A High-Spatial Resolution Dataset and Few-shot Deep Learning Benchmark for Image Classification

Mateus de Souza Miranda[1], Lucas Fernando Alvarenga e Silva[2], Samuel Felipe dos Santos[2],
Valdivino Alexandre de Santiago Júnior[1], Thales Sehn Körting[1], and Jurandy Almeida[3]

[1]*Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP – Brazil*
Email: {mateus.miranda, valdivino.santiago, thales.korting}@inpe.br

[2]*Universidade Federal de São Paulo (UNIFESP), São José dos Campos, SP – Brazil*
Email: {e.lucas, felipe.samuel}@unifesp.br

[3]*Universidade Federal de São Carlos (UFSCar), Sorocaba, SP – Brazil*
Email: jurandy.almeida@ufscar.br

Abstract—This paper presents a high-spatial-resolution dataset with remote sensing images of the Brazilian Cerrado for land use and land cover classification. The Biome Cerrado Dataset (Cerra-Data) is a large database created from 150 scenes of the CBERS-4A satellite. Images were created by merging the near-infrared, green, and blue bands. Moreover, pan-sharpening was performed between all the scenes and their respective panchromatic bands, resulting in a final spatial resolution of two meters. A total of 2.5 million tiles of 256x256 pixels were derived from these scenes. From this total, 50 thousand tiles were labeled. We also conducted a few-shot learning experiment considering a training set with only 100 samples, 11 deep neural networks (DNNs), and two traditional machine learning (ML) algorithms, i.e., support vector machine (SVM) and random forest (RF). Results show that the DNN DenseNet-161 was the best model but its performance can be improved if it is used only as a feature extractor, leaving the classification task for the traditional ML algorithms. However, by decreasing the size of the training set, smarter approaches are needed. The labeled subset of CerraData as well as the source code we developed to support this study are available on-line: **https://github.com/ai4luc/CerraData-code-data**.

I. INTRODUCTION

The Cerrado, the second largest Brazilian biome, covers 23% of the national territory, extending over two million square kilometers. It is characterized by typical phytophysiognomies in which woody plants have thick stems, a dark tone, and are twisted but, in other cases, the branches can be angled close to the ground and the tip facing upwards [2]. The term "Cerrado" has been used to refer to the biome, a set of vegetation physiognomies, as well as to a specific type of floristic composition that occurs in the formation of savannas [3], as depicted in Fig. 1.

This biome is located in the tropical zone where climate indirectly affects the characteristics and development of vegetation through the soil [2]. For almost all the Cerrado, the climate is defined in two seasons: wet, occurring more frequently between September and April, and dry, occurring mainly between April and September. It is important to add that at least 40% of the entire area has been converted into pastures and extensive agricultural fields, specifically annual crops such as soybeans and corn [4].

Hence, it is important to analyze the dynamics of land use and land cover (LULC) using methods, techniques, and also based on a robust dataset that offers diversity of data for each class. Several previous studies do not use high-spatial-resolution images set for classification using deep learning (DL) and remote sensing techniques for some Earth observation applications. Moreover, it is important that a significant number of images (tiles) is used in order to have a representative sample of a large biome like Cerrado.

In order to fulfil these gaps, we introduce a novel high-spatial-resolution dataset with optical remote sensing images of the Cerrado for LULC classification, aiming to facilitate access to data ready to support machine learning (ML) and DL models for classification and semantic segmentation. The Biome Cerrado Dataset (CerraData) is a large database, a total of 2.5 million tiles of 256x256 pixels, obtained from 150 scenes made by the Wide Panchromatic and Multispectral Camera (WPM) of the China-Brazil Earth Resources-4A (CBERS-4A) satellite.

Motivated by the success of DL algorithms and their applications to remote sensing tasks [5], an experimental evaluation was conducted considering 50 thousand tiles from CerraData and a few-shot learning setting. We compared the performance of 11 deep convolutional neural networks (CNNs) [6] considering two learning methods: training from scratch and fine-tuning the pre-trained model on ImageNet. We also compared the results by using the top-performing CNNs as a feature extractor only for two traditional ML algorithms: support vector machine (SVM) [7] and random forest (RF) [8]. Finally, we stressed the few-shot learning aspect within our experiment.

