sid.inpe.br/mtc-m21c/2021/03.30.18.49-TDI

HIERARCHICAL MAPPING OF BRAZILIAN SAVANNA (CERRADO) PHYSIOGNOMIES BASED ON DEEP LEARNING

Alana Kasahara Neves

Doctorate Thesis of the Graduate Course in Remote Sensing, guided by Drs. Thales Sehn Körting, and Leila Maria Garcia Fonseca, approved in March 22, 2021.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34R/44DTSUS>

INPE
São José dos Campos
2021

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

sid.inpe.br/mtc-m21c/2021/03.30.18.49-TDI

HIERARCHICAL MAPPING OF BRAZILIAN SAVANNA (CERRADO) PHYSIOGNOMIES BASED ON DEEP LEARNING

Alana Kasahara Neves

Doctorate Thesis of the Graduate Course in Remote Sensing, guided by Drs. Thales Sehn Körting, and Leila Maria Garcia Fonseca, approved in March 22, 2021.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34R/44DTSUS>

INPE
São José dos Campos
2021

Cataloging in Publication Data

Neves, Alana Kasahara.

N414h Hierarchical mapping of Brazilian Savanna (Cerrado) physiognomies based on Deep Learning / Alana Kasahara Neves. – São José dos Campos : INPE, 2021.

xxiv + 98 p. ; (sid.inpe.br/mtc-m21c/2021/03.30.18.49-TDI)

Thesis (Doctorate in Remote Sensing) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2021.

Guiding : Drs. Thales Sehn Körting, and Leila Maria Garcia Fonseca.

1. Spatial context. 2. High spatial resolution image. 3. GEOBIA. 4. Semantic segmentation. 5. Convolutional Neural Network. I.Title.

CDU 528(213.54)

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

Serviço de Pós-Graduação

DEFESA FINAL DE TESE DE ALANA KASAHARA NEVES

BANCA Nº 041/2021, REG 133558/2017

No dia 22 de março de 2021, as 09h, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Doutora em Sensoriamento Remoto. O trabalho precisa da incorporação das correções sugeridas pela Banca Examinadora e revisão final pelo(s) orientador(es).

Título: "HIERARCHICAL MAPPING OF BRAZILIAN SAVANNA (CERRADO) PHYSIOGNOMIES BASED ON DEEP LEARNING"

Eu, Luiz Eduardo Oliveira e Cruz de Aragão, como Presidente da Banca Examinadora, assino esta ATA em nome de todos os membros.

Membros da banca:

Dr. Luiz Eduardo Oliveira e Cruz de Aragão - Presidente - INPE

Dr. Thales Sehn Körting - Orientador - INPE

Dr. Leila Maria Garcia Fonseca - Orientador - INPE

Dr. Sidnei João Siqueira Sant'Anna - Membro da banca - INPE

Dr. Cleber Gonzales de Oliveira - Membro Externo - Visiona Tecnologia Espacial S.A.

Dra. Ane Auxiliadora Costa Alencar - Membro Externo – IPAM

Documento assinado eletronicamente por **Luiz Eduardo Oliveira E Cruz de Aragão**, **Chefe da Divisão de Observação da Terra e Geoinformática**, em 29/03/2021, às 08:31 (horário oficial de Brasília), com fundamento no art. 6º do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.mctic.gov.br/verifica.html, informando o código verificador **6601230** e o código CRC **BBEE4060**.

"I still believe, in spite of everything, that people are truly good at heart. It's utterly impossible for me to build my life on a foundation of chaos, suffering and death. I see the world being slowly transformed into a wilderness, I hear the approaching thunder that, one day, will destroy us too, I feel the suffering of millions. And yet, when I look up at the sky, I somehow feel that everything will change for the better, that this cruelty too shall end, that peace and tranquility will return once more. In the meantime, I must hold on to my ideals."

Anne Frank

*To my parents **Silvia** and **David** and to my grandmother **Agrippina***

ACKNOWLEDGEMENTS

(In Portuguese)

Agradeço a Deus por ter me dado condições físicas e mentais para enfrentar essa trajetória.

Ao meu orientador, Dr. Thales Körting, por ser paciente e generoso, por acreditar no meu potencial desde o mestrado, por me apoiar nas minhas decisões e por sempre contribuir ativamente com o meu trabalho.

À minha coorientadora, Dra. Leila Fonseca, por me acompanhar na trajetória acadêmica desde o mestrado, sempre me incentivando a conquistar novas metas e objetivos e não medindo esforços para promover todo o suporte necessário para a realização do trabalho.

Aos pesquisadores Dr. Luiz Aragão, Dr. Sidnei Sant'Anna, Dra. Ane Alencar e Dr. Cleber Oliveira por aceitarem participar da banca e contribuir com este trabalho. Ter meu trabalho avaliado por vocês é uma grande responsabilidade, mas também uma honra.

Ao INPE, seus funcionários, coordenação do curso de Sensoriamento Remoto e todos que colaboram direta ou indiretamente para que o instituto, mesmo em tempos difíceis, tenha ensino e pesquisa de qualidade.

Ao CNPq e ao DAAD pelas bolsas de doutorado e doutorado sanduíche, respectivamente. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

A todos os meus antigos mestres e orientadores que contribuíram com a minha trajetória acadêmica, ainda em Belém-PA: Dr. Claudio Almeida, Dr. Adriano Marlisom Sousa, Dra. Alessandra Gomes, Dr. Igor Narvaes, Dra. Janaina Maia, Dr. Francisco Berrêdo e Me. Daniel Bentes. A contribuição de vocês foi fundamental nessa caminhada.

A todos os meus amigos e colegas, inpeanos ou não, que estando próximos ou distantes, me incentivaram a concluir esse trabalho.

À toda minha família, em especial meus pais (Silvia e David), minha irmã Celine

e minha vó Agripina, por todo o suporte e incentivo incondicional. Uma menção honrosa à minha sobrinha Marjorie, que sempre proporciona momentos de alegria e descontração.

Ao meu namorado Cesare, pelo suporte, paciência e companhia em todos os momentos.

À Miki, minha felina, por ser minha inseparável companheira durante a redação dessa tese.

(In English)

I would like to thank Prof. Dr. Christian Heipke for all support and guidance during my time in Hannover, Germany. I am grateful to all IPIans, especially Dr. Sanaz Vajedian, Dennis Wittich, Lin Chen, Mirjana Voelsen, Dr. Franz Rottensteiner, Dr. Gilson Costa, Dominic Clermont, Christian Kruse, Claudia Sander and Annette Radtke. Thank you for the collaboration and the moments of fun (lunches, coffee time, company outing, Christmas market).

Special thanks to Anne Eggers, my flatmate in Hannover, for the friendship, the talks, the German food and all help to deal with the German language and bureaucracy.

ABSTRACT

The Brazilian Savanna, also known as Cerrado, is considered one of the global hotspots for biodiversity conservation and plays an important role as carbon stock, due to its above and below-ground biomass. The Cerrado vegetation is composed by a mosaic of ecosystems, which comprises since natural grasslands until dense forests. There is a vegetation gradient with a wide variation in structure, density and biomass, which generates several types of vegetation, known as physiognomies. According to the Ribeiro and Walter classification system, there are three major groups of ecosystems (Grassland, Savanna and Forest), which can be divided into 11 physiognomies and 14 additional sub-types of physiognomies, resulting in 25 physiognomic types. Monitoring the Cerrado vegetation cover in a large scale, using Remote Sensing imagery, is still a challenge due to the high spatial and temporal variability of the vegetation types and their spectral similarity. Two aspects of the Cerrado physiognomies are relevant to create a novel classification method: its classification system hierarchy and the relative context where each physiognomy occurs. Two classification techniques that considers the spatial context have been used in the Remote Sensing field: GEOBIA and Deep Learning. Thus, the general objective of this study is to develop and evaluate a novel method based on Deep Learning to hierarchically classify the Cerrado physiognomies, according to the classification system proposed by Ribeiro and Walter, in the Brasília National Park, a federal environmental Protected Area. Several spectral channels were tested as input datasets to evaluate their importance and contribution in the classification task and all experiments used a WorldView-2 multispectral image (2 meters spatial resolution). To demonstrate the potential of Deep Learning techniques in the Cerrado vegetation discrimination, hierarchical and non-hierarchical GEOBIA approaches were initially performed to classify seven physiognomies. In addition to the spectral bands, five vegetation indices, three fractions of the Linear Spectral Mixture Model, three components of the Tasseled Cap transformation and six texture features were used as features. Compared to a GEOBIA non-hierarchical approach, the GEOBIA hierarchical approach achieved an overall accuracy of 2.5 percentage points higher (66.4% and 68.9%, respectively). In the Deep Learning approach, an adapted U-net architecture was used to hierarchically classify the physiognomies. The dataset composed of RGB bands plus the 2-band Enhanced Vegetation Index (EVI2) achieved the best performance and was used to perform the hierarchical classification. In the first level, which identified Forest, Savanna and Grassland, the overall accuracy was 92.8%. For detailed Savanna and Grassland physiognomies (second level of classification), the overall accuracies were 86.1% and 85.0%, respectively. The Brasília National Park final map obtained in this study has ten physiognomies: Gallery Forest, Woodland Savanna, Typical Savanna, Shrub Savanna, Rupestrian Savanna, *Vereda*, Rupestrian Grassland, Shrub Grassland, Open Grassland and Humid Open Grassland. The misclassified areas are mainly related to transition regions between the physiognomies. Deep Learning techniques were able to understand and well represent the physiognomy patterns. To the best of our knowledge, this work was the first one that used

Deep Learning to discriminate the Cerrado physiognomies in this level of detail. Besides, the accuracy rates obtained here outperformed other works that applied traditional Machine Learning algorithms and GEOBIA for this task.

Keywords: Spatial context. High spatial resolution image. GEOBIA. Semantic segmentation. Convolutional Neural Network. U-net.

MAPEAMENTO HIERÁRQUICO DAS FITOFISIONOMIAS DA SAVANA BRASILEIRA (CERRADO) BASEADO EM *DEEP LEARNING* (APRENDIZAGEM PROFUNDA)

RESUMO

A Savana brasileira, conhecida como Cerrado, é considerada um *hotspot* global para a conservação da biodiversidade, e exerce um importante papel como estoque de carbono, devido à sua biomassa acima e abaixo do solo. A vegetação do Cerrado é composta por um mosaico de ecossistemas, que abrange desde campos naturais até densas florestas. Existe um gradiente de vegetação com ampla variação em estrutura, densidade e biomassa, que geram diferentes tipos de vegetação, chamados de fitofisionomias. De acordo com o sistema de classificação proposto por Ribeiro e Walter, existem três grupos principais de ecossistemas (Floresta, Savana e Campo), que podem ser divididos em 11 fitofisionomias e 14 subtipos adicionais, resultando em 25 tipos de fitofisionomias. O monitoramento da vegetação do Cerrado em larga escala, usando imagens de sensoriamento remoto, ainda é um desafio devido à alta variabilidade espacial e temporal e à similaridade espectral das fitofisionomias. Dois aspectos da vegetação do Cerrado são relevantes para a criação de um novo método de classificação: a hierarquia do sistema de classificação e o contexto espacial em que cada fitofisionomia ocorre. Duas técnicas de classificação que consideram o contexto espacial têm sido utilizadas na área de Sensoriamento Remoto: GEOBIA e *Deep Learning*. Assim, o objetivo geral deste trabalho é desenvolver e avaliar um novo método baseado em *Deep Learning* para classificar hierarquicamente as fitofisionomias do Cerrado, de acordo com o sistema de classificação proposto por Ribeiro e Walter, existentes no Parque Nacional do Brasília, uma Unidade de Conservação federal. Várias bandas e atributos espectrais foram testados como dados de entrada para avaliar suas contribuições na classificação e todos os experimentos usaram uma imagem multiespectral WorldView-2 (resolução espacial de 2 metros). Para demonstrar o potencial das técnicas de *Deep Learning* para discriminar a vegetação do Cerrado, inicialmente uma abordagem usando GEOBIA para classificar sete fitofisionomias foi realizada. Além das bandas espectrais, cinco índices de vegetação, três frações do Modelo Linear de Mistura Espectral, três componentes da transformação *Tasseled Cap* e seis atributos de textura foram usados como atributos. Em comparação com uma abordagem não hierárquica de GEOBIA, a abordagem hierárquica de GEOBIA obteve uma acurácia global 2,5 pontos percentuais maior (66,4% e 68,9%, respectivamente). Na abordagem com *Deep Learning*, uma arquitetura U-net adaptada foi usada para classificar hierarquicamente as fitofisionomias. O conjunto de dados composto pelas bandas RGB mais o *2-band Enhanced Vegetation Index* (EVI2) obteve o melhor desempenho e foi usado para realizar a classificação hierárquica. No primeiro nível, que identificou Floresta, Savana e Campo, a acurácia global foi 92,8%. Para as fitofisionomias detalhadas de Savana e Campo (segundo nível de classificação), as acurácias globais foram de 86,1% e 85,0 %, respectivamente. O mapa final do Parque Nacional de Brasília obtido neste trabalho possui dez fitofisionomias: Mata de Galeria, Cerrado Denso, Cerrado Típico, Cerrado Ralo, Cerrado Rupestre, Vereda, Campo Rupestre, Campo Sujo, Campo Limpo e Campo

Limpo Úmido. As áreas classificadas incorretamente estão relacionadas principalmente a regiões de transição entre as fitofisionomias. As técnicas de *Deep Learning* foram capazes de entender e representar bem os padrões das fitofisionomias. Até onde sabemos, esse foi o primeiro trabalho que usou *Deep Learning* para discriminar as fitofisionomias do Cerrado nesse nível de detalhamento. Além disso, as acurácias aqui obtidas superaram as de outros trabalhos que aplicaram algoritmos tradicionais de aprendizado de máquina e GEOBIA para essa tarefa.

Palavras-chave: Contexto espacial. Imagem de alta resolução espacial. GEOBIA. Segmentação semantica. Rede neural convolucional. U-net.

LIST OF FIGURES

LIST OF TABLES

Page

LIST OF ABBREVIATIONS

BNP	–	Brasília National Park
CEI	–	Crop Enhanced Index
CNN	–	Convolutional Neural Network
EVI	–	Enhanced Vegetation Index
EVI2	–	2-band Enhanced Vegetation Index
FCNN	–	Fully Convolutional Neural Network
FIP	–	Forest Investment Program
FLAASH	–	Fast Line-of-sight Atmospheric Analysis of Hypercubes
GEOBIA	–	Geographic Object-Based Image Analysis
GLCM	–	Grey Level Co-occurrence Matrix
IBGE	–	Brazilian Institute of Geography and Statistics
IFOV	–	Instantaneous Field of View
INPE	–	National Institute for Space Research
LAI	–	Leaf Area Index
LiDAR	–	Light Detection And Ranging
LSMM	–	Linear Spectral Mixture Model
LSTM	–	Long Short Term Memory
MDA	–	Multiple Discriminant Analysis
MLP	–	Multilayer Perceptron
MMA	–	Ministry of Environment
MODIS	–	Moderate Resolution Imaging Spectroradiometer
MSAVI2	–	Modified Soil-Adjusted Vegetation Index 2
NDVI	–	Normalized Difference Vegetation Index
NIR	–	Near Infrared
NN	–	Nearest Neighbor
OA	–	Overall Accuracy
PPCerrado	–	Action Plan for the Prevention and Control of Deforestation
	–	and Forest Fires in the Cerrado
ReLU	–	Rectified Linear Unit
RS	–	Remote Sensing
SAM	–	Spectral Angle Mapper
SAVI	–	Soil-Adjusted Vegetation Index
SLIC	–	Simple Linear Iterative Clustering
SVM	–	Support Vector Machine
TC	–	Tasseled Cap
UFG	–	Federal University of Goiás
UFMG	–	Federal University of Minas Gerais
VI	–	Vegetation Index

CONTENTS

Page

1 INTRODUCTION

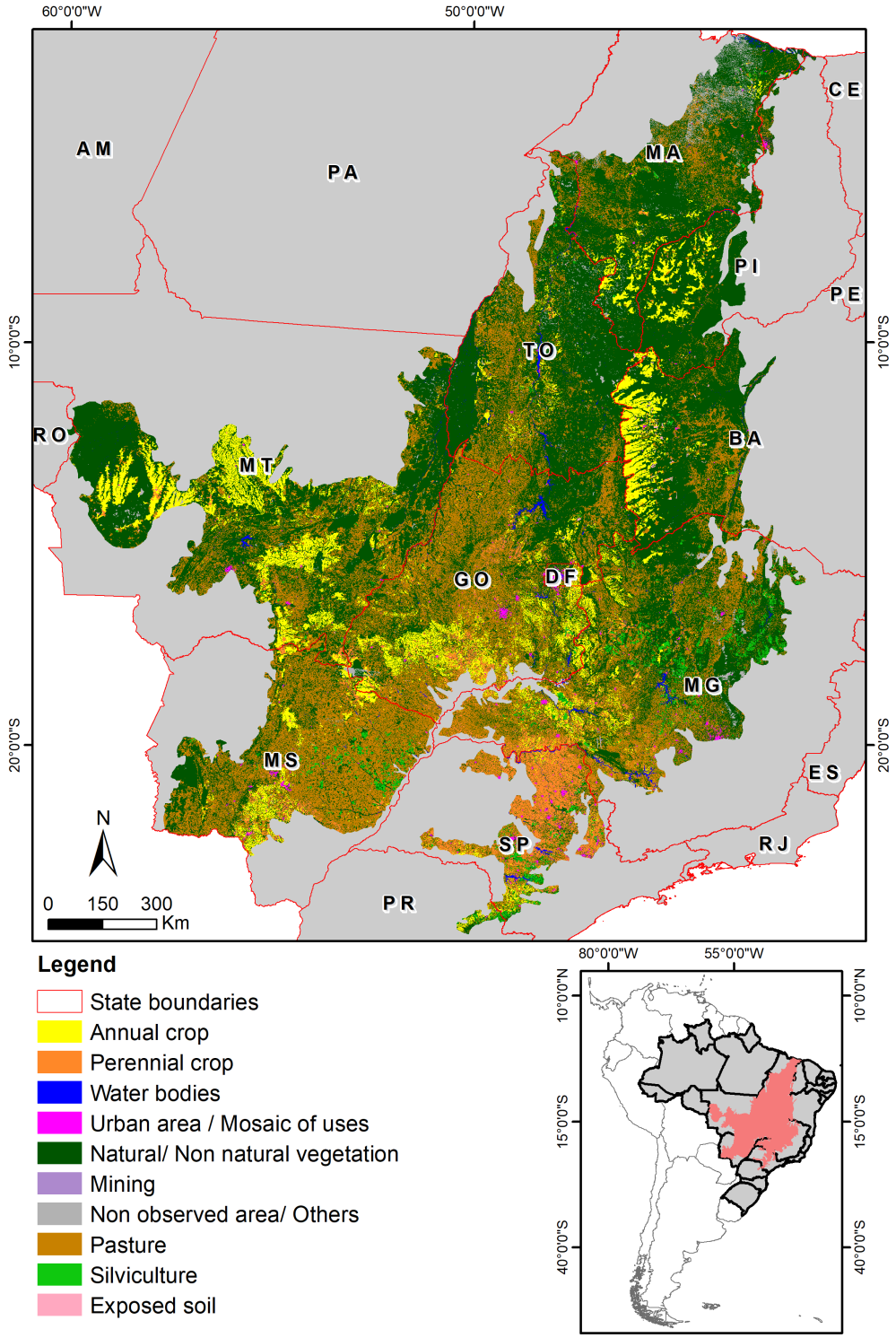
The ecosystems of Savanna cover approximately 20% of the Earth's terrestrial area. In tropical regions, they are rich in biodiversity (STRASSBURG et al., 2017) and water resources (FERREIRA et al., 2011) and play an important role in carbon stock, due to their content of above and below-ground biomass (RIBEIRO et al., 2011). However, many aspects of the Savannas (e.g., carbon cycles) are still not well understood when compared to other tropical ecosystems, like the Amazon forest. This lack of knowledge hinders the complete understanding of the global carbon cycle (ADUAN et al., 2003; GWENZI; LEFSKY, 2014) and justifies the necessity of further investigations in these topics.

