://www.udemy.com/share/101uu0AEETcFpXRX8B/
9. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S.: Deep learning for visual understanding: a review. Neurocomputing **187**, 27–48 (2016)
10. Keras: Adam. Keras (2021). https://keras.io/api/optimizers/adam/
11. Keras: Dropout layer. Keras (2021). encurtador.com.br/drxLT
12. Keras: Layer activation functions. Keras (2021). https://keras.io/api/layers/activations/
13. Keras: Maxpooling2d layer. Keras (2021). encurtador.com.br/hszCH
14. Keras: Probabilistic losses. Keras (2021). encurtador.com.br/rCJ23
15. Khatami, R., et al.: A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. Remote Sens. Environ. **177**, 89–100 (2016)
16. Körting, T.S., Namikawa, L., et al.: How to effectively obtain metadata from remote sensing big data? In: GEOBIA 2016: Solutions and Synergies (2016)
17. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. IEEE **86**(11), 2278–2324 (1998)
18. Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A.: Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J. Photogrammetry Remote Sens. **152**, 166–177 (2019)

19. Ma, L., Li, M., Ma, X., Cheng, L., Du, P., Liu, Y.: A review of supervised object-based land-cover image classification. ISPRS J. Photogrammetry Remote Sens. **130**, 277–293 (2017)
20. Mohammadi, M., et al.: A comprehensive survey and taxonomy of the svm-based intrusion detection systems. J. Netw. Comput. Appl. **178**, 102983 (2021)
21. Ribeiro, J. F.; Walter, B.M.T.: Fitofisionomias do bioma cerrado. Cerrado: Ecologia e Flora, EMBRAPA, pp. 152–212 (2008)
22. Santiago Júnior, V.A., Silva, L.A.R., Andrade Neto, P.R.: Testing environmental models supported by machine learning. In: Proceedings of the III Brazilian Symposium on Systematic and Automated Software Testing, SAST 2018, pp. 3–12. Association for Computing Machinery, New York (2018). https://doi.org/10.1145/3266003.3266004
23. Shawky, O.A., Hagag, A., El-Dahshan, E.S.A., Ismail, M.A.: Remote sensing image scene classification using cnn-mlp with data augmentation. Elsevier **221**, 165356 (2020). https://doi.org/10.1016/j.ijleo.2020.165356
24. Zhang, W., Tang, P., Zhao, L.: Remote sensing image scene classification using cnn-capsnet. Remote Sens. **11**, 494 (2019). https://doi.org/10.3390/rs11050494