II. RELATED WORK

In view of the great diversity of landscapes in the Cerrado, many works select a small study area. Nogueira *et al.* [9] put forward a dataset about the Cerrado, comprising the *Serra do Cipó* region. It has 1,311 images of 64x64 pixels,
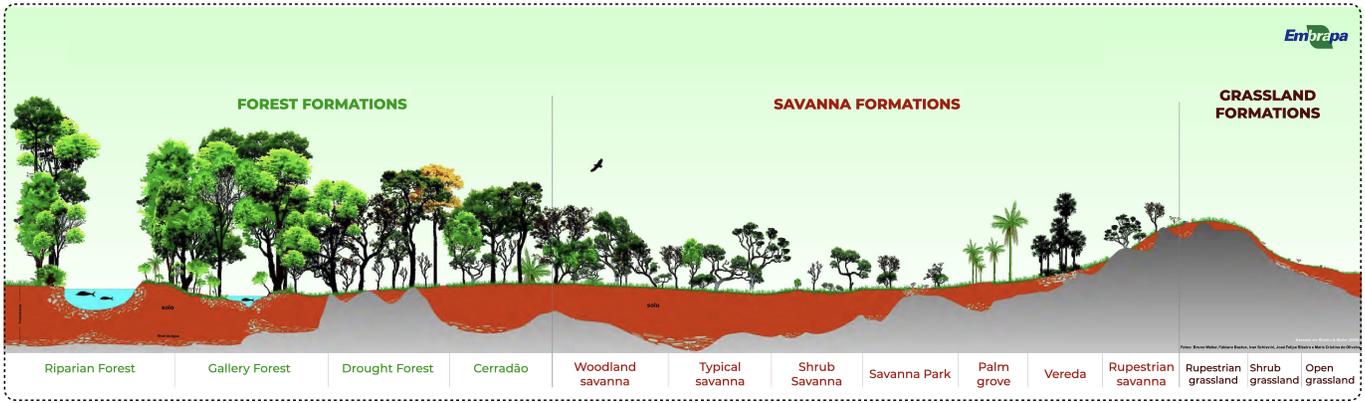
Fig. 1. Phytophysiognomies of the Cerrado biome. Adapted from: [1].

and five meters of spatial resolution, labeled as Agriculture, Arboreal, Herbaceous, and Shrubby Vegetation, showing a great diversity of landscapes and biodiversity. Tiles are formed by the near-infrared (NIR), green (G), and red (R) bands from the RapidEye satellite.

Lewis *et al.* [10] carried out a study on the *Chapada dos Veadeiros* National Park, located in the *Goiás* state, Brazil, aiming to classify eight native physiognomies. Sentinel-1 and 2 and Landsat-8 satellite images were used, considering all bands to calculate the Normalized Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), and Enhanced Vegetation Index 2 (EVI2). They used burned areas and also cloud covering masking has been considered to remove these pixels from the images. Aside from that, seven feature maps with 6,603 pixels with 20 meters spatial resolution on 1,686 sites were processed by the RF algorithm.

Bendini *et al.* [11] proposed a method for hierarchical classification taking into account the time series obtained from 186 observations with 8-day temporal resolution from Enhanced Thematic Mapper Plus (ETM+) and Operational Land Imager (OLI) cameras of the Landsat 7 and 8 satellites. These data cover the regions of western *Bahia*, southeastern *Mato Grosso*, and northeastern *São Paulo* which predominantly belong to the Cerrado biome. Data are composed mainly of Enhanced Vegetation Index (EVI) for biomass estimation and, from this dataset, a dense time series is extracted over cultivated areas.

On the other hand, Neves *et al.* [12] proposed a hierarchical mapping of the Cerrado physiognomies found in the *Brasília Nacional* Park, using a DL algorithm with eight different datasets composed of vegetation index and images in the visible spectrum. Three experiments were performed using different amounts of images and regions (8,415, 10,269, and 7,641 images of 161x161 pixels, including those created with data augmentation). The images were obtained from the WorldView-2 satellite and have two meters of spatial resolution.