The Brazilian Savanna, also known as Cerrado, is the second largest biome in the country, occupying an area of 1,983,017 km^2, approximately 23,3% of the national territory. This biome comprises the entire Federal District and parts of the following states: Goiás (98%), Tocantins (91%), Maranhão (65%), Mato Grosso do Sul (62%), Minas Gerais (54%), Piauí (53%), Mato Grosso (37%), São Paulo (19%), Bahia (18%), Paraná (2%) and Rondônia (1%) (IBGE, 2019). Considered one of the global hotspots for biodiversity conservation (STRASSBURG et al., 2017), Cerrado's flora has more than 12 thousand species, of which almost 40% are endemic (i.e., they do not occur in any other part of the planet) (JBRJ, 2015). Despite the ecological importance of Cerrado, only 8.6% of its natural vegetation is in environmental Protected Areas and just 3.1% in Protected Areas of Integral Protection (MMA, 2010).

Approximately 47% of the biome's vegetation has already been converted to other land uses, such as Planted Pasture (29%) and Annual Crop (9%) (Figure 1.1) (INPE, 2015). Moreover, deforestation rates in the Cerrado have been higher than those in the Amazon in 9 of the 12 years in the period from 2008 to 2019 (INPE, 2020). This heavy loss of native vegetation brings severe environmental consequences, such as vegetation fragmentation, habitat loss (FRANÇOSO et al., 2015), reduction of water yield and carbon stocks (GRACE et al., 2006; RESENDE et al., 2019) and several other negative impacts in Cerrado ecosystem services (RESENDE et al., 2019). In this scenario, accurate mapping of Cerrado vegetation is essential to support policies against deforestation and consequently maintain the provision of ecosystem services, since these maps are crucial for assessing biodiversity, improving Carbon stock estimation within the biome and guiding conservation policies.

The conservation of the Cerrado vegetation is beyond the creation and implementation of laws against deforestation. It is necessary that the decision makers stimulate

Figure 1.1 - Land use and land cover in Cerrado in 2013.



SOURCE: Adapted from INPE (2015).

the proper management of the remaining natural vegetation areas, their biodiversity and ecosystem services. In this scenario, in 2009, the National Policy on Climate Change (Brazilian Law nº 12.187) was created. It represents the Brazilian government's commitment to the United Nations to reduce the greenhouse gases emissions through the development of a low carbon emission economy. One of the axes of this Policy is the Action Plan for the Prevention and Control of Deforestation and Forest Fires in the Cerrado (PPCerrado). This plan has three main lines of action: sustainable production; monitoring and control; and protected areas and land use planning (MMA, 2014). To achieve them, many programs were created, including the Forest Investment Program (FIP) Cerrado, which focuses on promoting sustainable land use and improving forest management to reduce carbon emissions. The National Institute for Space Research (INPE), together with the Federal Universities of Goias and Minas Gerais (UFG and UFMG), are responsible for the execution of one of the four objectives created by FIP Cerrado, which aims to implement a system for forest fire prevention and Cerrado vegetation monitoring (THE WORLD BANK, 2016).

Monitoring the Cerrado vegetation cover in a large scale, using Remote Sensing (RS) imagery, is still a challenge due to the high spatial and temporal variability of the types of vegetation and their spectral similarity (JACON et al., 2017). The Cerrado vegetation is composed by a mosaic of ecosystems, which comprises since natural grasslands, with a predominance of herbaceous vegetation, until dense forests. There is a gradient of vegetation with a wide variation in structure, density and biomass, generating several types (classes) of vegetation. The types of vegetation are called physiognomies and they have been characterized by a multi-level classification system proposed by Ribeiro and Walter (2008). According to this system, there are three major groups of ecosystems (Grassland, Savanna and Forest), which can be divided into 11 physiognomies and 14 additional sub-types of physiognomies, resulting in a total of 25 physiognomic types.

The Ribeiro and Walter classification system (RIBEIRO; WALTER, 2008) may be arranged in a hierarchical structure, where the identification of more generalist classes (Grassland, Savanna and Forest) in the first level of classification facilitates the identification of more detailed physiognomies in the next level. The Shrub Grassland, for example, is a Grassland physiognomy usually misclassified as Shrub Savanna, which is a Savanna physiognomy (FERREIRA et al., 2007; JACON et al., 2017). This confusion occurs mainly due to the difficulty in identifying the boundaries of each physiognomy in transition areas. Thus, a successful separation of Grassland and Savanna in a first level of classification would reduce the physiognomies misclassification in

the next hierarchical level. Furthermore, the patterns of the physiognomies in the Ribeiro and Walter (2008) classification system were based on *in situ* observations of vegetation structure and environmental aspects. Thus, their identification using RS images relies on the existence of a proper environmental (spatial) context. Consequently, classification methodologies based only on local (pixel) spectral features may be inefficient to discriminate the Cerrado physiognomies.

In RS applications, among several techniques that take into account the spatial context information in the image classification process, two have been widely used: Geographic Object-Based Image Analysis (GEOBIA) (BLASCHKE, 2010) and Deep Learning (LECUN et al., 2015). Compared to a per-pixel analysis, GEOBIA allows the extraction of other features (e.g., spatial, textural and contextual) for each segment or object in the image to perform the classification. However, most studies based on GEOBIA to discriminate the Cerrado vegetation have not used the necessary level of detail to properly identify the physiognomies and have not applied a hierarchical approach (GRECCHI et al., 2013; OROZCO-FILHO, 2017).

In recent years, Deep Learning methods based on Convolutional Neural Networks (CNNs) have thrived in some RS applications (MA et al., 2019). CNNs are able to perform end-to-end classification, learning from an input dataset and extracting features whose complexity increases in each layer of the network (LECUN et al., 2015). CNN layers perform convolution operations on the image during training, which enable to take into account a rich spatial context information in the classification process (MA et al., 2019). The results achieved with such methods often outperform those obtained with traditional Machine Learning algorithms, such as Random Forest and Support Vector Machine (SVM) (GUIRADO et al., 2017; KUSSUL et al., 2017).

Each physiognomy has a unique biodiversity and is responsible for a specific amount of carbon stocked above and below the ground (RIBEIRO; WALTER, 2008; RIBEIRO et al., 2011). For this reason, improving the detailed Cerrado physiognomies mapping is crucial, since so far the mapping initiatives handled well the identification of the three major ecosystem groups, but reached only low accuracies for individual physiognomies (JACON et al., 2017; GIROLAMO-NETO, 2018). Additionally, the spectral behavior of the physiognomies and their respective major ecosystem groups rely on information contained in different wavelengths, represented by satellite spectral bands. The great variety of spectral bands is an important source of information in RS field to analyze and discriminate targets. In Deep Learning methods, originally developed in the computer vision field, the analysis of the contribution of differ-

ent spectral bands to improve the network accuracy is not yet well explored. The majority of studies use only Red, Green and Blue channels (GUIRADO et al., 2017; KATTENBORN et al., 2019), some of them include the Near InfraRed one (NOGUEIRA et al., 2016; JOZDANI et al., 2019). Additionally, very few initiatives in Deep Learning approaches have applied some hierarchical behavior in classification tasks (GUO et al., 2018; YANG et al., 2020).

Therefore, this work is based on the following **hypotheses**:

- The use of Deep Learning techniques enables the classification of Cerrado physiognomies with higher level of detail and better performance than traditional techniques of Machine Learning and GEOBIA;
- The adoption of a hierarchical classification enables that less complex tasks (i.e., differentiation of Forest, Savanna and Grassland) in a first level reduces the confusion between more detailed physiognomies in the subsequent levels;
- The use of several different combinations of spectral bands as input dataset in the Deep Learning network provides the assessment of the appropriate spectral information for the goal task.

1.1 Objectives

The general objective of this work is to develop and evaluate a novel method based on Deep Learning techniques to hierarchically classify the detailed physiognomies of the Brazilian Savanna (Cerrado), according to the classification system proposed by Ribeiro and Walter (2008). As specific objectives, this work aims to:

- Compare the hierarchical classification results achieved using GEOBIA and Deep Learning techniques;
- Compare the performance of the classification according to different approaches used to generate the training samples for the Deep Learning method;
- Evaluate different combinations of spectral bands taken as input dataset in the Deep Learning network;
- Generate detailed physiognomies mapping of an environmental Protected Area from Cerrado using Deep Learning.

1.2 Document organization

In order to achieve the proposed objectives, this work is structured as follows:

- Chapter 2: presents a literature review about the importance of the Cerrado vegetation and the main concepts of GEOBIA and Deep Learning. It also describes and analyses several works related to this thesis;
- Chapter 3: presents the description of the study site, the data used, the GEOBIA approach tested and explains the proposed hierarchical methodology based on Deep Learning techniques;
- Chapter 4: presents the results of the pixel-wise classification of Cerrado physiognomies achieved using GEOBIA and Deep Learning, and compare them. It also presents the performance of the Deep Learning networks when different combinations of spectral bands were used as input datasets;
- Chapter 5: assess the results presented in Chapter 4 and compare them to other related works;
- Chapter 6: explains the concluding remarks of the work and presents suggestions for future works.

2 RELATED WORK

In this chapter, initially the Cerrado vegetation is characterized, emphasizing the classification system proposed by Ribeiro and Walter (2008). Thereafter, main concepts related to this thesis regarding optical RS image classification are reviewed. Finally, several studies that classified the Cerrado vegetation employing RS techniques are analyzed.

2.1 Brazilian Savanna (Cerrado) and its vegetation cover

There are basically three well known classification systems available to describe the Cerrado vegetation. The first one was proposed by the RADAMBRASIL Project (VELOSO; GÓES-FILHO, 1982) and the second is the Technical Handbook of Brazilian Vegetation from the Brazilian Institute of Geography and Statistics (IBGE) (IBGE, 1992; IBGE, 2012), both representing the vegetation of the entire country. Actually, IBGE's classification system is a recasting of the system proposed by RADAMBRASIL. Among the proposed classes, IBGE's system does not present an unique class for Gallery Forest, which is commonly included in other classes. Savanna vegetation classes (e.g., *Savana Arborizada*, *Savana Parque* and *Savana Gramínio-Lenhosa*) may or may not contain Gallery Forests in their subdivisions. Following this system, the class *Savana Gramínio-Lenhosa com Mata de Galeria*, for instance, would include patches of Grassland and Forest in the same class.

The third system, which represents specifically the Cerrado vegetation, was proposed by Ribeiro and Walter (2008). Since it was created specifically to characterize the Cerrado vegetation, it is more detailed than the general systems previously presented and presents several aspects to characterize each vegetation type. The differentiation among the physiognomies in this system is based on shape (structure, types of dominant growth and possible seasonal changes), aspects of the environment (edaphic factors) and floristic composition. This system has been widely used in analysis based on RS imagery (FERREIRA et al., 2007; SCHWIEDER et al., 2016; JACON et al., 2017; RIBEIRO et al., 2020) and divides the physiognomies in three major ecosystem groups (also called *formations*): *Grassland*, *Savanna* and *Forest*.

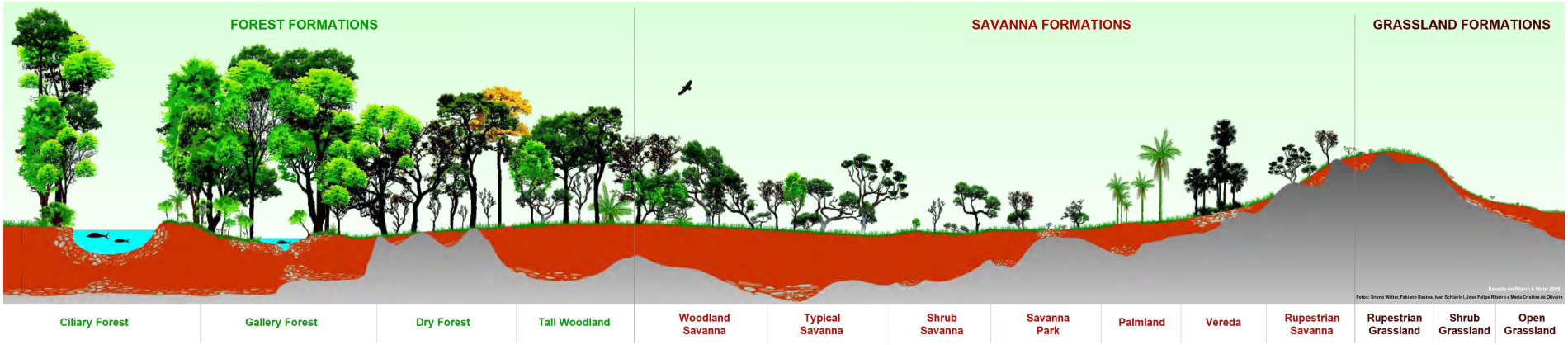
In this work, the types of vegetation will be described according to the classification system proposed by Ribeiro and Walter (2008). The Forest formations present the predominance of arboreal species, forming continuous or discontinuous canopy. There are four physiognomies included in this group: Ciliary Forest, Gallery Forest, Dry Forest and Tall Woodland. The Ciliary and Gallery Forests are both riparian

forests (e.g., occur adjacent to water bodies). They can be differentiated according to the water bodies that they follow, the deciduousness and the floristic composition. In terms of floristic composition, the Ciliary Forest is more similar to the Dry Forest, however this last one does not occur adjacent to water bodies. The fourth Forest physiognomy, Tall Woodland, is a lower forest, that presents several species of the *Sensu Stricto* Savanna.

In Savanna formations, the presence of continuous canopy is uncommon and there are trees and shrubs scattered over grasses. There are also four physiognomies in Savanna formations: *Sensu Stricto* Savanna, Savanna Park, Palmland and *Vereda*. The *Sensu Stricto* Savanna is comprised by four very relevant subtypes: Woodland Savanna, Typical Savanna, Shrub Savanna and Rupestrian Savanna. Woodland, Typical and Shrub Savanna are basically differentiated by their vegetation density, while Rupestrian Savanna typically occurs in shallow soils with rocky outcrops. The Savanna Park concentrates the vegetation in little terrain elevations, known as *murundus* or *monchões*. The other two physiognomies, Palmland and *Vereda*, are both mainly composed by palmtrees. However, the *Vereda* presents only one species (*Mauritua flexuosa*, known as Buriti) and in a lower density.

In Grassland formations, there are predominantly herbaceous species and some shrubs. Three physiognomies belongs to this group: Shrub Grassland, Open Grassland and Rupestrian Grassland. In the Shrub Grassland, the presence of shrubs is very abundant, while in the Open Grassland, the vegetation is mostly herbaceous. The Rupestrian Grassland is differentiated by the occurrence in areas with rocky outcrops. Therefore, it is possible to specify 11 physiognomies from the three ecosystem groups. Considering their subdivisions, the classification system has 25 physiognomies. These types of vegetation are illustrated in Figure 2.1, and their detailed description is presented in Table 2.1.

Figure 2.1 - The physiognomies according to the Ribeiro and Walter classification system, represented in a biomass gradient (growing from right to left).



SOURCE: Adapted from Ribeiro and Walter (2008).

Table 2.1 - Detailed description of the Cerrado physiognomies.

Formation	Physiognomy	Vegetation	Tree cover	Tree height	Soil type	Other aspects	Subdivisions
Forest	Ciliary Forest *(Mata Ciliar)*	Riparian forest, closed canopy, composed predominantly by semi-deciduous trees	From 50% to 90%	20 - 25m	Shallow (Cambisol, Plinthosol or Litholic Neosol) or deep (Latosol or Argisol) or Fluvic Neosol	Follows the medium and large size rivers, vegetation does not form galleries (closed corridors)	-
	Gallery Forest *(Mata de Galeria)*	Riparian forest, closed canopy, composed predominantly by evergreen (deciduous) trees	From 70% to 95%	20 - 30m	Cambisol, Plinthosol, Argisols, Gleysol, Neosol or Latosol (from adjacent areas)	Follows the small size rivers and streams, vegetation forms galleries over water bodies	Floodable and Non Floodable
	Dry Forest *(Mata Seca)*	Composed by various levels of deciduousness, which determine three subdivisions	From 70% to 95% (wet season) or less than 50% (dry season)	15 - 25m	high fertility (*Terra Roxa Estruturada*, *Brunizém* or Cambisol), Latosol or Calcareous Soil (in rocky outcrops)	Not associated with water bodies, occurs in interfluves and its subdivisions are also influentiated by floristic composition and soil type	Deciduous, Semi-deciduous and Evergreen

(To be continued)

Table 2.1 - Detailed description of the physiognomies (Continued).

Formation	Physiognomy		Vegetation	Tree cover	Tree height	Soil type	Other aspects	Subdivisions
	Tall Woodland *Cerradão*		Presents sclerophyll characteristics (rigid leaves and thick cuticle). Presence of shrub layer	From 50% to 90%	8 - 15m	Deep and well drained soils (Red and Red-yellow Latosols)	Physiognomically, it is a forest, but it is floristically more similar to a Savanna	Dystrophic (poor soil) and Mesotrophic (richer soil)
Savanna	*Sensu Stricto**	Woodland Savanna (*Cerrado Denso*)	Mostly arboreal, presence of shrub and herbaceous layer (less dense due to the tree shades)	From 50% to 70%	5 - 8m	Predominantly Red and Red-yellow Latosol or Cambisol	Presence of tortuous trees with twisted branches. Usually presents evidence of forest fires	-
		Typical Savanna (*Cerrado Típico*)	Mostly arboreal-shrubby, without continuous canopy. Intermediate Savanna type between Woodland and Shrub Savanna	From 20% to 50%	3 - 6m	Predominantly Red and Red-yellow Latosol or Cambisol	Presence of tortuous trees with twisted branches. Usually presents evidence of forest fires	-
		Shrub Savanna (*Cerrado Ralo*)	Mostly arboreal-shrubby. Presents a significant shrub-herbaceous layer	From 5% to 20%	2 - 3m	Predominantly Red and Red-yellow Latosol or Cambisol	Represents the lowest and least dense of the *Sensu Stricto* Savannas	-

(To be continued)

Table 2.1 - Detailed description of the physiognomies (Continued).

Formation	Physiognomy	Vegetation	Tree cover	Tree height	Soil type	Other aspects	Subdivisions
	Rupestrian Savanna (*Cerrado Rupestre*)	Arboreal-shrubby, presents shrub-herbaceous layer and occurs in rupestrian environments	From 5% to 20%	2 - 4m	Litolic Neossols	Presence of rocky outcrops. Usually occurs in mosaics, which includes other types of vegetation	-
	Savanna Park *(Parque de Cerrado)*	Trees grouped in small elevations of the terrain (*murundus* or *monchões*)	From 5% to 20%	3 - 6m	Gleysol	Soils are better drained in *murundus* than in adjacent depressions	-
	Palmland *(Palmeiral)*	Majority of a certain species of arboreal palmtree	Depend on the predominant species			Occurs in well-drained soils	*Babaçual*, *Buritizal*, *Guerobal* and *Macaubal*
	Vereda	Presents only one species of palmtree, the Buriti (*Mauritua flexuosa*), in lower density than in Palmland	From 5% to 10%	12 - 15m	Gleysol	Does not form a closed canopy and is usually surrounded by shrub-herbaceous vegetation	-

(To be continued)

Table 2.1 - Detailed description of the physiognomies (Continued).

Formation	Physiognomy	Vegetation	Tree cover	Tree height	Soil type	Other aspects	Subdivisions
Grassland	Shrub Grassland *(Campo Sujo)*	Shrub-herbaceous. Evident presence of shrubs, sub-shrubs and isolated low trees	No canopy formation	-	Shallow soils (Litolic Neosol, Cambisol and Plinthosol) or deep and low-fertility soils (medium texture Latosol or Quartzenic Neosol)	Subdivisions occur according to topographic and edaphic factors	Dry (deep water table), Humid (high water table) and with *Murundus*
	Open Grassland *(Campo Limpo)*	Mostly herbaceous, the presence of shrubs and sub-shrubs are insignificant and there are no trees	No canopy formation	-	Neosol and Plinthosol	Subdivisions occur according to topographic and edaphic factors	Dry (deep water table), Humid (high water table) and with *Murundus*
	Rupestrian Grassland *(Campo Rupestre)*	Shrub-herbaceous, possible presence of low trees	No canopy formation	up to 2m	Litolic Neosol	Usually occurs at altitudes above 900m and in areas with rocky outcrops	-

*Characterized according to its subdivisions.

SOURCE: Adapted from Ribeiro and Walter (2008).

2.2 Image classification

In RS field, image classification consists in extracting information from digital images in order to label the pixels or an object in an image as representatives of specific types of land cover or other classes (SCHOWENGERDT, 2006). The information may be related to several properties (spectral, spatial, temporal, geometric, statistical and so on) (CHEN et al., 2018). Except for Nogueira et al. (2016) and Neves et al. (2020), all works (described in Section 2.3) proposed methods to automatically map the Cerrado vegetation using traditional techniques with GEOBIA approaches or Machine Learning algorithms. *Traditional techniques*, in this study, refers to methods and algorithms related to GEOBIA and Machine Learning (e.g., Maximum Likelihood, SVM and Random Forest), which are already established and conventionally being used in RS applications over the past decades.

Both GEOBIA and Machine Learning techniques require features related to spectral, textural and geometrical characteristics, which are extracted from the images and used to classify the objects of interest. These features (also known as *shallow features*) are often created by mathematical formulations, such as the ratio between spectral bands of the sensor (i.e., vegetation indices), which transforms the input data into proper representations of it (ZHANG et al., 2016). In order to create features to correctly represent the classes in the image, the traditional methods require a high level knowledge about specific aspects of the classes. Even so, in some cases they are not adequate enough to represent all existing information in the input image (BALL et al., 2017). Despite the high classification accuracy presented in some studies, a detailed Cerrado vegetation mapping, considering the Ribeiro and Walter (2008) classification system, has not yet been achieved. Under this scenario, this Section provides a brief description of the image processing and pattern recognition techniques used in this work.

2.2.1 Geographic Object-Based Image Analysis (GEOBIA)

Based on established concepts (e.g., segmentation and classification), the GEOBIA created a relationship between them and, consequently, allowed the migration of analyzes based only on spectral information of pixels to contextual analyzes based on objects and their spectral, temporal and spatial characteristics (BLASCHKE, 2010). One of the main purposes of GEOBIA is to develop theories, methods and tools that are capable of reproducing or improving (automatically or semi-automatically) the interpretation performed by specialists (HAY; CASTILLA, 2008).

Many scientific works have demonstrated that classification approaches based on objects have achieved better results than those per pixel in some RS applications (BLASCHKE, 2010; WHITESIDE et al., 2011). This occurs especially in high spatial resolution images (JOHANSEN et al., 2010), where the variability of spectral information for each target is greater than in medium or low spatial resolution images. A classic GEOBIA framework mainly includes four steps: segmentation, feature extraction, classification and validation (CHEN et al., 2018; BLASCHKE, 2010). The segmentation step consists in partitioning the image into meaningful non-overlapping segments or objects based on some homogeneity criteria. For each object, several features are extracted according to its characteristics. In the classification step, relationships between these features are used to label each segment/object. Finally, in the validation step, the resulting classification is evaluated based on some reference data.

2.2.1.1 Features for vegetation patterns

In the feature extraction step, some techniques are commonly used to create features to represent the vegetation patterns, including the Cerrado vegetation. Some of them are the following:

• **Vegetation Indices (VI):** They are utilized to highlight vegetation targets, reducing the data dimensionality without losing spectral information. VI features can be obtained from the combination of two or more spectral bands, which is possible due to the distinct vegetation spectral behavior especially in the visible and near-infrared ranges. In the visible range, the low reflectance values are related to the absorption of electromagnetic radiation by the photosynthetic pigments (e.g., chlorophyll), while in the near infrared range, the high reflectance values occur due to the spread of radiation inside the leaves caused by their cellular structure (PONZONI; SHIMABUKURO, 2010).

These indices may be related to several biophysical variables of vegetation, such as biomass, Leaf Area Index (LAI) and evapotranspiration. They allow the seasonal and long-term monitoring of structural, phenological and biophysical parameters of vegetation (HUETE et al., 2002). However, they should be carefully used, since modifications in the image acquisition conditions (e.g., satellite viewing geometry) may affect the indices without any change in the vegetation (ADAMI, 2010).

• **Linear Spectral Mixture Model (LSMM):** The radiation captured in the Instantaneous Field of View (IFOV) of a sensor is the combination of radiation emitted and reflected by various targets (e.g., leaves, branches, soil and water) on the

Earth's surface. The mixture of radiation is related to the sensor spatial resolution; a higher spatial resolution and, consequently, a smaller IFOV, result in less mixture of elements at the same pixel (SHIMABUKURO; SMITH, 1991).

The spectral mixture models are used to estimate the contribution of each component, also called endmembers, within a pixel (ROBERTS et al., 1998). They generate image fractions from the "pure components" spectra. Several models have been proposed to estimate these contributions and the most common used endmembers are vegetation, soil and shade (or water). The LSMM assumes that components of a pixel have a linear relationship (SHIMABUKURO; SMITH, 1991). The resulting image fractions are highly dependent on the quality of the chosen "pure pixels", and the number of endmembers can not be higher that the number of input bands used in the analysis.

• **Tasseled Cap (TC) transformation:** The TC transformation is a linear transformation employed to reduce the data volume and complexity, highlighting some characteristics of the images (KAUTH; THOMAS, 1976). The sensor data may be represented in a multidimensional space, in which each dimension is related to a spectral band. In a sensor with three bands, for example, the space would be a cube (CRIST; KAUTH, 1986).

Every type of target (e.g., vegetation and soil) has a specific spectral behavior, since their individuals share physical properties. Therefore, sensor data do not have equal probabilities of occurrence in a multidimensional space. The data mainly concentrates in certain portions of this multidimensional space, creating what is called "data structure". If each band was correlated to only one type of target and uncorrelated to others, modifications in that target would change only that band. The TC transformation identifies the data structure for a certain sensor and application and, by rotating the band axes, defines new directions (axes) associated exclusively with the physical characteristics of a scene target. These new axes will be the components resulting from the TC transformation (CRIST; KAUTH, 1986).

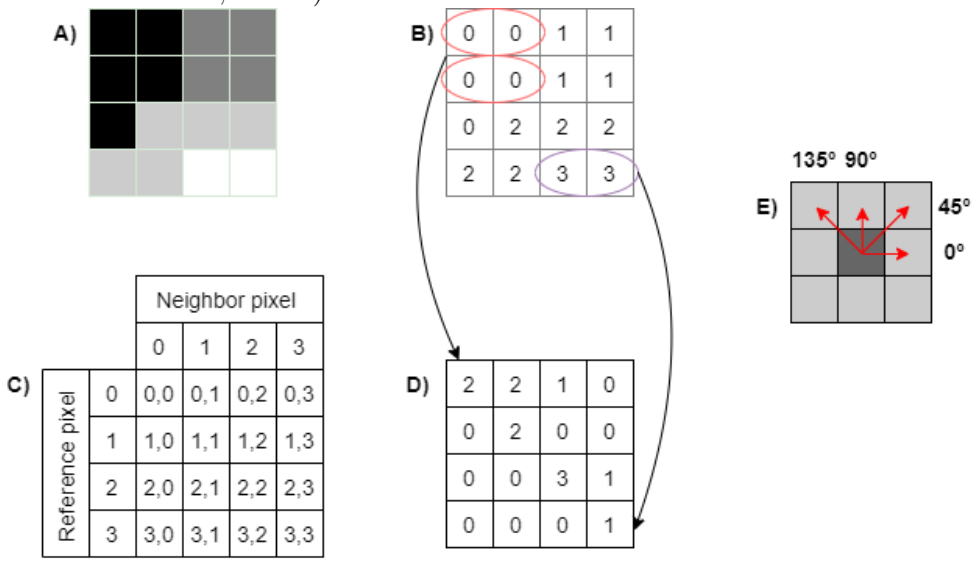
For each sensor, a different transformation suitable for its spectral bands must be proposed. Three commonly components created for several sensors are the following: greenness, brightness and wetness. The greenness component is related to the green vegetation reflectance, while the brightness component is defined in the direction of soil reflectance variation and may be related to uncovered or partially covered soils and variations in topography. Finally, the wetness component corresponds to the direction observed by the variation of soil moisture, and may also be related to

water (CRIST; KAUTH, 1986; BAIG et al., 2014; YARBROUGH et al., 2014).

• **Texture - Grey Level Co-occurrence Matrix (GLCM):** The texture is related to the differences between the grey levels that occur in an image, their spatial correlation and distribution. Small differences generate a smooth texture, while great differences are related to a rough texture. To represent the image texture and its features, Haralick et al. (1973) created the Grey Level Co-occurrence Matrix (GLCM). It is a tabulation of all different combinations of pixel brightness (grey levels) that occur in an image or region of image.

Given an image (Figure 2.2A) and its respective grey levels (Figure 2.2B), the matrix structure (Figure 2.2C) is created based on the possible combinations between the reference pixels and their neighbors. The GLCM (Figure 2.2D) is filled according to how many times each combination occurs. To choose which pixel will be considered as neighbor, a direction (Figure 2.2E) must be considered. In this example, the 0° direction and a counterclockwise rotation were considered (HALL-BEYER, 2017).

Figure 2.2 - Construction of a GLCM. A) Image; B) Grey levels; C) Matrix structure; D) GLCM; and E) Directions.



SOURCE: Adapted from Hall-Beyer (2017).

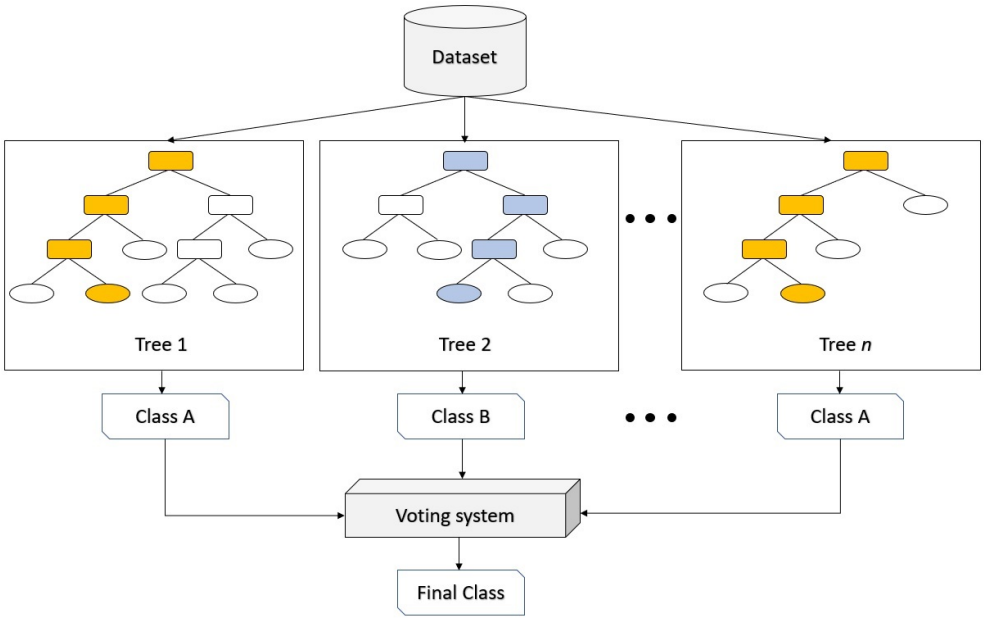
The GLCM is a square matrix and has the same number of lines and columns that the image quantization level. A image with 256 grey levels, for instance, will create a GLCM with 256 lines, 256 columns and 65.536 cells. Based on the GLCM,

texture features of three groups are created: contrast, order and descriptive statistics. The contrast group is composed by three features: contrast, dissimilarity and homogeneity. In the order group, there are the entropy and the angular second moment. Finally, the descriptive statistics group is composed by mean, variance and correlation (HALL-BEYER, 2017).

2.2.1.2 Random Forest algorithm

Usually, in GEOBIA, the relationships and hierarchies between objects are adaptable and can be modified according to the needs and interpretation of specialists (HAY; CASTILLA, 2008). The classification step allows the use of several techniques, such as empirical decision trees created based on the specialist knowledge and Machine Learning algorithms. Among them, the Random Forest algorithm (BREIMAN, 2001), represented in Figure 2.3, has achieved higher accuracy when compared to other traditional Machine Learning algorithms in several RS applications, especially when using GEOBIA approaches (MA et al., 2017).

Figure 2.3 - Illustration of a Random Forest algorithm, where a set of decision trees creates a voting system to choose the class.



SOURCE: Adapted from Oshiro (2013).

This algorithm is an ensemble method, i.e., composed by a set of classifiers and, in general, is more robust to deal with noisy data and outliers. The Random Forest

algorithm creates a set of decision trees based on subsets of the training data. The subsets are generated by a statistical technique called bootstrap, that performs a random sampling of the training data with replacement. Each decision tree is responsible for a vote and the most popular class will be assigned to the sample (an object, in GEOBIA) (BREIMAN, 2001). The performance of the classification using Random Forest is highly dependent on the input features, also known as shallow features.

2.2.2 Deep Learning

Machine Learning is a research field within Artificial Intelligence, which seeks to extract patterns from raw data to acquire some knowledge (GOODFELLOW et al., 2016). In the last years, a Machine Learning subarea has thrived and allowed pattern representations to be learned through a process called Deep Learning. In Deep Learning, multiple levels of the data representation are built hierarchically: the first level detects edges, corners and grey levels, the next ones detect patterns and then the following ones learn the association of these patterns. The complexity increases in each level (LECUN et al., 2015).

In the last decades, two aspects were fundamental to allow the recent success of Deep Learning techniques: the great increase in the capacity of data generation, which are used as samples; and the improvement of computational infrastructure, both in software and in hardware (GOODFELLOW et al., 2016). Deep Learning has reached better results than traditional Machine Learning techniques in many research fields, such as speech recognition (SAINATH et al., 2013), object detection (GIRSHICK et al., 2015) and medical images analysis (LITJENS et al., 2017). In RS field, the use of Deep Learning also became a trend in the development of digital image processing techniques, especially in three categories: (1) image classification, whereby entire image patches are labeled with one class (NOGUEIRA et al., 2016); (2) object recognition, whereby objects of interest in an image are located mostly using bounding boxes (GUIRADO et al., 2017); and (3) semantic segmentation or pixel-wise classification, whereby each pixel of the image is labeled (NEVES et al., 2020; ADARME et al., 2020). In supervised methods, the most used Deep Learning algorithm is the CNN (LECUN et al., 2015).

2.2.2.1 Convolutional Neural Network (CNN)

The CNNs are composed by multiple hierarchical stages for feature extraction. Each stage is commonly composed by three parts: (1) Convolutional layers, (2) Activa-

tion function and (3) Pooling layer. After these stages, there are one or more Fully-Connected layers and, in the end, a classifier (ZHANG et al., 2016). It is possible to perform a pixel-wise classification (also known as semantic segmentation in Deep Learning applications) using CNNs. However, the samples for a supervised classification in this case would not be only pixels. CNNs receive images with defined size as samples and, in this way, the existing context among the pixels is still considered in the classification.

The most common parts of a CNN are the following:

a) Convolutional layer: It is composed by a set of filters that will be applied to the input image. The filter (or kernel) consists in a $k \times k$ matrix of weights w_i, whose values are learned by the network using the backpropagation algorithm. The $k \times k$ matrix computes the filter in a neighborhood called local receptive field (Figure 2.4), where the output value is a combination of the input pixel values. After applying the filter to the entire image, a feature map is generated. The stacked feature maps resulting from a convolutional layer are the input data for the next layer. The number of convolutional layers depends on the CNN architecture being used (PONTI et al., 2017).

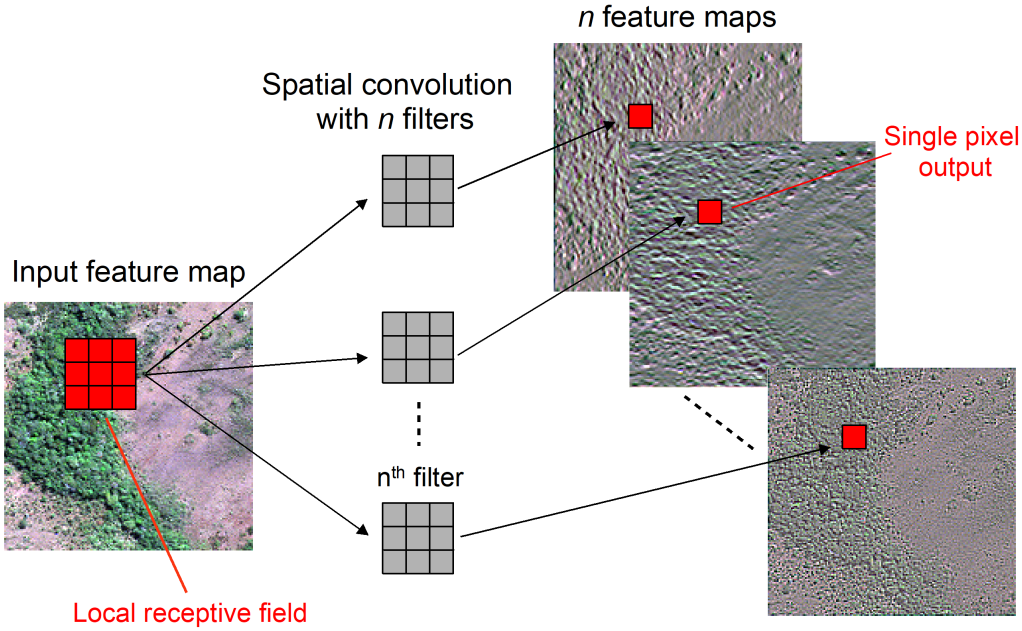
Each filter has a unique set of weights, which is shared by the local receptive fields of the image. This aspect reduces the number of parameters that should be learned and, consequently, the network training time (LECUN et al., 2015). For each convolutional layer, it is necessary to define the number of filters and their respective size, stride and padding. The stride represents the interval of pixels that will be considered when applying the filters. The padding occurs when the filters consider additional edges in the input image and it is optional (PONTI et al., 2017).

b) Activation Function: After a convolutional layer, an activation function is usually applied. The most common in CNNs is the Rectified Linear Unit (ReLU - Figure 2.5a):

$$f(x) = max(0, x) \tag{2.1}$$

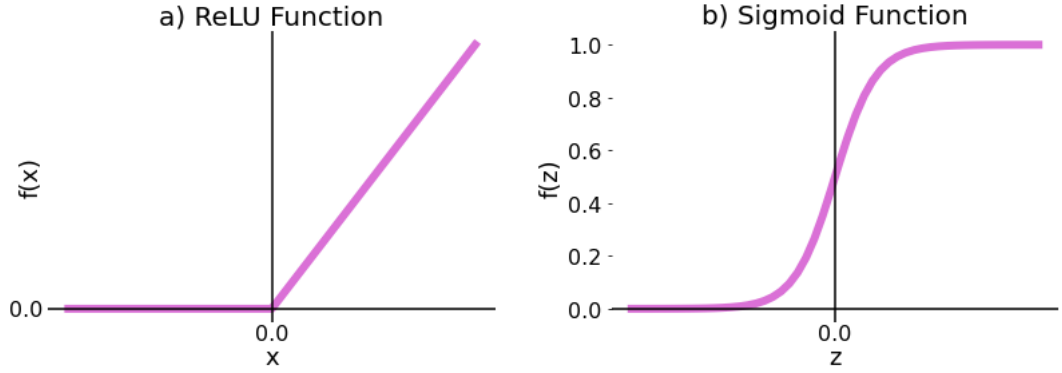
where x is the feature map pixel value.

Figure 2.4 - Illustration of a convolutional process, where the values from the local receptive field are used as input in the filters to produce single pixels output in the feature maps.



SOURCE: Adapted from Ponti et al. (2017).

Figure 2.5 - Representations of: a) Rectified Linear Unit function; b) Sigmoid function.

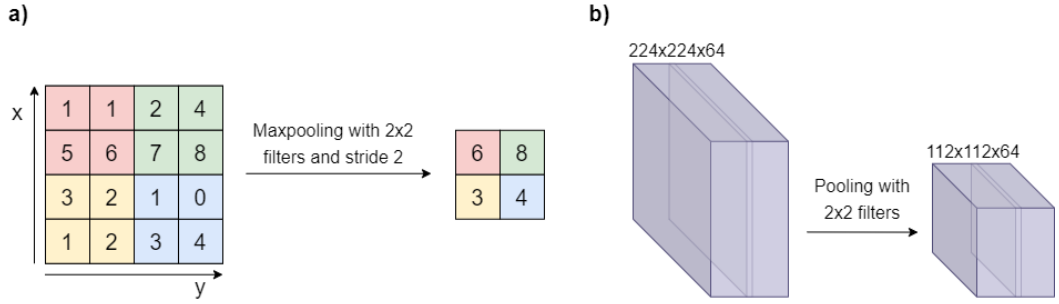


SOURCE: Author's production.