In these previous studies, we may highlight that the spatial and temporal resolution of satellite images is the main aspect to create the datasets. However, regarding classification, it is noted that the set of images with the finest spatial resolution (less than three meters) ensures better accuracy, once this gives the ML and DL models more features that better detail the vegetation of the biome, making it easier to distinguish types of savanna or crops in agricultural areas, such as demonstrated in Neves *et al.* [12]. Moreover, a higher number of images (tiles) is beneficial since the Cerrado is a large biome, presenting to the models more features about each class. We address both of these issues in our research.

III. THE CERRADATA DATASET

1) Data Collection: CerraData has scenes from the Cerrado biome which are shown in green color in Fig. 2. However, we considered a study area delimited by the red line, covering the states *Bahia* (BA), *Goiás* (GO), *Maranhão* (MA), *Mato Grosso* (MT), *Tocantins* (TO), and the unit *Distrito Federal* (DF). We gathered 150 scenes without cloud covering, each comprising one path/row in the study site, recorded between early February 2020 to late February 2022 using the WPM of the CBERS-4A satellite. This camera was chosen due to the spatial resolution of the green, blue, and near-infrared bands, which have eight meters of resolution, and due to the panchromatic (PAN) band, since it can be combined with other channels in order to improve the resolution even more. In our case, the final spatial resolution is two meters. Moreover, WPM provides orthorectified scenes, i.e., images with radiometric and geometric correction of the system refined by the use of control points and a digital model of terrain elevation.

The data were preprocessed as follows. We merged the spectral bands of near-infrared (0.77 - 0.89μm; NIR) to R channel, green (0.52 - 0.59μm; G) to G channel, and blue (0.45 - 0.52μm; B) to B channel, from its respective scene. This false-color composition of NIR, G, and B bands was chosen because it highlights the vegetation, in shades of red, from the other objects in the scenes, such as water, soil, and fire scars. Next, we applied pan-sharpening with the Hue Saturation Value (HSV) method using the panchromatic (0.45 - 0.50μm; PAN) band, producing a final image with a spatial
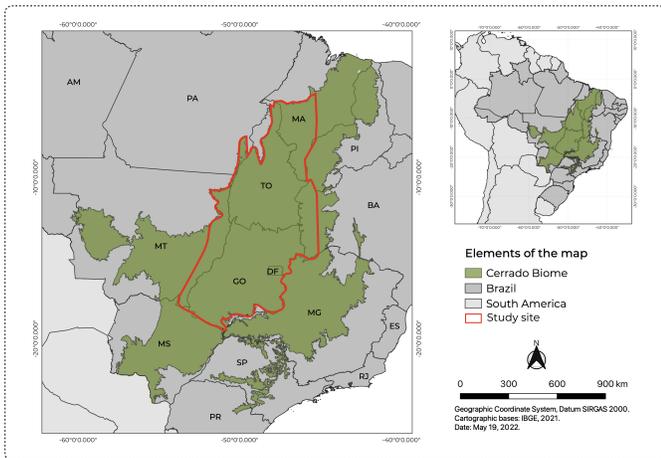
Fig. 2. Study site.

resolution of two meters. After that, the high-spatial-resolution multi-spectral scenes were cropped producing tiles of 256x256 pixels, preserving the geospatial information. Thus, we created approximately 45 thousand tiles per scene. About 20 thousand non-data tiles were removed, i.e., areas of the image with no data. In the end, we created a large dataset with 2.5 million usable tiles. All these processes were performed in Python, except the pan-sharpening, which used the QGIS platform.

These unlabeled data are organized in sections, identified by their origin states, due to data amount. Furthermore, we organized another dataset from this big one containing five LULC classes with 10 thousand tiles for each class, as shown in Fig. 3. Details of such classes are: (*i*) *Cultivated Area*: samples comprising pasture, agriculture, and planted trees; (*ii*) *Forest Formation*: samples are characterized by the predominance of arboreal formation and riparian forests; (*iii*) *Non-Forest Area*: images of urban areas, mining, fire scars, and dunes; (*iv*) *Savanna Formation*: samples of five different phytophysiognomies, i.e., woodland savanna, typical savanna, rupestrian savanna, shrub savanna, and *vereda*; and (*v*) *Water*: river, small lakes, dams, and fish farming ponds.