In ReLU function, all negative values are reset to zero and all positive values are maintained, without losing spatial resolution (GOODFELLOW et al., 2016). It reduces the data dimensionality and selects which neurons remain active for the next layer (LECUN et al., 2015).

c) Pooling layer: A Pooling layer is usually applied in order to reduce the data dimensionality. Frequently, the operator used is the maxpooling, which selects the maximum value of the chosen window. The application of maxpooling with window size 2×2 and stride 2 is illustrated in Figure 2.6. The resulting images is 2 times smaller, but the volume depth is preserved, when existing. This process downsamples the input image and preserves the higher frequencies.

Figure 2.6 - Pooling layer. a) Applying maxpooling operator; and b) Downsampling the image, but preserving the volume depth.



SOURCE: Adapted from Li (2018).

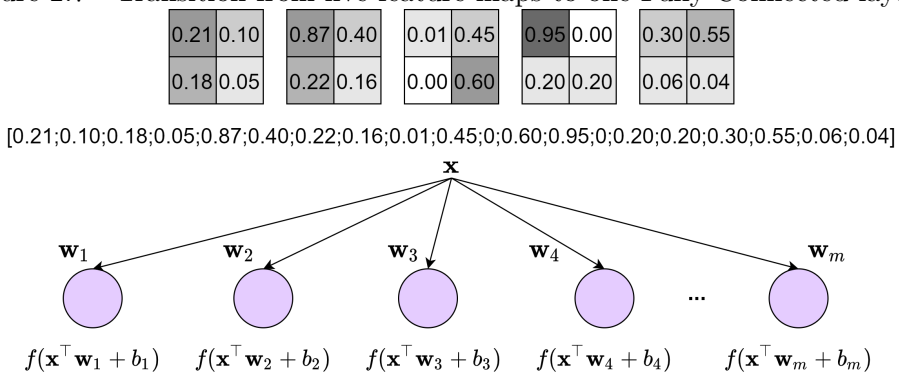
d) Fully-Connected layer: It is similar to the hidden layers of Multilayer Perceptron (MLP), an Artificial Neural Network algorithm. Each neuron from the Fully-Connected layer is connected to all layers from the previous layer. In Figure 2.7, a transition from five feature maps (2×2) to one Fully-Connected layer with m neurons is presented. Each neuron produces an output based on $f(x^\top w + b)$, where x is a vector created based on the feature maps, w represents the weights associated to each neuron from the previous layer and b is the bias.

e) Classification: Usually, after the Fully-Connected layer, the last part of a CNN is the one that computes probabilities (from 0 to 1) of belonging to each class for each element (pixel). For each pixel, the class that achieved the highest probability is assigned. In order to convert predicted values into probabilities, a Sigmoid function may be used:

$$f(z) = \frac{1}{1 + e^{-z}} \tag{2.2}$$

where $f(z)$ is the estimated probability between 0 and 1, e is the natural logarithm

Figure 2.7 - Transition from five feature maps to one Fully-Connected layer.



SOURCE: Adapted from Ponti et al. (2017).

base and z is the input for the function (i.e., the previous value predicted by $f(x^\top w + b)$).

2.2.2.2 U-Net

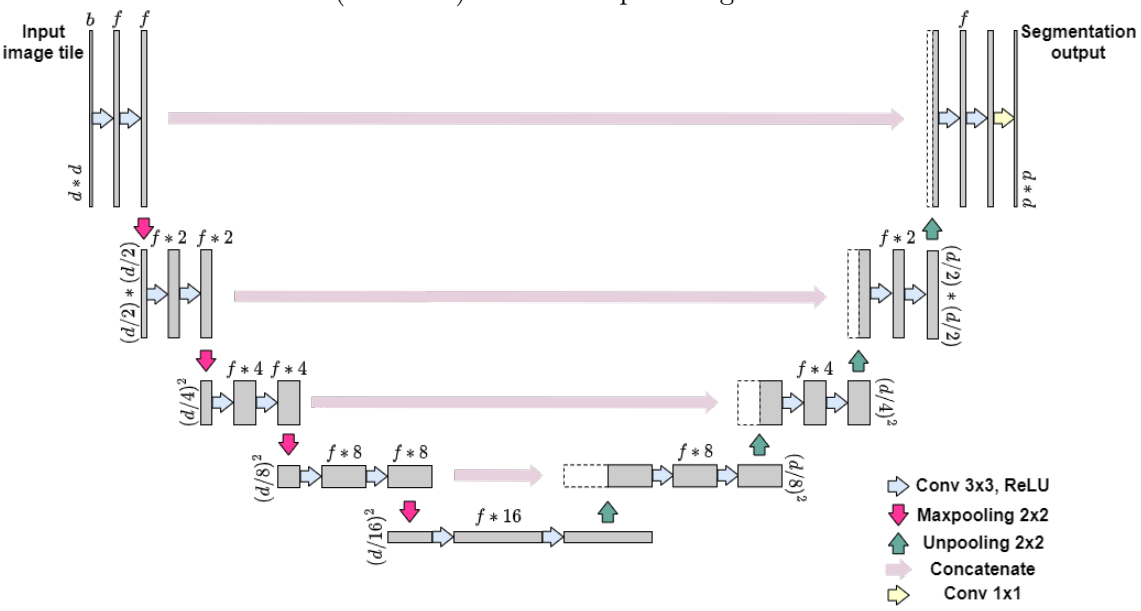
Traditional CNNs, such as LeNet (LECUN et al., 1990) and AlexNet (KRIZHEVSKY et al., 2012), predict a single class for each image. As a subgroup of CNNs, Fully Convolutional Neural Networks (FCNNs - Long et al. (2015)) are tailored to the task of semantic segmentation. In particular, they take an image patch with an arbitrary number of channels as input and predict a map usually of the same size as the input or a little smaller, with a label class assigned for each pixel. The U-net (RONNEBERGER et al., 2015) belongs to the group of FCNNs and has achieved great results especially for vegetation targets (FLOOD et al., 2019; WAGNER et al., 2019; NEVES et al., 2020).

U-net is composed of two usually symmetrical parts, which creates a U-shaped architecture (Figure 2.8). The first part (left side of Figure 2.8) consists in a multi-layer convolutional encoder that successively reduces the spatial resolution and increases the number of filters per layer. This contracting path is able to capture the context associated to the studied classes. The second part (right part of Figure 2.8) is a multi-layer convolutional decoder, that up-scales the features to the original spatial resolution through unpooling layers in order to preserve the information localization. They use skip-connections between contracting and expansive paths in order to concatenate the precise localization of the information (from contracting path) and the high level features (from the expanding path). These concatenations enables

the accurate prediction of object boundaries (RONNEBERGER et al., 2015).

In the U-net contracting path, convolutions of 3×3 followed by ReLU layers and 2×2 maxpooling layers with stride two are applied. In each of these downsampling steps, the number of feature layers (and filters, consequently) is doubled. In the expansive path, each step corresponds to an upsample of the feature map (unpooling layer), a concatenation with the corresponding feature maps from contracting path and two 3×3 convolutions followed by ReLUs. At the end, a 1×1 convolution is used to map the feature layers into the desired number of classes (RONNEBERGER et al., 2015). The unpooling layer is also known as a 2×2 transposed convolution. It is not exactly a deconvolution layer, since the operation is not the mathematical inverse of the convolution applied in the contracting path. The transposed convolution applies some padding in the input image before the computation of 2×2 normal convolutions, reconstructing the desired spatial resolution.

Figure 2.8 - Representation of a U-net. d is the image size, f is the number of filters and b is the band (or bands) from the input image tile.



SOURCE: Adapted from Ronneberger et al. (2015).

2.3 Initiatives of Cerrado vegetation mapping

The study and characterization of the Cerrado native vegetation in the field are essential to assess its ecological functions. However, the *in situ* analysis is generally restricted to small areas and also requires more time and financial and human resources. For this reason, RS techniques have been used as an efficient tool to study the Cerrado vegetation, enabling analysis of large areas and places that are hard to reach (INPE, 2015; ALENCAR et al., 2020). These analyzes can be performed visually or using automatic classification methods and several types of RS images. The following sections present a brief review about the main projects and studies related to Cerrado vegetation mapping. They are separated based on the main techniques employed.

2.3.1 Visual interpretation

Project of Conservation and Sustainable Use of the Brazilian Biological Diversity (PROBIO) (SANO et al., 2009) and Terraclass (INPE, 2015) were the first initiatives to map the vegetation in the entire Cerrado using Landsat satellite imagery. Both projects adopted a methodology based on image segmentation and visual interpretation to map the Cerrado vegetation. PROBIO produced a map for the year 2002 based on the classification system proposed by IBGE (2012). Differently, Terraclass adopted the Ribeiro and Walter (2008) classification system to produce the map for the year 2013.

Due to the difficulty in differentiating detailed physiognomies employing 30m spatial resolution images, both projects grouped them. PROBIO obtained 71% of overall accuracy to classify three vegetation types: Forest, Savanna and Grassland (SANO et al., 2009). TerraClass Cerrado reached an overall accuracy of 80.2%, considering all of its land use and land cover classes (i.e., including classes such as Agriculture and Pasture). However, it mapped Forest and Non-Forest (comprises Savanna and Grassland) classes with an overall accuracy ranging from 60% to 65% (INPE, 2015). One possibility to reduce the confusion is using better spatial resolution images.

Pinheiro and Durigan (2009) used aerial and QuickBird images (0.65m - fused) to discriminate Humid Open Grassland, Shrub Savanna, Typical Savanna, Woodland Savanna and Tall Woodland. Their purpose was to evaluate the succession stages of the physiognomies from 1962 to 2006, therefore they did not develop a classification methodology and also did not provide accuracy measures. The authors claimed to be possible to differentiate physiognomies in QuickBird images. Despite that, map-

ping through visual interpretation is a time-consuming process and has an intrinsic subjectivity due to different interpreters who are performing the task. Nowadays, there are other methodologies to partially or totally automate the mapping process.

2.3.2 Machine Learning

MapBiomas Project has performed the automatic land use and land cover mapping for the entire Cerrado biome, using Landsat imagery (SOUZA et al., 2020; ALENCAR et al., 2020). In 2020, MapBiomas released its fifth collection, which includes a series of annual (from 1985 to 2019) maps produced by a classification method based on Random Forest algorithm to discriminate Forest, Savanna and Grassland. For each collection, MapBiomas methodology is modified to improve the resulting maps. The maintenance of consistent maps over time is fundamental for analysis that imply the application of penalties, such as the Soy Moratorium in the Amazon.

Other initiatives of semi-automatic or automatic Cerrado vegetation mapping have been performed in small study sites. Some of them included the seasonality, an important aspect of the Cerrado vegetation. In this case, high temporal but low spatial resolution (250m) imagery from MODIS (Moderate Resolution Imaging Spectroradiometer) were used (ACERBI-JUNIOR et al., 2006; BORGES; SANO, 2014; COSTA et al., 2015; ABADE et al., 2015). MODIS data is commonly used to estimate vegetation indices, such as EVI (Enhanced Vegetation Index) (BORGES; SANO, 2014; COSTA et al., 2015) and NDVI (Normalized Difference Vegetation Index) (ABADE et al., 2015).

The differentiation of the physiognomies through vegetation indexes times series can be performed using algorithms that calculate similarities between the temporal profiles (e.g., Euclidean distance and Spectral Angle Mapper - SAM) (ABADE et al., 2015; BORGES; SANO, 2014) or using classification algorithms, such as SVM and MLP (ABADE et al., 2015; COSTA et al., 2015). However, due to the low spatial resolution, MODIS data generate mixture of classes in the same pixel and hinders the proper detailing of the vegetation classes. Abade et al. (2015) grouped the physiognomies into only one class "Cerrado" to separate them from Caatinga vegetation in a frontier between both biomes. They achieved an accuracy of 80.65%. Costa et al. (2015) obtained an overall accuracy of 85.96% in the classification of three classes: Native Grasslands, Pasture (cultivated) and Others.

Differently, Borges and Sano (2014) achieved better detailing of vegetation classes, differentiating the following ones: Shrub Grassland, Shrub Savanna, Typical Savanna, Woodland Savanna, Seasonal Semideciduous Forest, Seasonal Deciduous For-

est (Dry Forest), Secondary Vegetation, Agriculture and Cultivated Pasture. However, they used images from 2000 to 2011 (276 images) and considered only the predominant pattern of each pixel. Areas converted to other land covers in this period were classified as the predominant class, which overrates the natural vegetation area. The validation was performed using only 90 points in a RapidEye orthoimage with 5m of spatial resolution and the overall accuracy obtained was 84%. Considering that, in the validation process, only 1 and 5 points were selected for Shrub Grassland and Shrub Savanna classes, respectively, their patterns may not have been well represented in the classification.

Acerbi-Junior et al. (2006) had the purpose of creating an approach with the Landsat spatial resolution and the MODIS temporal resolution to map Cerrado vegetation. For that, they combined MODIS and Landsat data in a hybrid image through fusion based on Wavelet transform. However, the highest accuracy in the classification of Forest, Savanna and Grassland was 87.69% with only Landsat images, using the Maximum Likelihood estimation. Using the hybrid image to classify the same classes, the accuracy was 82.71%. Thus, despite presenting a temporal resolution favorable to the analysis of the Cerrado vegetation seasonality, the spatial resolution of MODIS images is not suitable to differentiate the vegetation patterns.

Having a better spatial resolution, Landsat imagery have been widely used to map Cerrado vegetation, even in general applications that do not require the individualization of physiognomies (MÜLLER et al., 2015; GRECCHI et al., 2013). Müller et al. (2015) employed NDVI time series metrics (from 2009 to 2012) and Random Forest algorithm to distinguish Forest and Savanna. They achieved an overall accuracy of 92.19%. Despite they obtained high classification accuracy, their methodology was not designed to distinguish detailed physiognomies. Similarly to Borges and Sano (2014), Müller et al. (2015) used time series in the classificaton process but did not identify land cover conversions in the analyzed period, which may have caused misclassifications.

This same problem was observed in Schwieder et al. (2016), who classified the following physiognomies based on the Ribeiro and Walter (2008) classification system: Open Grassland, Shrubby Grassland, Shrubby Savanna (*Campo Cerrado*), Wooded Savanna (*Cerrado Sensu Stricto*), Savanna Woodland (*Cerrado Denso*), Dense Cerrado Woodland (*Cerradão*) and Gallery Forests. They used phenological metrics from 2000 to 2014 calculated based on TC components, which were computed using Landsat images. The overall accuracy with SVM algorithm was 63%. Despite

the long period analyzed, the lack of data in some months may have affected the classification and generated artefacts in the final map.

Additionally, Schwieder et al. (2016) observed considerable confusions in transition areas between the Open Grassland and Shrubby Grassland and also between Shrubby Grassland and Shrubby Savanna. To overcome this problem and increase the accuracy, the authors considered the classification errors between adjacent classes as correct classifications. For example, if there were 290 correct validation points in a total of 451 (64% of accuracy) for Shrubby Grassland, in the new approach the correct validation points in adjacent classes were added and the total correct points turned to 433 (Open Grassland + Shrubby Grassland + Shrubby Savanna) in 451, achieving an accuracy of 96%. Despite the significant increase, the new accuracies do not truly represent the method's ability to differentiate the physiognomies.

Misclassifications in transition areas were also observed by Ferreira et al. (2007) and Jacon et al. (2017). The first one classified 5 physiognomies (Cerrado Grassland - *Campo Limpo*, Shrub Cerrado - *Campo Sujo*, Wooded Cerrado - *Cerrado Ralo*, Cerrado Woodland - *Cerrado Típico* and Gallery Forest) using Mahalanobis Distance and two types of LSMM created from Landsat images: a general one, including the soil, vegetation and water fractions, and a specific one, with Grassland, Savanna and Forest fractions. Despite the accuracy of 89% using the general LSMM, confusions in transition areas occurred. Shrub Cerrado (21%) was classified as Wooded Cerrado and Wooded Cerrado (18%) was classified as Shrub Cerrado. The spectral similarity (and consequent confusion) between the physiognomies is higher in transition areas, considering the vegetation cover gradient presented in Figure 2.1, therefore their patterns are not linearly separable (FERREIRA et al., 2007). Besides, images with 30m spatial resolution are not appropriate to deal with the spectral similarity between two or more physiognomies in transition areas.

Using 6 hyperspectral images from Hyperion/EO-1 (30m spatial resolution) acquired in different dates, Jacon et al. (2017) used three types of metrics (reflectance, vegetation indices and absorption parameters) and Multiple Discriminant Analysis (MDA) to differentiate seven Cerrado physiognomies: Gallery Forest, Dense Woodland Savannah (*Cerrado Denso*), Woodland Savannah (*Cerrado Típico*), Open Woodland Savannah (*Cerrado Ralo*), Wooded Savannah (*Campo Cerrado*), Shrub Savannah (*Campo Sujo*) and Savannah Grasslands (*Campo Limpo*). The overall accuracy was higher in the dry season than in the wet season: 84.2% and 81.4%, respectively. During the dry season, the spectral separability of the physiognomies increases.

However, even using hyperspectral imagery, confusions were observed between Savannah Grassland and Shrub Savannah, and between Woodland Savannah and Open Woodland Savannah in transition areas.

Using a high spatial resolution image from WorldView-2 (2 meters), Girolamo-Neto (2018) tested different levels of legend. In the first level (Forest, Savanna and Grassland), they obtained accuracy of 88.95%. However, the accuracy dropped to 68.60% in the detailed level, which classified 10 physiognomies: Gallery Forest, *Vereda*, Woodland Savanna, Typical Savanna, Shrub Savanna, Shrub Grassland, Open Grassland, Rupestrian Grassland, Humid Open Grassland and Humid Open Grassland with *murundus*. Girolamo-Neto (2018) showed the potential of texture features to separate some classes (e.g., Open Grassland and Shrubby Grassland).

2.3.3 GEOBIA

Using Landsat image for the years 1985, 1995 and 2005, Grecchi et al. (2013) adopted a methodology based on GEOBIA to identify Cerrado natural vegetation, a general class. They used the Nearest Neighbor (NN) algorithm, which classifies the objects based on a similarity measure. The reflectance, the Crop Enhanced Index (CEI) and the terrain slope were used as features and the overall accuracy was of 86%. With a better spatial resolution, Silva and Sano (2016) and Orozco-Filho (2017) used RapidEye images to classify Cerrado vegetation types.

Silva and Sano (2016) performed image segmentation and classification based on Euclidean distance to discriminate Forest, Savanna and Grassland classes. They obtained accuracy values of 91% (Forest), 90% (Savanna) and 71% (Grassland). However, in each of the three classified images, only one class was predominant. Thus, the classification was between Forest and Non-Forest in one scene, Savanna and Non-Savanna in the other and Grassland and Non-Grassland in the last. The confusion between different classes of vegetation barely occurred and might have not interfered in the classification. Orozco-Filho (2017) obtained an accuracy of 86% with decision trees and an object-based approach. To accomplish that, he grouped some classes (Humid Open Grassland and *Veredas*) and did not subdivided any Forest formation.

Altought Ribeiro and Walter (2008) proposes a hierarchical structure, which simplifies the classification of Cerrado vegetation types, just few works have explored this hierarchical aspect (NEVES et al., 2019; RIBEIRO et al., 2020). Neves et al. (2019) used GEOBIA techniques and Random Forest algorithm to compare a hierarchical and

a non-hierarchical approach to differentiate seven Cerrado physiognomies. Despite showing the superiority of the hierarchical approach over the non-hierarchical one, the accuracy for some physiognomies were still low. More recently, Ribeiro et al. (2020) also used GEOBIA and Random Forest algorithm and furthermore included a spatial contextual ruleset to represent other environmental factors (e.g., soil type, slope and elevation) to improve the identification of Cerrado vegetation classes. They classified 13 classes (11 of which were vegetation types) and achieved an overall accuracy of 87.6%. However, their semi-automatic methodology still relies on some subjective tasks, such as the determination of segmentation image parameters.

2.3.4 Deep Learning

To improve the detailing of Cerrado vegetation mapping, a large variety of Machine Learning techniques such as Random Forest and Decision Trees have been employed. For Savanna vegetation, some efforts have already been made using Deep Learning to delineate tree crowns (BRANDT et al., 2020; TORRES et al., 2020). However, Nogueira et al. (2016) were the first to employ a Deep Learning-based method to identify Cerrado vegetation patterns, which include several tree heights, tree cover and shrub and herbaceous vegetation.