The tiles were manually selected and annotated based on visual support from the samples of Cerrado physiognomies published by Neves *et al.* [12], as well as the descriptions of vegetation types presented by Ribeiro and Walter [3]. This task is time-consuming and requires a lot of attention. Each file is named with data, position geography, and id, which refers to the crop sequence. It took three months to complete all the aforementioned steps: the first two to merge, pan-sharpening, crop, and remove any non-data value in the images; and the last for labeling, which required an entire month.

IV. Few-shot Learning Experiment

The significant advances obtained by DL algorithms usually depend on huge databases, demand considerable training time, and make use of an expensive computing infrastructure containing graphics processing units (GPUs). Thus, large models are usually costly to train and tune not only in financial terms,

due to the cost of electricity and/or cloud computing time, but also with respect to the environment, due to the carbon footprint required to enable the hardware infrastructure [13].

Moreover, having an adequate amount of labeled data is critical for developing high-performance ML and DL models. However, data labeling is time-consuming and costly too, for instance, it took a whole month to label 2% of CerraData (i.e., 50 thousand out of 2.5 million tiles).

Therefore, training a model with a very small amount of labeled training data is really relevant to address both previous issues. This has been the goal of a subfield known as few-shot learning where there is a very limited number of samples with supervised information for a specific task, such as classification [14]–[16].

Based on these previous motivations, we conducted an extensive experimental evaluation to assess the ability of ML and DL models to learn from a few labeled samples and accurately classify unseen data. These experiments were carried out on 50 thousand labeled tiles from CerraData and considering a few-shot learning setting, in which only 20 samples for each category were used for training.

Using the holdout method and random stratified sampling, we split the 50 thousand labeled tiles from dataset into training, validation, and test sets with 100, 100, and 49,800 tiles, respectively. For a fair comparison, the same splits were used by all the evaluated models. F1-score and accuracy (Acc) were chosen as performance measures. Five replications were performed to ensure statistically sound results. The mean and standard deviation of the performance measures for the test set of all the replications were reported.

We compared the performance of 11 different CNNs: VGG-11, VGG-16, ResNet-18, ResNet-50, SqueezeNet, DenseNet-161, InceptionV3, ShuffleNetv2 1.0, ResNeXt-50, EfficientNet B4, and ConvNeXt-Tiny [6], [17]–[19].

All these models were trained using the following hyper-parameters: 100 epochs; early-stopping monitoring of the F1-score of the validation split for 10 epochs with Δ set to 0 (i.e., any amount of improvement reset the early-stopping counter); batches of 32 images; and stochastic gradient descent (SGD) optimizer with a learning rate of 0.001 and a momentum of 0.9. In addition, two different learning strategies were considered: (*i*) randomly initializing the weights, training from scratch; and (*ii*) initializing the weights from the publicly available ImageNet weights, in this case, fine-tuning the pre-trained model in our new classification task. Also, as a pre-processing step, all the images were normalized using the Z-score normalization. When trained from scratch, the means and standard deviations of the three RGB channels were computed from the training and validation sets. Otherwise, ImageNet statistics were used.

In addition to DNNs, we also tested the SVM and RF classifiers [20]. Firstly, for feature extraction, we passed all the images through the first layers of the best performing CNNs in terms of F1-score. Then, the hyper-parameters of such classifiers were tuned using a grid search on the validation set. As for SVM, we varied the C hyper-parameter between

Fig. 3. Samples from each class.

10^{-1} and 10^3 and the Gamma between 10^{-4} and 10^0, both in steps of powers of 10. Also, three different kernels were considered: linear, polynomial of degree three, and Radial Basis Function (RBF). As for RF, we varied the number of trees in the forest between 10^0 to 10^3 in steps of powers of 10, the number of features used to split a node was searched between 100%, 75%, 50%, 25%, square root and the $\log_2$ from the total amount of features. Both *Gini* impurity and entropy were tested as criteria to measure the quality of the splits. Finally, after the models created by the different settings were evaluated on the validation set, we took the best one according to the F1-score and used it on the test set.

All the experiments were executed in nodes of the SDumont supercomputer (V100 NVIDIA GPUs). We used the release #22.04-py3 of the PyTorch container consisting of the Ubuntu 20.04 distribution, CUDA 11.6, and PyTorch 1.12.

V. RESULTS AND DISCUSSION

1) Best Features and Classifiers: Tables I and II present, at the top, the results on the test set obtained by all 11 DNNs considering the two learning strategies, from-scratch and fine-tuning, respectively. The best DNN, considering the F1-score, is highlighted with * while the second best is with **. We also present, at the bottom, the results for the SVM and RF classifiers with the features extracted by the two best CNNs. The best result for all 15 ML/DL techniques is shown in **bold**.

Regarding the from-scratch strategy (Table I), we can clearly observe that the DNN DenseNet-161 achieved the best F1-score (76.07%) and Acc (76.13%), being approximately 2% better than the second best, ResNet-18. It is also noted that deeper models (e.g., VGG-16 and ResNet-50) underperformed their shallow versions (e.g., VGG-11 and ResNet-18). Although some approaches performed very well in terms of F1-score and Acc, others presented a low mean and a high standard deviation for both measures (e.g., VGG-11, VGG-16, ResNet-50, ShuffleNetv2 1.0, and EfficientNet B4).

However, we can see improvements when we use CNNs as feature extractors only and rely on classical ML algorithms as classifiers. Enhancement of at least 2% were detected for DenseNet-161 and almost 3% for ResNet-18. Especially, the combination of DenseNet-161 as feature extractor and RF as classifier reached the best result (78.18%).

TABLE I
PERFORMANCE ASSESSMENT: FROM-SCRATCH APPROACH.

Feature Extraction	Classifier	F1-score	Acc
VGG-11	DNN	45.50 ± 22.1	50.97 ± 18.9
VGG-16	DNN	51.68 ± 20.1	56.00 ± 17.5
ResNet-18**	DNN	74.58 ± 3.09	74.86 ± 3.06
ResNet-50	DNN	59.44 ± 16.9	61.51 ± 14.5
SqueezeNet	DNN	58.32 ± 7.10	62.23 ± 5.94
DenseNet-161*	DNN	76.07 ± 1.55	76.13 ± 1.63
InceptionV3	DNN	64.81 ± 7.26	66.48 ± 5.45
ShuffleNetv2 1.0	DNN	49.57 ± 12.5	54.71 ± 9.51
ResNeXt-50	DNN	70.11 ± 3.65	70.66 ± 3.41
EfficientNet B4	DNN	49.10 ± 15.5	51.47 ± 12.6
ConvNeXt-Tiny	DNN	54.77 ± 1.50	58.89 ± 1.14
DenseNet-161	**RF**	$\mathbf{78.18 \pm 1.31}$	$\mathbf{78.13 \pm 1.37}$
DenseNet-161	SVM	77.49 ± 1.87	77.47 ± 1.94
ResNet-18	RF	77.47 ± 3.20	77.42 ± 3.32
ResNet-18	SVM	77.22 ± 3.23	77.17 ± 3.39

As for the fine-tuning strategy (Table II), except for ShuffleNetv2 and EfficientNet B4, all other methods presented consistent results with great enhancement compared to the from-scratch strategy. Among all 11 DNNs, VGG-16 has the best F1-score (86.41%) followed by DenseNet-161 (86.38%). Regarding Acc, they switch positions: DenseNet-161 was the best (86.45%) and VGG-16 was the second best (86.38%).

We can also note that F1-score and Acc values are close to each other for all models (except ShuffleNetv2), give out that they are precise, have good recall, and correctly classify most unseen data when knowledge from a general classification task is transferred to a remote sensing classification task. Also, the issue of high standard deviation values observed in the from-scratch strategy is alleviated when pre-trained models are used.

We observe further that using the best CNNs as feature extractors only and performing the classification with the classical ML algorithms was a good strategy. Even though the gaps are small, the best result was achieved when DenseNet-161 extracted the features and SVM performed the classification.