Using RapidEye imagery, Nogueira et al. (2016) classified entire regular image patches as only one class (Forest, Savanna or Grassland), resulting in a considerable mixture of classes in a single patch. A semantic segmentation (i.e., the assignment of a separate class per pixel, also known as pixel-wise classification) of the three major ecosystem groups was performed by Neves et al. (2020), using a modified U-net architecture and 8 WorldView-2 spectral bands. Compared to the classification approach performed by Nogueira et al. (2016), the semantic segmentation proposed by Neves et al. (2020) resulted in a much better delineation of each class.

Table 2.2 presents a summary about the projects and studies on Cerrado vegetation mapping analyzed in Section 2.3.

Table 2.2 - Summary of initiatives for Cerrado vegetation mapping analyzed in Section 2.3.

Reference	Image	Classes*	Method	Accuracy	Contribution	Limitation
Abade et al. (2015)	MODIS (250m)	Cerrado (*Stricto Sensu* and Grassland)	Euclidean distance, SAM and SVM	80.75%	Classification in transition areas with Caatinga biome	Low spatial resolution
Acerbi-Junior et al. (2006)	MODIS and Landsat (fusion - 30m)	Forest, Savanna and Grassland	Maximum likelihood	87.59%	Integrated Landsat spatial resolution and MODIS temporal resolution	Highest accuracy achieved with Landsat only
Alencar et al. (2020) (MapBiomas)	Landsat (30m)	Forest, Savanna and Grassland	Decision trees and Random Forest	71%	Three decades of automatic mapping of the entire biome	Confusions between Grassland and Pasture and between Savanna and Forest
Borges and Sano (2014)	MODIS (250m)	6 physiognomies	SAM	84%	Used time series to represent seasonality	Did not detect class conversions during the period (11 years). Few validation points

(To be continued)

Table 2.2 - Summary of initiatives for Cerrado vegetation mapping analyzed in Section 2.3 (Continued).

Reference	Image	Classes*	Method	Accuracy	Contribution	Limitation
Costa et al. (2015)	MODIS (250m)	Grasslands, Pasture and Others	SVM, MLP and Autoencoder	85.96%	Investigated the separation of Pasture and Grasslands	Reference data for Pasture and Grasslands are from different years (2006 and 2009, respectively)
Ferreira et al. (2007)	Landsat (30m)	5 physiognomies	Mahalanobis distance	89%	Used LSMM fractions	Confusion between Shrub Cerrado and Wooded Cerrado
Girolamo-Neto (2018)	WorldView-2 (2m)	10 physiognomies	Random Forest	68.52%	Used image texture	Low accuracy for detailed level
Grecchi et al. (2013)	Landsat (30m)	Cerrado natural vegetation	NN	86%	Considerated slope data	No detailing of natural vegetation types
INPE (2015) (TerraClass Cerrado)	Landsat8/OLI (30m)	Forest and Non-Forest	Segmentation and visual interpretation	Between 60% and 65%	Entire biome mapping	Grouped Grassland and Savanna. Low accuracy

(To be continued)

Table 2.2 - Summary of initiatives for Cerrado vegetation mapping analyzed in Section 2.3 (Continued).

Reference	Image	Classes*	Method	Accuracy	Contribution	Limitation
Jacon et al. (2017)	Hyperion/EO-1 (30m)	7 physiognomies	MDA	84.2%	Differentiated vegetation patterns in dry and wet season	Misclassification in transition areas
Müller et al. (2015)	Landsat (30m)	Forest and Savanna	Random Forest	92.19%	Used time series of 30m spatial resolution	Did not detect class conversions during the period
Neves et al. (2019)	WorldView-2 (2m)	7 physiognomies	Random Forest	68.95%	Used a hierarchical approach	Low accuracy
Neves et al. (2020)	WorldView-2 (2m)	Forest, Savanna and Grassland	Deep Learning	87%	Used Deep Learning approach and performed pixel-wise classification	No physiognomies detailing
Nogueira et al. (2016)	RapidEye (5m)	Forest, Savanna and Grassland	Deep Learning	90.54%	Used Deep Learning approach	Classified entire image patches. Poor delineation
Orozco-Filho (2017)	RapidEye (5m)	5 physiognomies	Decision trees (J48)	86.1%	Used object-based approach	Grouped some physiognomies

(To be continued)

Table 2.2 - Summary of initiatives for Cerrado vegetation mapping analyzed in Section 2.3 (Continued).

Reference	Image	Classes*	Method	Accuracy	Contribution	Limitation
Pinheiro and Durigan (2009)	Aerial photos and QuickBird (0.6m)	5 physiognomies	Visual Interpretation	-	High Spatial resolution	Did not perform validation
Ribeiro et al. (2020)	RapidEye (5m)	11 physiognomies	GEOBIA and Random Forest	87.6%	Included a spatial context ruleset	Relies on subjective steps
Sano et al. (2009) (PROBIO)	Landsat7/ETM+ (30m)	Forest, Savanna and Grassland	Segmentation and visual interpretation	71%	Entire biome mapping	Confusions between Grassland and Pasture and between Savanna and Forest
Schwieder et al. (2016)	Landsat (30m)	7 physiognomies	SVM	63%	Used time series of 30m spatial resolution	Lack of data for some months and confusion in transition areas
Silva and Sano (2016)	RapidEye (5m)	Forest, Savanna and Grassland	Segmentation and Euclidean distance	Between 71% and 91%	High Spatial resolution	Predominance of one class in each scene

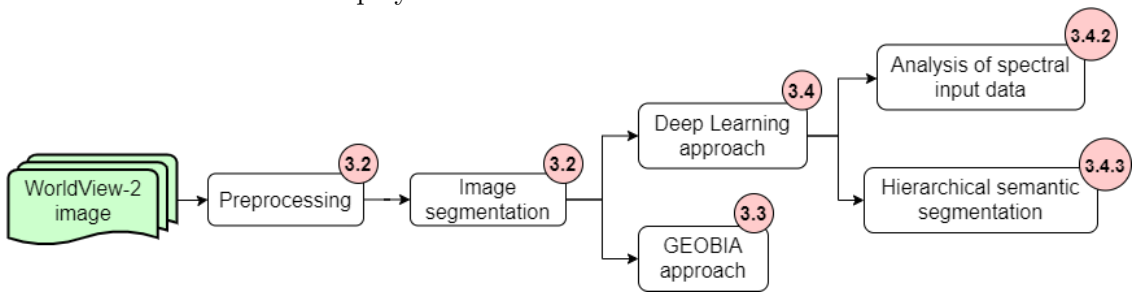
* Considering only the classes of vegetation.

SOURCE: Author's production.

3 METHODOLOGY

This research was developed using two different classification approaches: GEOBIA (Section 3.3) and Deep Learning (Section 3.4). Figure 3.1 shows a simplified methodology flowchart, where WorldView-2 image is preprocessed and segmented. The resulting image and the segmentation are used by both GEOBIA and Deep Learning approaches. This last approach is divided into two parts: analysis of spectral input data and hierarchical semantic segmentation. Each step will be further explained in the following sections.

Figure 3.1 - Simplified methodology flowchart. Inside the circles, the respective section numbers are displayed.



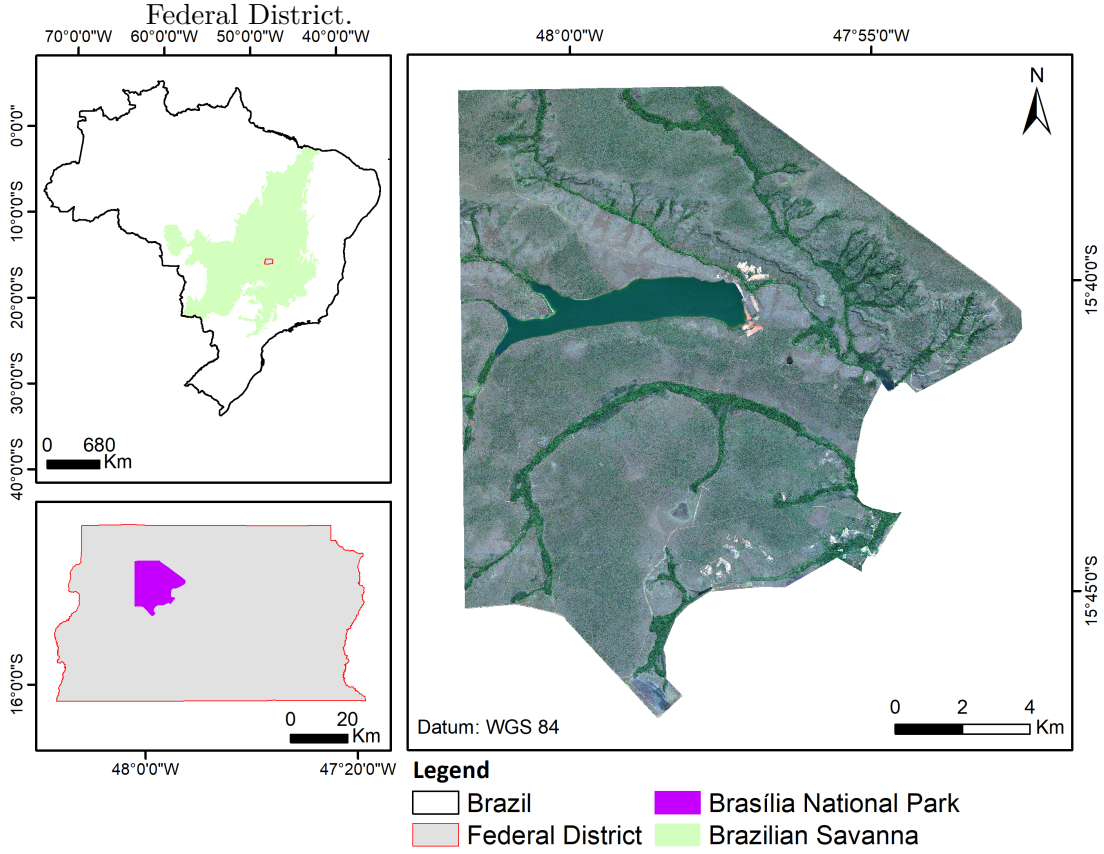
SOURCE: Author's production.

3.1 Study site

As study site, a Brazilian Protected Area was chosen to ensure the analysis of native Cerrado vegetation. The Brasília National Park (BNP) (Figure 3.2), located in the Federal District, Brazil, comprises approximately 423 km^2 of preserved Cerrado vegetation. It is a representative area of Cerrado native vegetation, since it encompasses the major physiognomies found in Cerrado biome (FERREIRA et al., 2001; FERREIRA et al., 2007) and presents high ecological relevance, containing several endangered species (ICMBIO, 1998) and a dam that is responsible for 25% of the Federal District's water supply. This Protected Area was also used as study site in several other works (FERREIRA et al., 2001; FERREIRA et al., 2007; SCHWIEDER et al., 2016; GIROLAMO-NETO, 2018), which attests its representativeness and facilitates the comparison of results.

According to the existing physiognomies that occur in the study site and the classification system proposed by Ribeiro & Walter (RIBEIRO; WALTER, 2008), we dif-

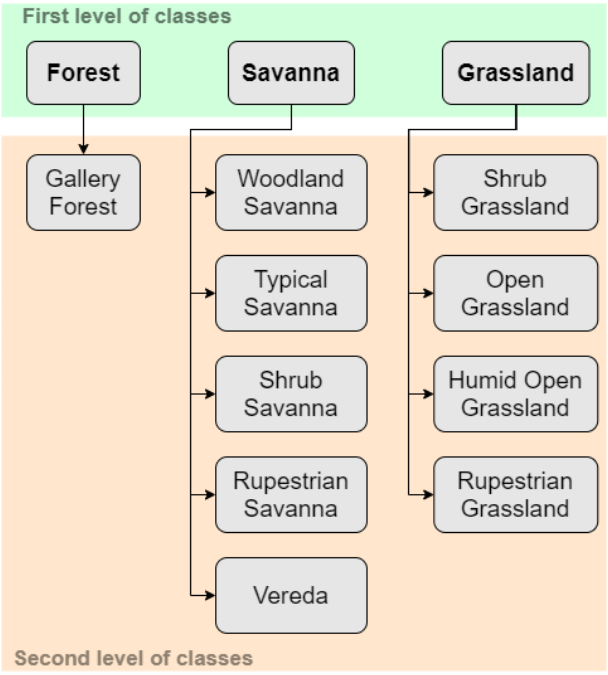
Figure 3.2 - Location of the Brasília National Park image (true color composite), in the Federal District.



SOURCE: Author's production.

ferentiated two hierarchical levels of classes (Figure 3.3). In the first level, the three major ecosystem groups (also known as formations) were classified: Forest, Savanna and Grassland. The Forest formations present the predominance of arboreal species, forming continuous or discontinuous canopy. In the second level, the Forest was maintained as Gallery Forest (*Mata de Galeria*), since it is the only Forest physiognomy with significant presence in this area. In Savanna formations, the presence of continuous canopy is uncommon and there are trees and shrubs scattered over grasses. The areas identified as Savannas in the first level were subdivided into Woodland Savanna (*Cerrado Denso*), Typical Savanna (*Cerrado Típico*), Rupestrian Savanna (*Cerrado Rupestre*), Shrub Savanna (*Cerrado Ralo*) and *Vereda* in the second level. Woodland, Typical, Rupestrian and Shrub Savanna are commonly defined as subtypes of a class named Savanna *Sensu Stricto* (*Cerrado Sensu Stricto*), but such a definition was not used in this work.

Figure 3.3 - Class hierarchical levels, based on Ribeiro and Walter (2008) and Table 2.1.



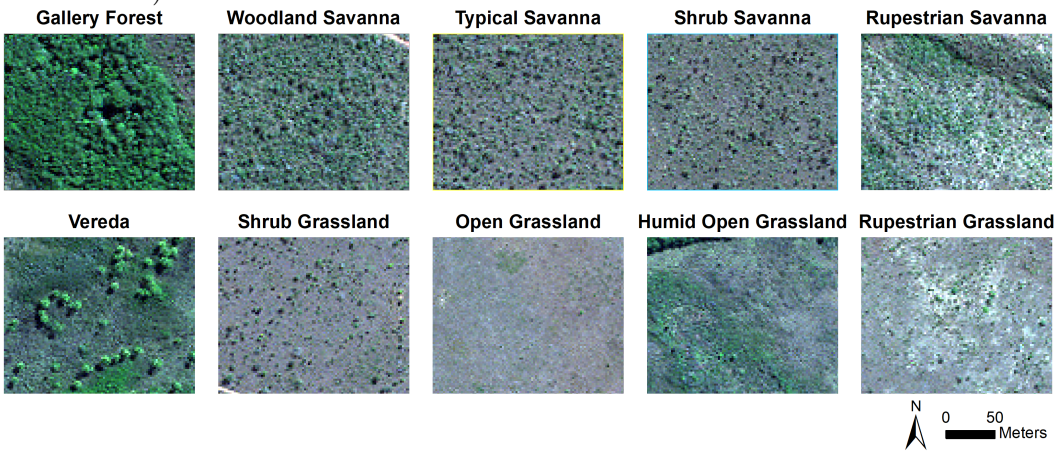
SOURCE: Author's production.

In the Grasslands, there are predominantly herbaceous species and some shrubs. Four sub-classes were differentiated in the second level: Shrub Grassland (*Campo Sujo*), Open Grassland (*Campo Limpo*), Rupestrian Grassland (*Campo Rupestre*) and Humid Open Grassland (*Campo Limpo Úmido*). The Humid Open Grassland is a subtype of Open Grassland, but it was considered an independent class due to its significant presence in the study site. The individual definitions and characteristics of the physiognomies are presented in Table 2.1. Their patterns in a true color composite of the WorldView-2 image can be seen in Figure 3.4.

3.2 Remote sensing data, preprocessing and segmentation

The high spatial resolution (2 meters) image used in this work is a WorldView-2 multispectral image (tile ID: 103001003373A600; processing level: standard 2A) acquired in July 22, 2014. With 16-bit radiometric resolution, this image has 8 multispectral bands, whose wavelength ranges are presented in Table 3.1 (DIGITALGLOBE, 2021). Initially represented in Digital Numbers (DN), the image was converted to surface reflectance using the Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) algorithm (PERKINS et al., 2005) available in the ENVI 5.2 software (EXELIS, 2013).

Figure 3.4 - Patterns of the physiognomies in the WorldView-2 image (true color composite).



SOURCE: Author's production.

Table 3.1 - Wavelength ranges of the WorldView-2 multispectral bands.

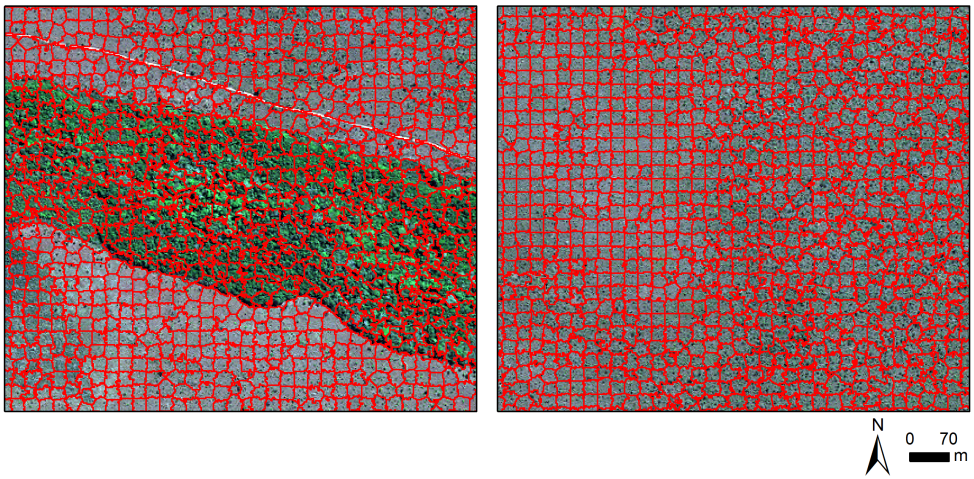
Spectral band	**Wavelength (nm)**
Coastal	400-450
Blue	450-510
Green	510-580
Yellow	585-625
Red	630-690
Red-Edge	705-745
Near Infrared 1	770-895
Near Infrared 2	860-1040

SOURCE: DigitalGlobe (2010).

The WorldView-2 image was partitioned into meaningful segments (objects), called Superpixels (ÇIğLA; ALATAN, 2010). The algorithm used was the Simple Linear Iterative Clustering (SLIC), implemented in the Scikit-Learn Python package (PEDREGOSA et al., 2011). This is an adaptation of the k-means algorithm (MACQUEEN et al., 1967), which computes the weighted distance measure through a combination of color (in the CIELAB color space) and spatial proximity. As input for SLIC, we use Red, Green and Blue bands, since the CIELAB color space is defined by the lightness (color brightness). It is also possible to control the compactness of the Superpixels. If its value is large, spatial proximity is more relevant, therefore

Superpixels are more compact (close to a square in shape). However, when the compactness value is small, they adhere more to image boundaries and have less regular size and shape (ACHANTA et al., 2012). Chosen empirically, in this study, compactness equals to 400 was used (Figure 3.5). This value depends on the range of data value used, and it was chosen empirically to create Superpixels that adhere well to the class patterns.

Figure 3.5 - SLIC result in regions of the WorldView-2 image (true color composite) using compactness equal to 400. Image on the left is a region of Gallery Forest and image on the right is a region of Grassland. Both are on the same scale.



SOURCE: Author's production.