Based on all these results, considering the CerraData and the proposed experimental procedures, we may conclude: (*i*) DenseNet-161 can be regarded as the best CNN among all the

Feature Extraction	Classifier	F1-score	Acc
VGG-11	DNN	83.84 ± 2.60	83.93 ± 2.66
VGG-16*	DNN	86.41 ± 1.22	86.38 ± 1.17
ResNet-18	DNN	83.87 ± 1.91	83.98 ± 1.81
ResNet-50	DNN	85.94 ± 2.18	86.03 ± 2.11
SqueezeNet	DNN	84.49 ± 2.82	84.51 ± 2.80
DenseNet-161**	DNN	86.38 ± 1.45	86.45 ± 1.41
InceptionV3	DNN	77.85 ± 2.92	78.22 ± 2.65
ShuffleNetv2 1.0	DNN	15.16 ± 3.14	24.08 ± 5.90
ResNeXt-50	DNN	84.85 ± 2.08	84.89 ± 2.11
EfficientNet B4	DNN	56.42 ± 5.89	57.81 ± 5.43
ConvNeXt-Tiny	DNN	86.04 ± 2.35	86.10 ± 2.32
VGG-16	RF	82.51 ± 1.05	82.95 ± 0.91
VGG-16	SVM	83.59 ± 0.97	83.87 ± 0.84
DenseNet-161	RF	86.16 ± 0.98	86.22 ± 0.90
DenseNet-161	**SVM**	$\mathbf{86.57 \pm 1.36}$	$\mathbf{86.58 \pm 1.30}$

DNNs evaluated, since it was the best in both metrics for the from-scratch strategy, obtaining the highest Acc and second best F1-score in the fine-tuning strategy, in addition to being the most suitable feature extractor; (ii) using CNNs as feature extractor only and classical ML algorithms as classifiers may worth the effort; (iii) transfer learning by fine-tuning the CNNs showed again to be a promising direction.

2) Limits of Learning from Few Samples: We also explored the few-shot learning capabilities of the evaluated models by varying the number of training samples, decreasing it until reaching the minimum, i.e., 1 sample per class. In particular, we evaluated the models that achieved the best results when training from scratch (DenseNet-161+DNN and DenseNet-161+RF) and also fine-tuning based on ImageNet (VGG-16+DNN and DenseNet-161+SVM). They were tested with smaller and smaller training sets, i.e., containing 20, 15, 10, 5, 4, 3, 2 samples, and, finally, only 1 sample per class.
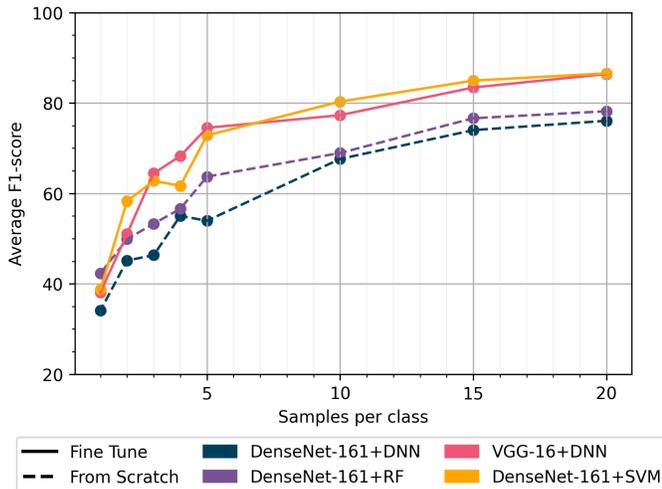

Fig. 4. Few-shot stressing of the best evaluated models.

The results are presented in Fig. 4. As expected, the fewer training examples, the worse the model performance, dropping the F1-score, on average, from 81% (20 samples per class) to 38% (1 sample per class). Moreover, by increasing the training set size by a small amount (from 1 to 4 samples per class), we could observe increases of almost 25% in the F1-score. Despite the remarkable characteristics of the DNNs and classical ML algorithms evaluated in this study, it is clear that smarter strategies are important to obtain the maximum benefit according to the few-shot learning philosophy.

3) Visualization Analysis: The difficulty level of a dataset can also be grasped by visualizing the class separability of the learned feature representations. Therefore, a common strategy is to project the features learned by a DL model into a lower-dimensional space, typically 2D or 3D, using a dimensionality reduction method, like UMAP [21] or t-SNE [22].

As for feature extraction, we passed all the 50 thousand labeled tiles through the first layers of the VGG-16+DNN model, which achieved the highest F1-score among all 11 DNNs. Then, we used UMAP to project the original (high-dimensional) feature space into the 2D space.

The projections of the feature spaces spanned by the VGG-16+DNN models learned from training sets with 20, 4, and 1 samples per class, respectively, are plotted in Fig. 5. We notice that these results are in agreement with those obtained by varying the size of the training set (Fig. 4), in which the more samples per class used for training, the better the class separability, improving the classification performance. Although the larger training set (i.e., 20 samples/class) yields a feature space with tighter and more spaced groups, they clearly overlap for most classes, showing that CerraData is a challenging dataset.

VI. CONCLUSIONS

This paper presented CerraData, a high-spatial-resolution dataset with satellite images covering the Brazilian Cerrado. The motivation to develop this dataset was because we believe that there is a lack of high-spatial-resolution benchmarks particularly tailored to the Cerrado biome. Overall it has 2.5 million tiles providing a significant amount of images for future approaches addressing remote sensing classification.

We performed an extensive few-shot learning experiment with two settings: from-scratch and fine-tuning. Altogether, 11 DNNs were considered as feature extractors plus classifiers, and the two best DNNs played the role of feature extractors, leaving the classification itself for two traditional ML algorithms, i.e., RF and SVM. Results show that the DNN DenseNet-161 was the best model but its performance can be improved if it is used only as a feature extractor.

In the initial design of our few-shot learning experiment, where we defined 20 samples per class in the training set, we might say that, in general, these traditional supervised learning algorithms (DNNs, RF, SVM) yield good results, particularly in the fine-tuning approach. But stressing the limits of few-shot, the performances of the best strategies were poor, and hence smarter approaches are important.
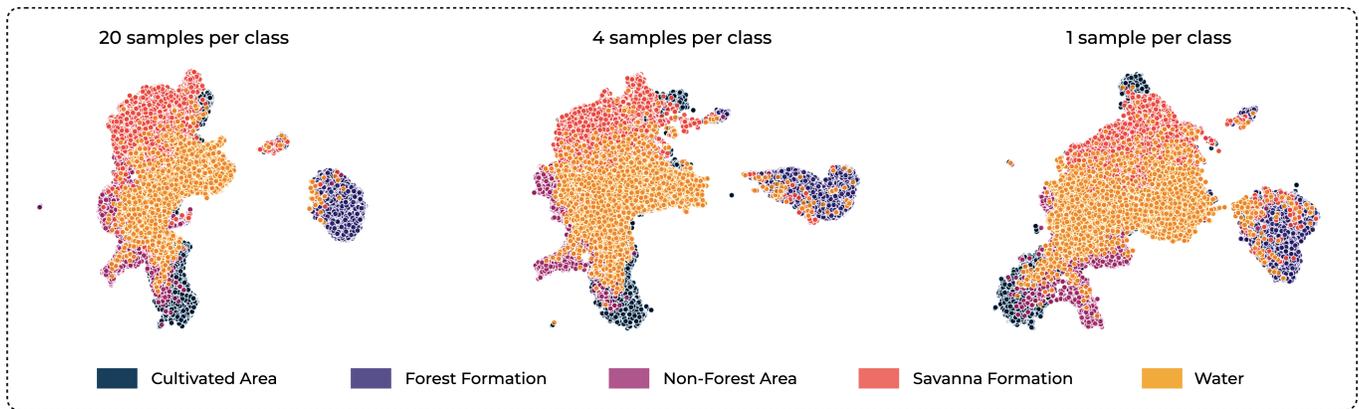
Fig. 5. UMAP projections (2D) of VGG-16+DNN model from 6 different splittings.

Future work includes exploiting all the 2.5 million tiles of CerraData with methods that do not require additional manual labeling efforts. We will investigate other DNNs more suitable for few-shot learning. There is also room for open set learning (OSL) where we may consider new classes of the biome Cerrado and submit them to the algorithms and perceive their performances.