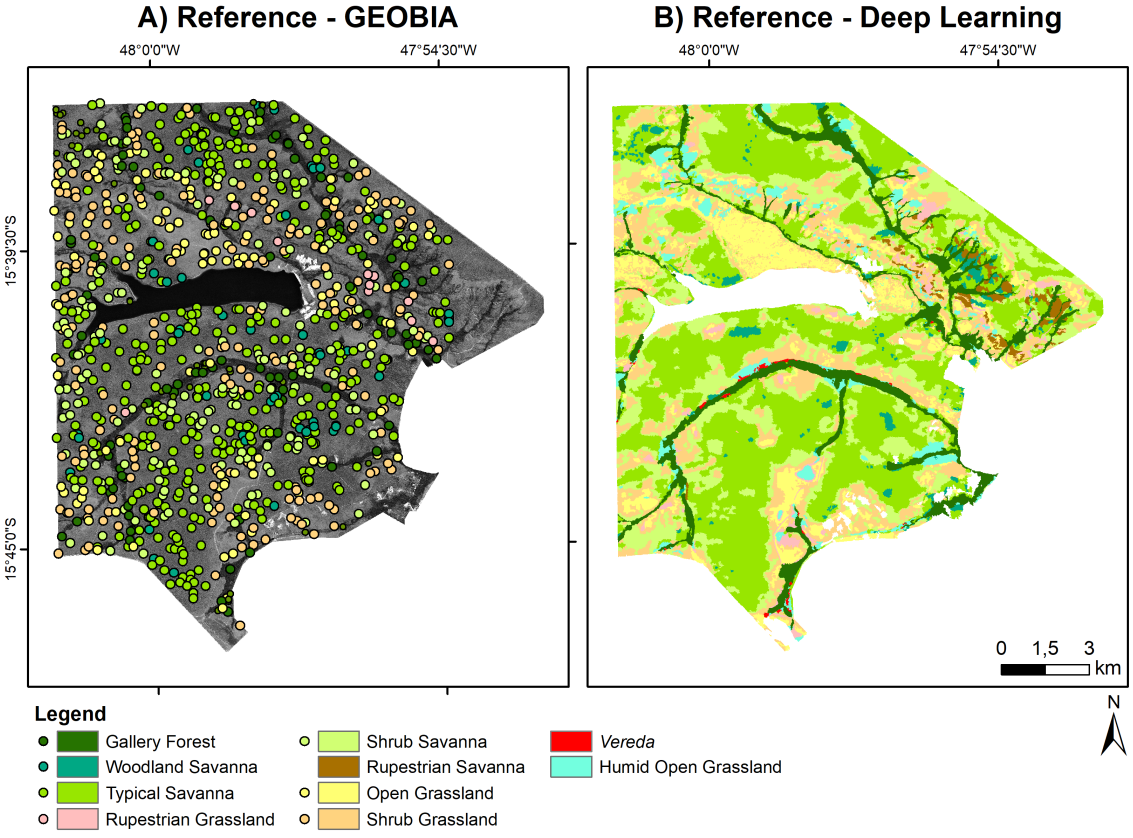
3.2.1 Reference data

A Cerrado physiognomies map produced from high spatial resolution at this level of detail is not available yet. Therefore, reference data to validate our results had to be produced. To execute the GEOBIA approach, a set of points collected by Girolamo-Neto (2018) in a fieldwork was used. The fieldwork was realized in July, 2017 (in the dry season, the same period of the WorldView-2 image acquisition), in which 141 sample points were collected. Since the target classes of this work are natural vegetation patterns in a Protected Area, it is expected that the classes in the image date (2014) are still the same during fieldwork, in 2017. However, to make sure of that, each fieldwork sample point was double-checked in the WorldView-2 image and also more 877 sample points were obtained through visual interpretation. This resulted in a total of 1018 samples.

As the FCNN requires classified patches (rather than points) as samples, the entire

WorldView-2 image was classified through visual interpretation, and then it was used as reference data. Considering our intention to differentiate types of natural vegetation, some minority areas were masked in the reference data (built-up areas, water bodies, bare soil and burned areas), then remaining ten classes of interest, described in Table 2.1 and illustrated in Figure 3.4. In Figure 3.6, both references used in this research are presented.

Figure 3.6 - Reference data used in A) GEOBIA approach; and B) Deep Learning approach. Classes without an associated point in the legend were not used in GEOBIA approach.



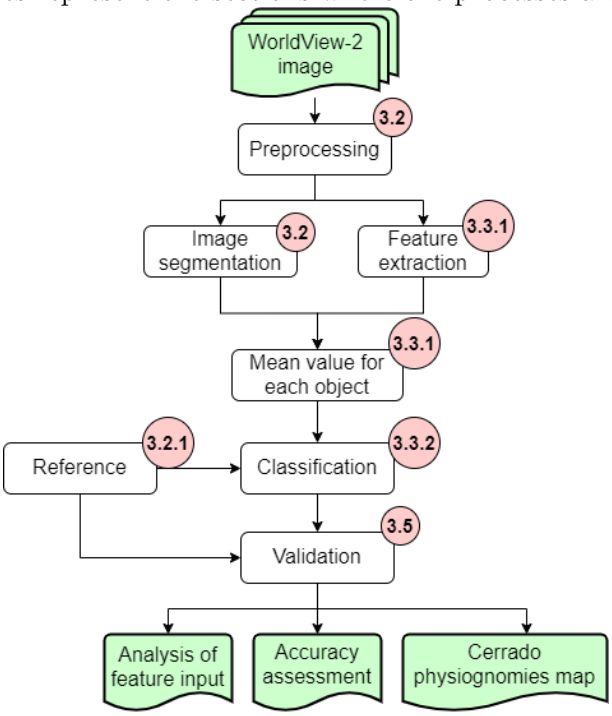
SOURCE: Author's production.

3.3 GEOBIA

GEOBIA approach was performed according to the flowchart presented in Figure 3.7. The preprocessed WorldView-2 image was taken as input in the feature extraction step. In order to integrate the features and the segmentation, the mean value for each feature was computed for each object to further perform the classification. In

this approach, three results were achieved: feature selection and analysis, accuracy assessment, and Cerrado physiognomies map. Each step will be further explained.

Figure 3.7 - Flowchart of the classification approach using GEOBIA. Numbers inside pink circles represent the sections where the processes are described.



SOURCE: Author's production.

3.3.1 Feature extraction

From WorldView-2 surface reflectance image, the following types of features were extracted: Vegetation Indices, fractions of the Linear Spectral Mixture Model, components of the Tasseled Cap transformation and texture images. In order to transform the pixel information into an object-based information, the average value of each feature was computed on each Superpixel. This process was not necessary for the texture images, since they were already extracted considering the Superpixels. More details about each type of feature will be presented in the following sections.

3.3.1.1 Vegetation Indices (VI)

These indices are commonly used to highlight vegetation targets in satellite images. They are based on the distinct spectral response that the vegetation has especially

in the regions of visible and Near Infrared (NIR) (PONZONI; SHIMABUKURO, 2010). Five VIs were extracted: Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), 2-band Enhanced Vegetation Index (EVI2), Soil-Adjusted Vegetation Index (SAVI) and Modified Soil-Adjusted Vegetation Index 2 (MSAVI2). Their respective equations and references are described in Table 3.2.

Table 3.2 - Equations and references of the five extracted Vegetation Indices.

Vegetation Index	**Equation**	**Reference**
Normalized Difference Vegetation Index (NDVI)	$\frac{\rho NIR - \rho R}{\rho NIR + \rho R}$	Rouse et al. (1974)
Enhanced Vegetation Index (EVI)	$2.5\frac{\rho NIR - \rho R}{\rho NIR + C_1 * \rho R - C_2 * \rho B + L}$	Huete et al. (1997)
2-band Enhanced Vegetation Index (EVI2)	$2.5\frac{\rho NIR - \rho R}{\rho NIR + 2.4\rho R + 1}$	Jiang et al. (2008)
Soil-Adjusted Vegetation Index (SAVI)	$\frac{\rho NIR - \rho R}{\rho NIR + \rho R + L_1} * (1 + L_1)$	Huete (1988)
Modified Soil -Adjusted Vegetation Index 2 (MSAVI2)	$\frac{2 * \rho NIR + 1 - \sqrt{(2 * \rho NIR + 1)^2 - 8 * (\rho NIR - \rho R)}}{2}$	Qi et al. (1994)

where ρNIR is the Near Infrared reflectance, ρR is the reflectance in the Red band, ρB is the reflectance in the Blue band, L is the soil line adjustment parameter (1), C_1 and C_2 are the aerossol resistence terms (6 and 7.5) and L_1 is a correction factor that depends on vegetation cover (0.5).
SOURCE: Author's production.

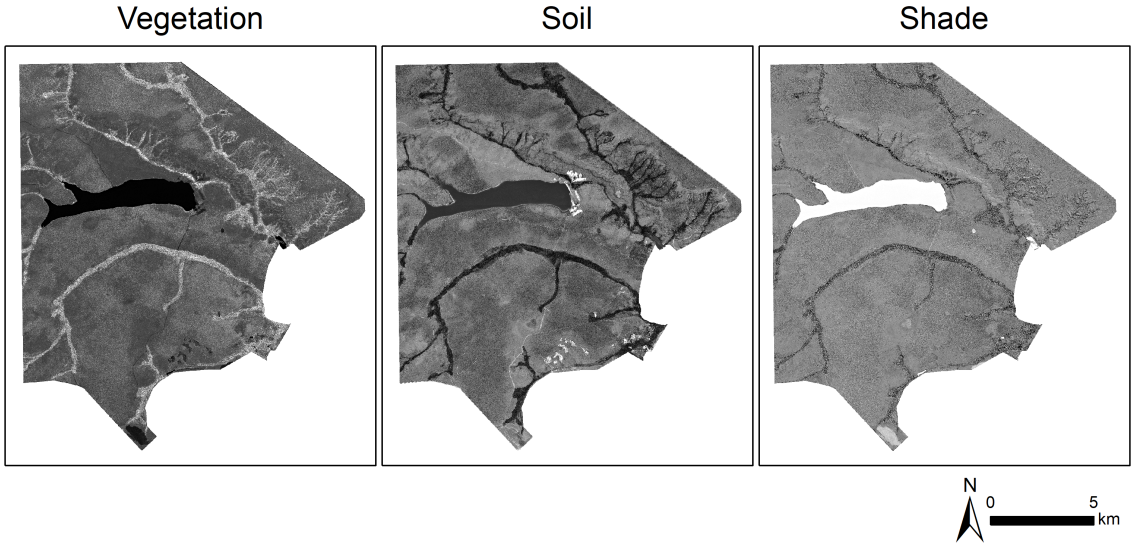
3.3.1.2 Linear Spectral Mixture Model (LSMM)

The LSMM was applied to estimate the proportions of three components in each pixel: Vegetation, Soil and Shade - Equation 3.1 (SHIMABUKURO; SMITH, 1991). To create the LSMM fraction images, ten endmembers (pure pixels) were selected for each component. The resulting feature images are presented in Figure 3.8.

$$r_i = \sum_{j=1}^{N} (a_{ij} \times x_j) + e_i \tag{3.1}$$

where r_i is the spectral reflectance mean for the i_{th} band of a pixel with N components; j is the number of components; i is the number of bands; a_{ij} is the spectral reflectance of the j_{th} component of a pixel for the i_{th} band; x_j is the proportional value of the j_{th} component of the pixel; and e_i is the error for the i_{th} band.

Figure 3.8 - Vegetation, Soil and Shade fractions resulting from the LSMM.



SOURCE: Author's production.

3.3.1.3 Tasseled Cap (TC) transformation

Through the Tasseled Cap (TC) transformation, the information from the eight bands of the WorldView-2 image was aggregated into components that are strongly related to the physical characteristics of the image. This process reduces the data dimensionality and results in three components: Greenness, Wetness and Brightness. The coefficients proposed by Yarbrough et al. (2014) were employed in the transformation.

Table 3.3 - Coeffients of the Tasseled Cap trasformation.

	WorldView-2 band							
TC component	**Coastal**	**Blue**	**Green**	**Yellow**	**Red**	**Red-Edge**	**NIR1**	**NIR2**
Brightness	-0.060436	0.012147	0.125846	0.313039	0.412175	0.482758	-0.160654	0.673510
Greenness	-0.140191	-0.206224	-0.215854	-0.314441	-0.410892	0.095786	0.600549	0.503678
Wetness	-0.270951	-0.315708	-0.317263	-0.242544	-0.256463	-0.096550	-0.742535	0.202430

SOURCE: Adapted from Yarbrough et al. (2014).

3.3.1.4 Texture features

Based on the GLCM proposed by Haralick et al. (1973) (see details in Section 2.2.1.1), six texture features were extracted for each superpixel: Contrast, Dissimilarity, Entropy, Homogeneity, Correlation and Second Angular Moment. They were calculated in the software eCognition Developer 9.01 and their equations are presented in Table 3.4.

Table 3.4 - Equations of the texture features extracted based on the GLCM.

GLCM feature	Equation
Contrast	$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$
Dissimilarity	$\sum_{i,j=0}^{N-1} P_{i,j}\lvert i-j\rvert$
Entropy	$\sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j})$
Homogeneity	$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$
Correlation	$\sum_{i,j=0}^{N-1} P_{i,j}\left(\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right)$
Second angular moment	$\sum_{i,j=0}^{N-1} P_{i,j}^2$

where P is the element in the position i, j of the GLCM with N gray levels, μ is the mean and σ^2 is the variance.
SOURCE: Adapted from Haralick et al. (1973) and Hall-Beyer (2017).

3.3.2 Classification

The classification with GEOBIA was performed using two approaches: non-hierarchical and hierarchical. The hierarchical approach mainly follows the hierarchy presented in Figure 3.3. However, the second level is composed by seven physiognomies (Gallery Forest, Woodland Savanna, Typical Savanna, Shrub Savanna, Shrub Grassland, Open Grassland and Rupestrian Grassland). In the non-hierarchical approach, the same seven physiognomies are differentiated in one step, without detecting Forest, Savanna and Grassland first. The number of samples for each physiognomy is presented in Table 3.5. The number of Savanna and Grassland samples, from the first level, corresponds to the sum of the samples of their respective physiognomies (594 for Savanna and 335 for Grassland).

Table 3.5 - Number of samples for each classified physiognomy.

Physiognomy	**Number of samples**
Gallery Forest	89
Shrub Savanna	186
Typical Savanna	362
Woodland Savanna	46
Rupestrian Grassland	12
Shrub Grassland	211
Open Grassland	112
Total	1018

SOURCE: Author's production.

Rupestrian Savanna and *Vereda* were not detected in the GEOBIA approach, since there were not enough training and validation points representing them. The Humid Open Grassland was included in the Open Grassland physiognomy. The result of the first classification level was used as input data to the next level. The misclassified superpixels at the first level were also considered as errors at the second level. In order to outline the importance of each group of features for the classification, several experiments with different combinations of features as input data were performed in the hierarchical approach (Table 3.6). Except for the first experiment, that includes all groups of features, the others use all features except one at a time. The five groups of features were represented as follows:

• **Spectral:** reflectance values from 7 spectral bands - Coastal band was not included in order to avoid the atmosphere influence in the classification (DIGITALGLOBE, 2010).

• **VegInd:** five vegetation indices presented in Section 3.3.1.1.

• **SMA:** three image fractions resulting from the spectral mixture analysis performed with the LSMM according to Section 3.3.1.2.

• **TCap:** three image components resulting from the TC transformation - Section 3.3.1.3.

• **TX:** six texture features presented in Section 3.3.1.4.

Table 3.6 - Experiments of hierarchical classification combining five groups of features.

Experiment	Composition of input data
All	All five groups of features
$\text{Except}_{\text{TCap}}$	Spectral, VegInd, SMA and TX features
$\text{Except}_{\text{SMA}}$	Spectral, VegInd, TCap and TX features
$\text{Except}_{\text{VegInd}}$	Spectral, SMA, TCap and TX features
$\text{Except}_{\text{Spectral}}$	VegInd, SMA, TCap and TX features
$\text{Except}_{\text{TX}}$	Spectral, VegInd, SMA and TCap features

SOURCE: Author's production.

The classification experiments (both non-hierarchical and hierarchical) were performed in the software Weka 3.8.4 (HALL et al., 2009) using the Random Forest algorithm (BREIMAN, 2001). According to the recommendation of Breiman (2001), the number of trees was set to 100 in all experiments. The classification accuracies were generated using a 10-fold cross validation, which means that the samples are randomly partitioned into ten groups. Nine groups are used as training set, while the remaining group corresponds to the test set that will evaluate the model. This procedure is repeated until each one of the ten groups is used as test set (i.e., ten times), and the final accuracy corresponds to the average value.

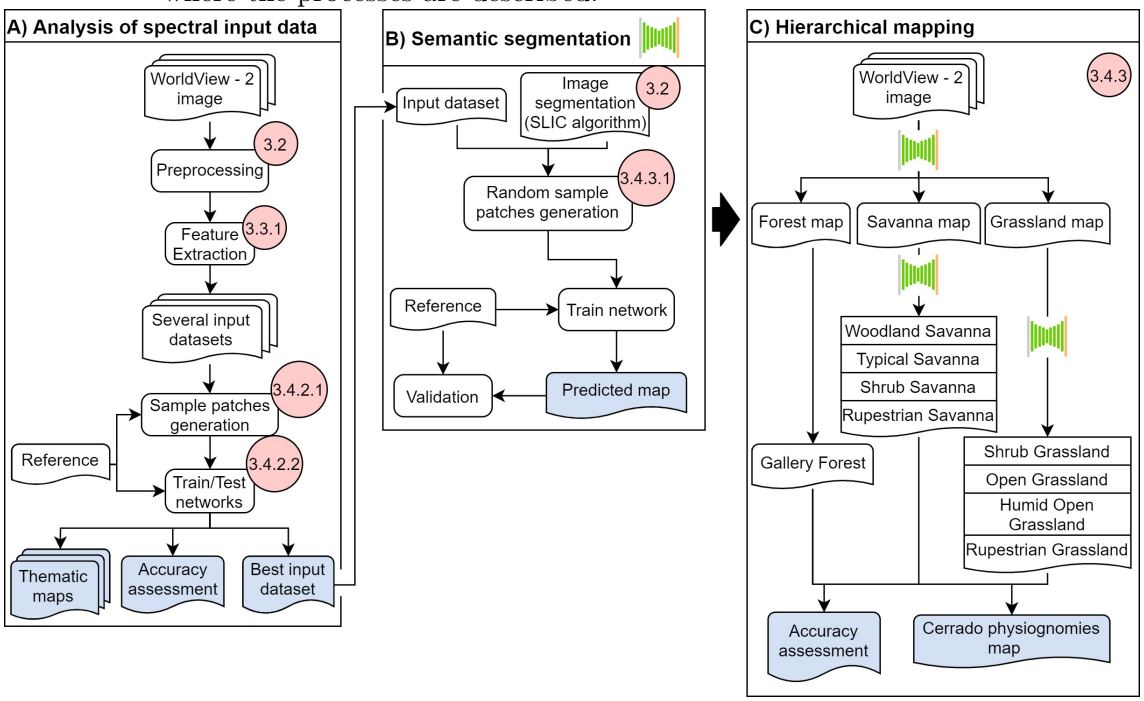
3.4 Deep Learning

The Deep Learning approach was performed according to the flowchart presented in Figure 3.9. The first part of the methodology (Figure 3.9A) corresponds to the investigation of several input datasets and results in the best combination of features to be used as input in the part B (Figure 3.9B). This part presents the steps required to perform the semantic segmentation approach. Hierarchical physiognomies mapping, using the semantic segmentation presented in Figure 3.9B, is explained in Figure 3.9C. The semantic segmentation icon is used to indicate every time it is performed. Every procedure of the flowchart is described in detail in the following sections.

3.4.1 Network Architecture

In this work, a variation of the U-net architecture (RONNEBERGER et al., 2015), proposed by Kumar (2018), was used in all tasks of pixel-wise classification (semantic segmentation). The architecture mainly follows the design-choices of Ronneberger

Figure 3.9 - Methodological flowchart presenting: A) Data spectral analysis generating the best input dataset; B) Semantic segmentation approach; and C) Hierarchical mapping methodology. Numbers inside pink circles represent the sections where the processes are described.



SOURCE: Author's production.

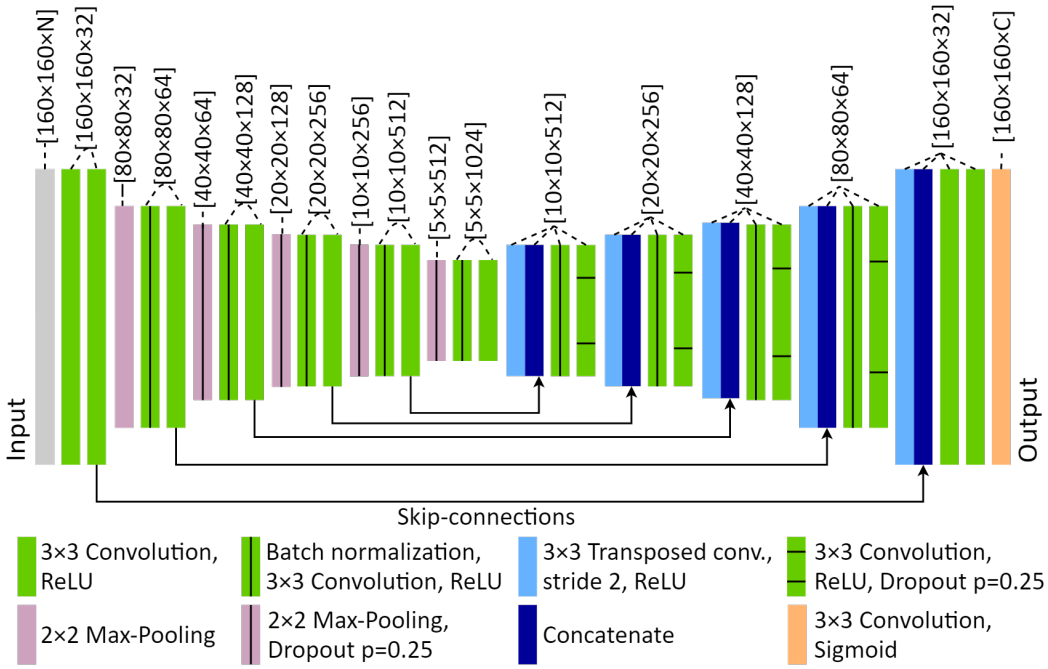
et al. (2015). However, it was modified as follows: while the original version uses unpadded convolutions, zero-padding (addition of zeros) was used to preserve the spatial size along the network. The up-sampling is based on transposed convolutions with a stride of two along both spatial dimensions.