REFERENCES

[1] *Bioma Cerrado*, Empresa Brasileira de Pesquisa Agropecuária, Parque Estação Biológica - PqEB, Brasília, 70770-901, 2008. [Online]. Available: https://www.embrapa.br/cerrados/

[2] G. Eiten, *Vegetação do Cerrado In: Pinto, M. (ORG.). Cerrado: caracterização, ocupação e perspectivas*. Brasília, DF, Brazil: Editora Universidade de Brasília, 1990, 657 p.

[3] J. F. Ribeiro and B. M. T. Walter, *As Principais Fitofisionomias do Bioma Cerrado In: SANO, S. M.; ALMEIDA, S. P. de; RIBEIRO, J. F. (Ed.). Cerrado: ecologia e flora*. Brasília, DF, Brazil: EMBRAPA, 2008, 876 p.

[4] A. Reatto, J. a. R. Correia, S. T. Spera, and E. Martins, *Solos do Bioma Cerado: aspectos pedológicos. In: SANO, S. M.; ALMEIDA, S. P. de; RIBEIRO, J. F. (Ed.). Cerrado: ecologia e flora*. Brasília, DF, Brazil: EMBRAPA, 2008, 876 p.

[5] G. Tsagkatakis, A. Aidini, K. Fotiadou, M. Giannopoulos, A. Pentari, and P. Tsakalides, "Survey of deep-learning approaches for remote sensing observation enhancement," *Sensors*, vol. 19, no. 18, 2019.

[6] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial intelligence review*, vol. 53, no. 8, pp. 5455–5516, 2020.

[7] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 1999.

[8] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[9] K. Nogueira, J. A. Dos Santos, T. Fornazari, T. S. F. Silva, L. P. Morellato, and R. d. S. Torres, "Towards vegetation species discrimination by using data-driven descriptors," in *2016 9th IAPR Workshop on Pattern Recogniton in Remote Sensing (PRRS)*. Ieee, 2016, pp. 1–6.

[10] K. Lewis, F. de V. Barros, M. B. Cure, C. A. Davies, M. N. Furtado, T. C. Hill1, M. Hirota, D. L. Martin, G. G. Mazzochini, E. T. A. Mitchard, C. B. R. Munhoz, R. S. Oliveira, A. B. Sampaio, N. A. Saraiva, I. B. Schmidt, and L. Rowland, "Mapping native and non-native vegetation in the brazilian cerrado using freely available satellite products," *Scientific Reports*, 2022.

[11] H. do Nascimento Bendini, L. M. G. Fonseca, M. Schwieder, T. S. Körting, P. Rufin, I. D. A. Sanches, P. J. L. ao, and P. Hostert, "Detailed agricultural land classification in the brazilian cerrado based on phenological information from dense satellite image time series," *Int J Appl Earth Obs Geoinformation*, 2019.

[12] A. K. Neves, T. S. Körting, L. M. G. Fonseca, A. R. Soares, C. D. Girolamo-Neto, and C. Heipke, "Hierarchical mapping of brazilian savanna (cerrado) physiognomies based on deep learning," *Journal of Applied Remote Sensing*, vol. 15, no. 4, 2021.

[13] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for modern deep learning research," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 09, pp. 13 693–13 696, Apr. 2020. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/7123

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[15] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F.-F. Li, "Imagenet large scale visual recognition challenge," *IJCV*, vol. 115, no. 3, pp. 211–252, 2015.

[16] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *ECCV*, 2010, pp. 213–226.

[17] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size," *arXiv:1602.07360*, 2016.

[18] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[19] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," *arXiv preprint arXiv:2201.03545*, 2022.

[20] L. M. Fonseca, T. S. Körting, H. do N. Bendini, C. D. Girolamo-Neto, A. K. Neves, A. R. Soares, E. C. Taquary, and R. V. Maretto, "Pattern recognition and remote sensing techniques applied to land use and land cover mapping in the brazilian savannah," *Pattern Recognition Letters*, vol. 148, pp. 54–60, 2021.

[21] L. McInnes and J. Healy, "UMAP: uniform manifold approximation and projection for dimension reduction," *CoRR*, vol. abs/1802.03426, 2018.

[22] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.