Network parameters, such as the number of layers and number of filters per layer, are depicted in Figure 3.10. The input layer, represented in grey, has the size of the sample patches (160×160). The depth (N) is the number of bands. The output layer has the same size of the input layer, but the depth is represented by the number of classes C. Every other layer is represented according to the legend; a 2×2 Max-Pooling layer, for instance, is illustrated in pink. The numbers in brackets are the image sizes in each layer followed by the number of filters. As in the original U-net, skip-connections are used to concatenate information of high spatial resolution but low complexity with information of low spatial resolution but high complexity.

While the last layer of a network for semantic segmentation is usually modelled by

Softmax function, here Sigmoid function is used. This output layer was preferred here because it presented higher overall accuracies in preliminary tests. This allows the model to predict independent probabilities per class and per pixel. Final class predictions are obtained by choosing the respective classes with the highest probabilities. The Network was implemented in a Python environment, using Keras (CHOLLET et al., 2015) with TensorFlow (ABADI et al., 2015) as backend. The NVIDIA GeForce RTX 2070 super (8GB) GPU was used.

Figure 3.10 - Modified U-net architecture. N (in the input size) is the number of bands, while C (in the output size) corresponds to the number of classes.



SOURCE: Adapted from Kumar (2018).

3.4.2 Analysis of spectral input data

To test the network performance according to the input spectral data, eight datasets were created: RGB, RGB+EVI2, LSMM, RGB+LSMM, RGB+RedEdge, RGB+NIR1+NIR2+RedEdge, RGB+NIR1+NIR2 and the 8 bands. Table 3.7 presents the composition of each dataset. The EVI2 and the LSMM fractions are the same used in the GEOBIA approach, described in the Sections 3.3.1.1 and 3.3.1.2, respectively.

Table 3.7 - Description of the datasets used in the analysis of spectral input data.

Dataset	Composition
RGB	Red, Green and Blue spectral bands
RGB+EVI2	Red, Green and Blue spectral bands and EVI2
LSMM	Vegetation, Soil and Shade fractions from LSMM
RGB+LSMM	Red, Green and Blue spectral bands and the three fractions of LSMM
RGB+RedEdge	Red, Green, Blue and Red-Edge spectral bands
RGB+NIR1+NIR2+RedEdge	Red, Green, Blue, Red-Edge, NIR1 and NIR2 spectral bands
RGB+NIR1+NIR2	Red, Green, Blue, NIR1 and NIR2 spectral bands
8 bands	All eight bands from WorldView-2 image

SOURCE: Author's production.

3.4.2.1 Sample patches generation

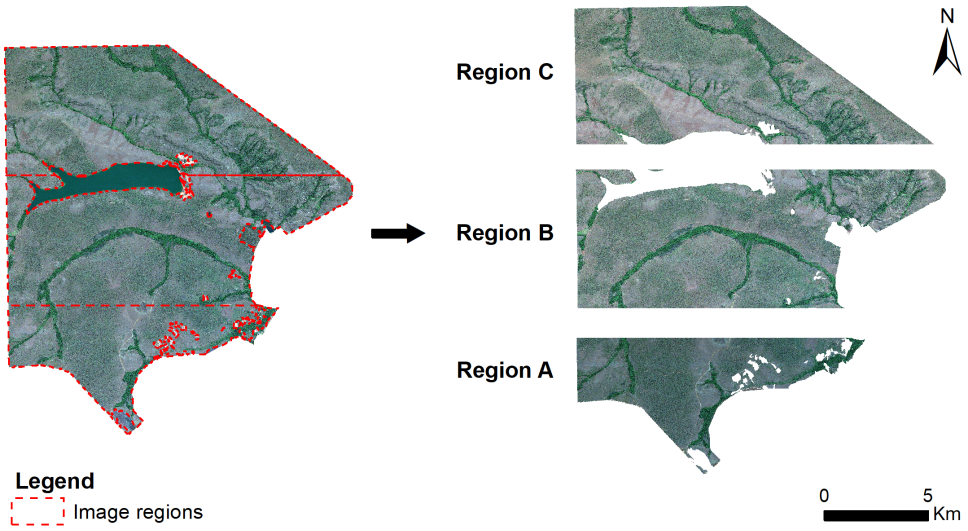
All datasets were divided into regions A, B and C (Figure 3.11), which contain roughly similar distributions of Forest, Savanna and Grassland classes. Thereafter, the datasets and the reference data (described in Section 3.2.1) were cropped into non-overlapping and adjacent tiles of 160×160 pixels to be used as samples. Tiles with any "no data" value (pixels originally covering built-up areas, water bodies, bare soil and burned areas) were removed from the samples.

In order to increase the amount of samples, data augmentation was performed. It consists in creating new samples based on slight modifications in the original data. In this research, six data augmentation techniques were employed: transposition, horizontal and vertical flips and three rotations (clockwise direction): 90, 180 and 270 degrees. The representation of each technique is presented in Figure 3.12.

3.4.2.2 Training, validation and test

The three regions samples were combined as follows: 70% of the samples from two regions (e.g., A and B) were randomly selected for training, and 30% for the network validation. The resulting network was then tested in the remaining region (e.g., C). This experiment was repeated three times, i.e., a cross-validation approach was performed. Table 3.8 shows the number of samples used in each experiment. For training and validation sets, those numbers already include samples generated by data augmentation.

Figure 3.11 - Regions A, B and C used to generate sample patches in the spectral data analysis.



SOURCE: Author's production.

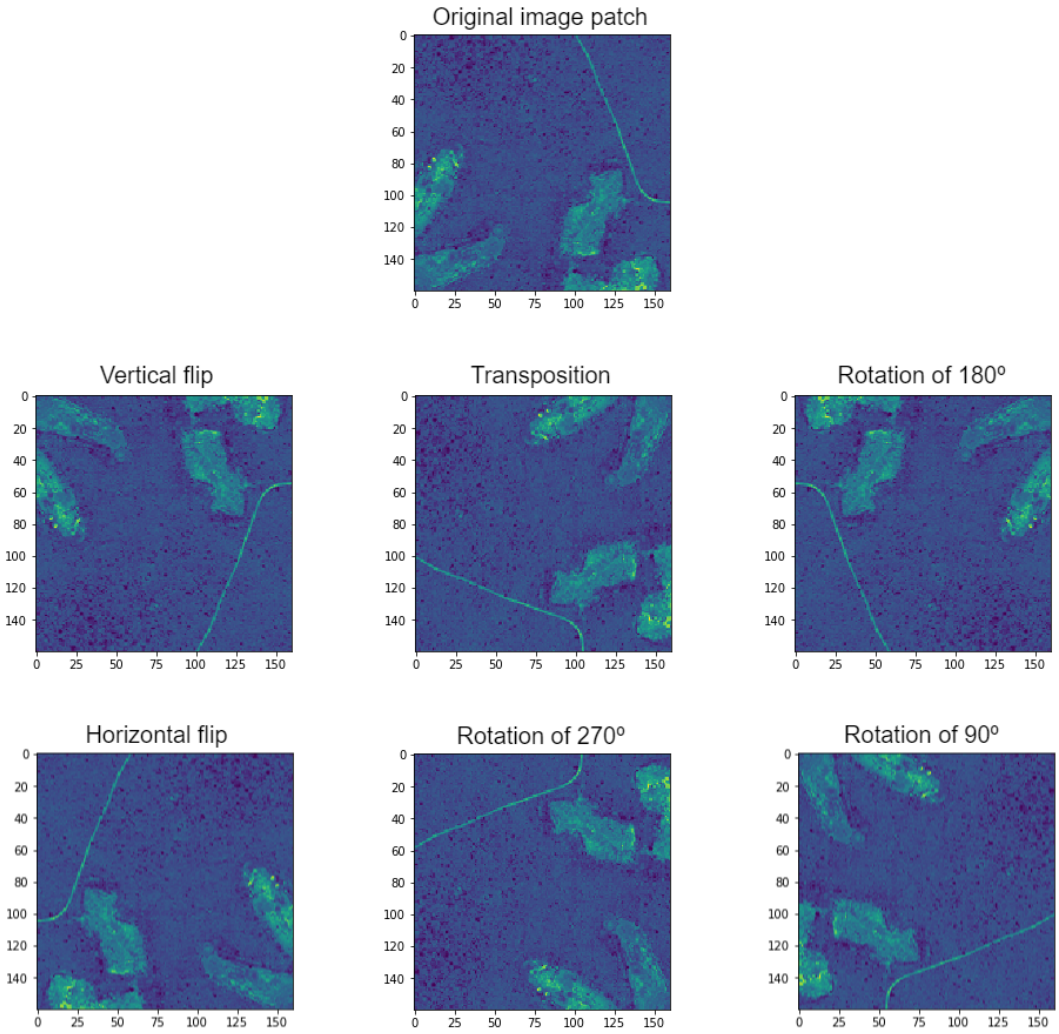
During training, the early stopping criterion (also known as "patience" in Keras) was set to 50, i.e., if after 50 epochs the validation accuracy did not increase, training is halted. In order to reduce the misclassification in the borders of the tiles, the predicted image of the semantic segmentation was created through a sliding window approach with steps of 20 pixels. All procedures (see also the first part of the methodology flowchart, Figure 3.9A) were executed for each of the eight datasets to classify the three classes (Forest, Savanna and Grassland). In addition to the thematic maps and the accuracy measures obtained in this approach, the best input dataset for our application is revealed. The best input dataset will be used in the next steps for the hierarchical classification.

Table 3.8 - Regions and number of samples used for training, validation and testing in each cross-validation experiment.

Experiment	Training Regions	Validation Regions	Testing Region
1	A + B: 70% (5439 samples)	A + B: 30% (2331 samples)	C (645 samples)
2	B + C: 70% (6951 samples)	B + C: 30% (2982 samples)	A (336 samples)
3	A + C: 70% (4802 samples)	A + C: 30% (2065 samples)	B (774 samples)

SOURCE: Author's production.

Figure 3.12 - Data augmentation techniques.



SOURCE: Author's production.

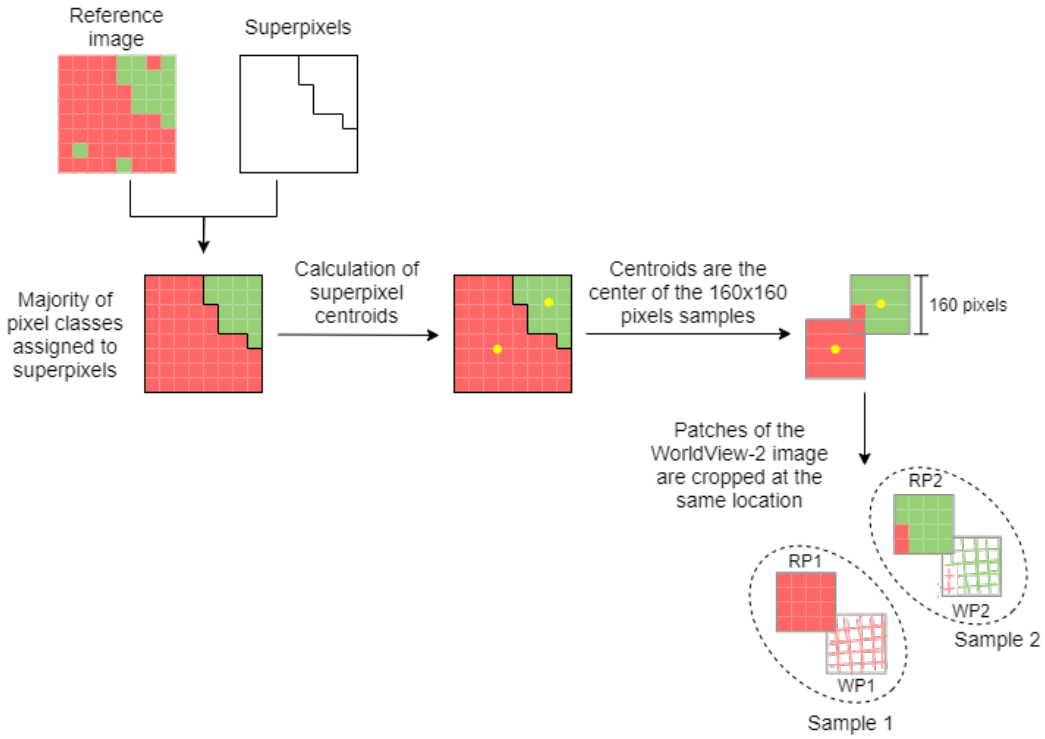
3.4.3 Hierarchical semantic segmentation

Using the input dataset that yielded the best performance in the previous spectral analysis step, a semantic segmentation approach was performed to hierarchically classify the physiognomies. Different from the first level, the classes in the second level are unbalanced, i.e., they do not present similar distributions across the three regions of the image (Figure 3.11). As the unbalanced class distribution can create artefacts, a different approach for sample generation was employed.

3.4.3.1 Random sample patches generation

To generate sample patches in the hierarchical semantic segmentation, the segmentation generated by the SLIC method (Section 3.2) was used. Figure 3.13 illustrates how sample patches are generated. For each Superpixel, a class was assigned based on the majority of corresponding pixel classes in the reference image. Thereafter, centroids of the Superpixels were calculated and used as the center point for each sample of 160 × 160 pixels. Each sample corresponds to the pair composed of one patch for reference image and one patch for WorldView-2 image.

Figure 3.13 - Scheme of how sample patches are generated. RP1 and WP1 are the Reference and WorldView-2 Patches of Sample 1, respectively, and RP2 and WP2 are the Reference and WorldView-2 Patches of Sample 2.



SOURCE: Author's production.

For each class of interest, 1000 centroids were randomly selected to generate a sample patch. The sample patches may contain areas of transition with other physiognomies, which is a positive aspect, because it will enable the network to learn the context where each physiognomy occurs. Similar to the previous experiment (Section 3.4.2), all samples that contained any no data value were excluded and, for the remaining

training and validation samples, the same six data augmentation techniques were applied.

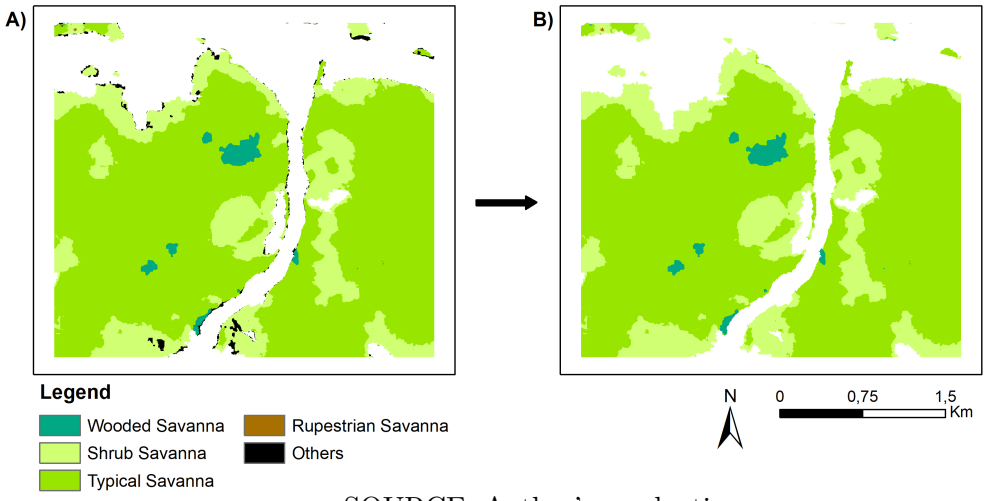
3.4.3.2 Semantic segmentation

For each class, the complete set of samples was randomly split, 70% being assigned for training and 30% for validation. In order to use the same sample generation approach in the entire hierarchical process, the semantic segmentation was repeated for the first level (Forest, Grassland and Savanna). With the resulting Savanna and Grassland maps, the same semantic segmentation approach was employed for each one of them. Using the trained model, the sliding window approach was also employed to create the entire predicted image. The final Cerrado physiognomies map (and the respective accuracy measures) is then composed of the Forest map (Gallery Forest) generated in the first level of classification, the Savanna map (Shrub Savanna, Typical Savanna, Woodland Savanna, Rupestrian Savanna and *Vereda*) and the Grassland map (Open Grassland, Shrub Grassland, Rupestrian Grassland, Humid Open Grassland), these last two generated in the second level of classification. These methodological steps are represented in Figure 3.9B and C.

The *Vereda* physiognomy has a minority presence in the BNP, so its area is not enough to be included into Deep Learning classification. Therefore, this physiognomy was included in the first level of classification (as part of the overall Savanna physiognomy), but it was manually identified in the second level. Consequently, it is present in the final map, but it was not considered in the confusion matrices and accuracy measures.

Another relevant detail concerns the generation of samples for Savanna and Grassland physiognomies in the second level of classification. When generating the samples of the four Grassland types, for instance, any other class present in the sample patch (e.g., Forest or Savanna classes) was considered as Others, a temporary class. If pixels corresponding to Others had simply been excluded, the network would be unable to understand patterns of transitions between Grassland physiognomies and other classes of Forest and Savanna, for example. As the network output is a probability for each pixel and each class, when the network tried to classify any pixel as Others in the predicted image, the second highest probability was considered, according to the example of Figure 3.14.

Figure 3.14 - Example of a predicted image of Savanna (second level of classification) demonstrating: A) Minority edges predicted as Others; and B) Replacement of Others class by the second highest probability of the network output.



SOURCE: Author's production.

3.5 Accuracy assessment

The obtained classification maps were then compared with the respective reference data, and then confusion matrices were generated. Since we are performing a hierarchical classification, misclassifications on the first level will directly influence the results of the second level, i.e., if a pixel of Shrub Savanna was classified as Grassland in the first level, it will still be considered as an error in the confusion matrix of the second level. Based on each confusion matrix (Table 3.9), the following measures were computed: Overall Accuracy (OA) (Equation 3.2), Recall (Equation 3.4), Precision (Equation 3.3) and F1-score (Equation 3.5).

The OA corresponds to the percentage of pixels (Deep Learning) or objects (GEOBIA) with the respective labels assigned correctly, considering the entire classified image. Precision is the proportion of pixels (or objects) predicted for a class, and actually belongs to that class; it is the complement of a commission error. Recall is the proportion of pixels (or objects) of a particular class successfully identified; it is the complement of an omission error. F1-score is the harmonic mean of Precision and Recall, computed for each class.

Table 3.9 - Example of a confusion matrix.

		Classification	
		Yes	No
Reference	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

SOURCE: Author's production.

$$OverallAccuracy(OA) = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.3}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.4}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3.5}$$

4 RESULTS

The experiments using GEOBIA approach, presented in Section 4.1, are based on Neves et al. (2019). The experiments using Deep Learning approach (Section 4.2) had as its starting point the results published in Neves et al. (2020).

4.1 GEOBIA approach

In order to test the contribution of several feature sets, six experiments were performed for both levels of classification. The OA values are presented in Table 4.1. On the first level (distinguished Forest, Savanna and Grassland), the highest OA was 88.2%. It was achieved considering all features except the VIs ($\text{Except}_{\text{VegInd}}$). Removing the features of TC components ($\text{Except}_{\text{TCap}}$), LSMM fractions ($\text{Except}_{\text{SMA}}$) and spectral bands ($\text{Except}_{\text{Spectral}}$) provided similar OA values, ranging from 87.6% to 87.9%. On the other hand, when texture features ($\text{Except}_{\text{TX}}$) were not included, the OA dropped to 85.2%, indicating that they play an important role in the classification of Cerrado physiognomies in GEOBIA approach.

Table 4.1 - Overall accuracy (%) of each level of hierarchical classification in the six experiments testing several features.

Experiment	**First Level**	**Second Level**	
		Savanna	**Grassland**
All (Spectral + TCap + SMA + VegInd + TX)	87.7	65.7	**64.8**
$\text{Except}_{\text{TCap}}$	87.6	65.2	64.2
$\text{Except}_{\text{SMA}}$	87.8	66.8	64.5
$\text{Except}_{\text{VegInd}}$	**88.2**	**67.3**	62.4
$\text{Except}_{\text{Spectral}}$	87.9	66.0	63.6
$\text{Except}_{\text{TX}}$	85.2	62.3	62.7

SOURCE: Author's production.

The confusion matrix for the first classification level using the $\text{Except}_{\text{VegInd}}$ dataset, which achieved the highest OA value, is presented in Table 4.2. The values in the matrix represents the number of superpixels. Forest had the highest F1-score (0.95), while Savanna and Grassland achieved 0.90 and 0.83, respectively. As Gallery Forest is the only relevant Forest physiognomy in the BNP, the patterns variability in this class is lower than in Savanna and Grassland, enabling the highest accuracy value.

Table 4.2 - Confusion matrix (in number of superpixels), Precision, Recall and F1-score for the first level of classification, using $\text{Except}_{\text{VegInd}}$ dataset. OA = 88.21%.

	Reference				
Predicted	**Grassland**	**Savanna**	**Forest**	**Total**	**Precision**
Grassland	275	51	0	326	**0.84**
Savanna	60	538	4	602	**0.89**
Forest	0	5	85	90	**0.94**
Total	335	594	89		
Recall	**0.82**	**0.91**	**0.96**		
F1-score	**0.83**	**0.90**	**0.95**		

SOURCE: Author's production.

For the Savanna physiognomies classification, the highest OA (67.3%) was also obtained using the experiment $\text{Except}_{\text{VegInd}}$. The largest drop in the OA occurred when TX features were excluded from the dataset (62.3%). The use of TX features proved to be essential for discriminating these classes, once they are capable of capturing the alternation between different types of vegetation, such as trees, shrub vegetation and herbaceous vegetation. Initially, in Table 4.2, there were 594 samples of Savanna. In the second level of classification for Savanna physiognomies (Table 4.3), 538 samples (46 for Woodland Savanna, 354 for Typical Savanna and 138 for Shrub Savanna) were considered, since they were correctly classified in the first level.

Despite the low F1-score of Woodland Savanna (0.30), the hierarchical classification was able to separate this physiognomy from Forest in the first level, since its samples were classified as Savanna and moved to the second level. However, in the second level, more than half of its samples were classified as Typical Savanna. This physiognomy achieved the highest F1-score (0.82) and only eight of its superpixels were misclassified in the first level. The majority of errors of Savanna in the first level was regarding the Shrub Savanna, with 48 misclassified superpixels, resulting in a F1-score of 0.51. To considerate the errors in the physiognomies accuracy measures, superpixels of Savanna misclassified in the first level were included in the Recall calculation in the second level (Table 4.3).

Table 4.3 - Confusion matrix (in number of superpixels), Precision, Recall and F1-score for the Savanna physiognomies, in the second level of classification, using the experiment $\text{Except}_{\text{VegInd}}$.

	Reference				
Predicted	**Woodland Savanna**	**Typical Savanna**	**Shrub Savanna**	**Total**	**Precision**
Woodland Savanna	9	6	0	15	**0.60**
Typical Savanna	37	322	65	424	**0.76**
Shrub Savanna	0	26	73	99	**0.74**
Total	46	354	138		
Recall	**0.20**	**0.89**	**0.39**		
F1-score	**0.30**	**0.82**	**0.51**		

SOURCE: Author's production.

The second level for Grassland physiognomies was generated using the experiment with all features, since it achieved the highest OA (64.78% - Table 4.1). Removing VIs from the feature set dropped the OA value to the lowest one (62.4%). Differently from what happened to the first level and the discrimination of Savanna physiognomies, the VIs were essential to improve the classification of Grassland physiognomies, once higher values of VIs were noticed for the Shrub Grassland when compared to the Open and Rupestrian Grassland.

The superpixels correctly classified as Grassland in the first level (275 from 335) were differentiated into Open, Shrub and Rupestrian Grasslands (Table 4.4). From these physiognomies, 16 (14.3%), 40 (18.9%) and 4 (33.4%) superpixels, respectively, were misclassified in the first level and considered in Recall calculation in the second level. The Rupestrian Grassland had the lowest F1-score (0.27). The reduced amount of samples may have affected the performance for this class. The Open and Shrub Grasslands presented considerable confusions between each other, probably in their transition areas, resulting in F1-scores of 0.63 and 0.77, respectively.

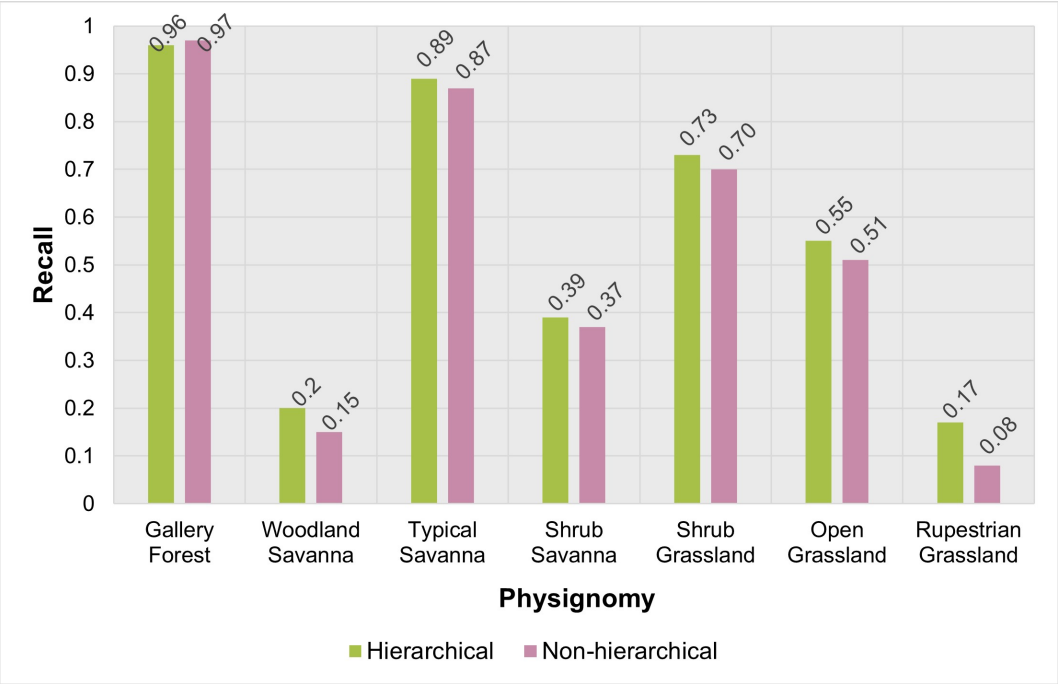
Table 4.4 - Confusion matrix (in number of superpixels), Precision, Recall and F1-score for the Grassland physiognomies, in the second level of classification, using the experiment with all features.

	Reference				
Predicted	**Open Grassland**	**Shrub Grassland**	**Rupestrian Grassland**	**Total**	**Precision**
Open Grassland	62	18	4	84	**0.74**
Shrub Grassland	33	153	2	188	**0.81**
Rupestrian Grassland	1	0	2	3	**0.67**
Total	96	171	8		
Recall	**0.55**	**0.73**	**0.17**		
F1-score	**0.63**	**0.77**	**0.27**		

SOURCE: Author's production.

In order to investigate the superiority of hierarchical approach, the Recall values of each physiognomy were compared to a non-hierarchical classification. The comparison is presented in Figure 4.1. Gallery Forest was the only class with a slightly better accuracy using the non-hierarchical approach. However, the presence of a unique Forest class on the study site hinders the evaluation of the hierarchy for the classification of Forest physiognomies. The remaining six classes had higher accuracies in the hierarchical approach. Therefore, hierarchical OA (68.9%) was also higher than the non-hierarchical OA (66.4%).

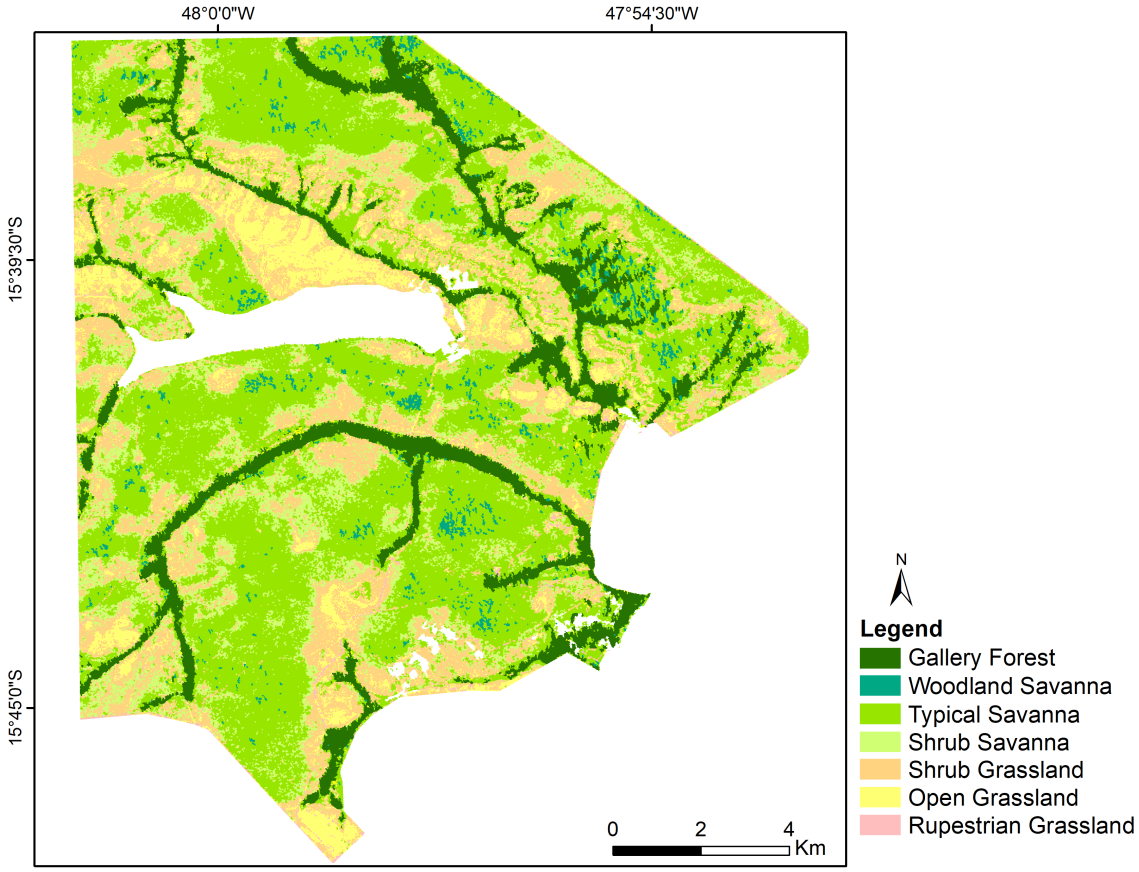
Figure 4.1 - Recall of the physiognomies and in the hierarchical and the non-hierarchical approach. Hierarchical classification OA = 68.9% and Non-hierarchical classification OA = 66.4%.



SOURCE: Author's production.

Classification results based on Random Forest models with the highest OA (Table 4.1), are presented in Figure 4.2. The vegetation map shows a consistent result, with correct transitions between the studied physiognomies. However, some physiognomies (e.g., Woodland Savanna and Rupestrian Grassland) still present accuracies below 50% and need more sophisticated techniques.

Figure 4.2 - Cerrado physiognomies map using GEOBIA hierarchical approach.



SOURCE: Author's production.

4.2 Deep Learning approach

4.2.1 Assessment of the spectral input information

In Table 4.5, the accuracies of the training and validation steps for the assessment of the spectral input information are presented, as well as the number of epochs needed to stabilize the network in each experiment. It took about 5 hours to train each network. Due to the patience parameter of 50 (see Section 3.4.2.2 for more details), all experiments had their training stopped before 100 epochs were reached, stabilization of the networks occurred between epoch 5 and epoch 93. If the training had been kept longer, the training accuracy values would have improved, but the validation accuracies started to become worse, indicating that the model might be overfitted. In general, all accuracy values were higher than 87.5%. The highest accuracies in training were reached using the RGB+NIR1+NIR2+RedEdge dataset

in Experiments 1 and 3, while, in Experiment 2, the highest accuracy was obtained for RGB+EVI2. In the validation step, the highest accuracies for Experiments 1 and 2 were reached using RGB+EVI2, while the highest accuracy in Experiment 3 was observed with the RGB+NIR1+NIR2+RedEdge dataset.

Table 4.5 - Training and validation accuracies for all datasets in the three experiments (see description in Table 3.8). The highest values in training and validation for the three experiments are given in bold.

Input Dataset	Exp.	Epochs	Training Acc. (%)	Validation Acc. (%)
RGB+LSMM	1	23	88.5	89.8
	2	52	91.7	88.2
	3	16	89.1	87.9
8 band	1	47	91.3	88.6
	2	34	90.6	89.0
	3	77	92.1	89.4
LSMM	1	25	89.9	89.8
	2	31	89.8	89.4
	3	89	93.5	89.3
RGB+NIR1+NIR2	1	5	87.7	88.9
	2	35	89.4	88.6
	3	93	93.0	89.5
RGB+NIR1+NIR2+RedEdge	1	55	**91.7**	89.9
	2	29	88.7	89.2
	3	87	**93.7**	**90.4**
RGB+RedEdge	1	10	88.6	89.4
	2	28	90.5	89.2
	3	53	91.4	89.8
RGB	1	47	89.2	89.1
	2	70	92.5	88.7
	3	28	90.1	89.2
RGB+EVI2	1	47	89.7	**90.4**
	2	58	**92.7**	**89.7**
	3	32	89.4	88.6

SOURCE: Author's production.

For the test step, the OAs and the F1-score per class for each input dataset are presented in Table 4.6. The OA varied from 87.4%, using RGB+LSMM, to 89.3% with the RGB+EVI2 dataset. It could be expected that the 8 band dataset would achieve the highest performance, since it has the most bands and, consequently, the most spectral information. However, it obtained the second worst OA value of 87.6%. Despite presenting the lowest OA, the RGB+LSMM dataset had the highest F1-score (0.91) for the Forest class. This is also reflected in the classes delineation

in the thematic map. For the Savanna and Grassland classes, the highest F1-scores (0.92 and 0.84, respectively) were achieved with the same dataset that had the best OA, RGB+EVI2.

Table 4.6 - Overall Accuracies - OA (%) and classes F1-score for all input datasets.

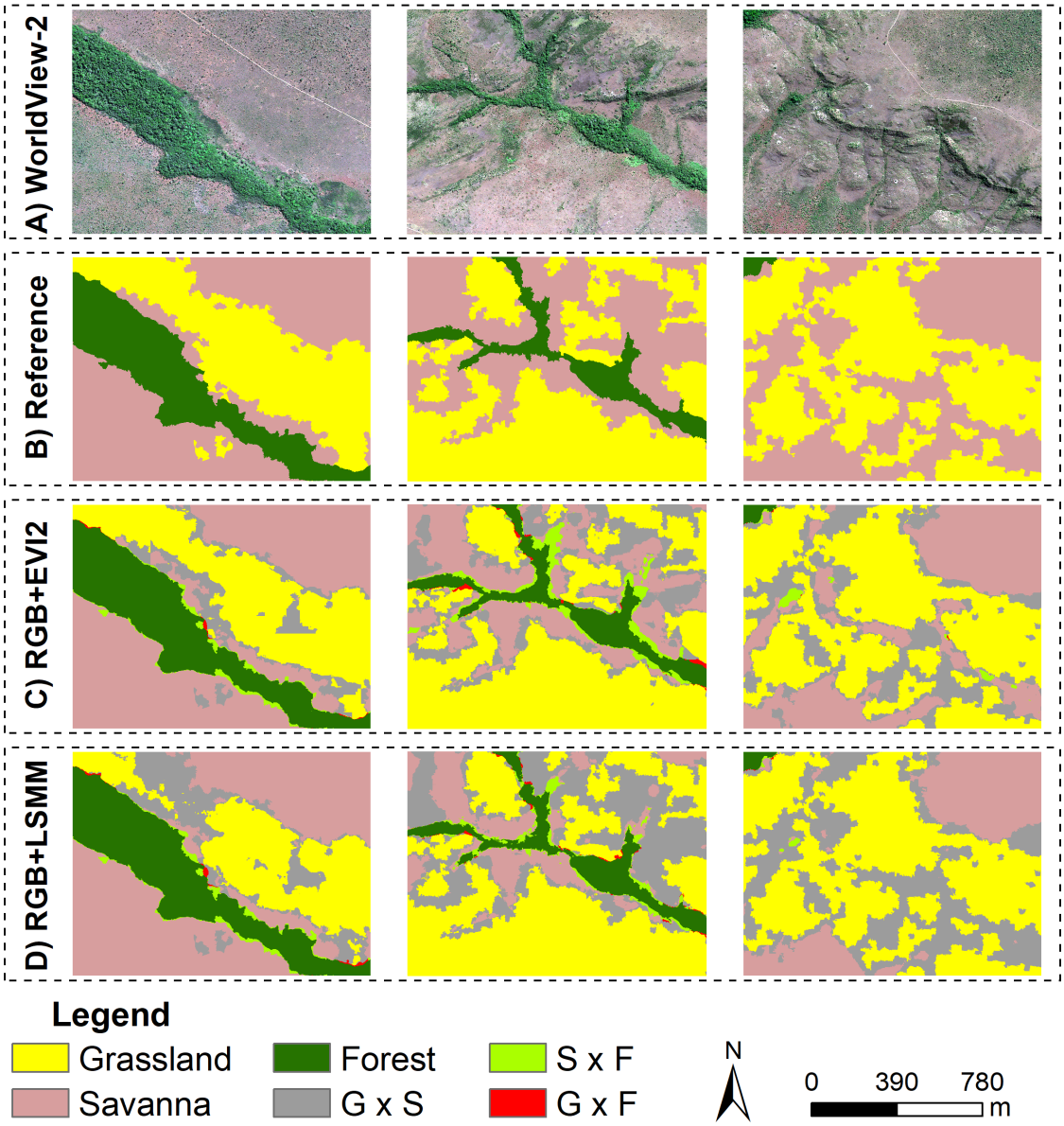
Input dataset	**OA (%)**	**Classes F1-score**		
		Grassland	**Savanna**	**Forest**
RGB+LSMM	87.4	0.81	0.90	**0.91**
8 band	87.6	0.81	0.90	0.89
LSMM	87.8	0.82	0.90	**0.91**
RGB+NIR1+NIR2	87.9	0.82	0.90	0.89
RGB+NIR1+NIR2+RedEdge	88.3	0.83	0.91	0.90
RGB+RedEdge	88.4	0.82	0.91	0.89
RGB (532)	88.6	0.83	0.91	0.89
RGB+EVI2	**89.3**	**0.84**	**0.92**	**0.91**

SOURCE: Author's production.

In order to analyze in more detail the results of the Table 4.6, Figure 4.3 shows some selected patches of the WorldView-2 image, the reference and the thematic maps using the RGB+EVI2 and the RGB+LSMM datasets. In this scale, the misclassified areas between Grassland and Savanna (GxS), Grassland and Forest (GxF) and Savanna and Forest (SxF) are highlighted in different colors. Despite the small difference between the best (RGB+EVI2) and the worst (RGB+LSMM) datasets of 1.9 percentage point, the resulting maps show significant dissimilarities.

In all maps, the major areas of misclassification occur between Grassland and Savanna (GxS), followed by Savanna and Forest (SxF). There are only few areas of confusion between Grassland and Forest (GxF). This behavior is expected, since the confusions of classification occur mainly in transition areas, considering an increasing scale of vegetation density in the existing physiognomies (i.e., GxS and SxF). Additionally, the higher Forest F1-score with RGB+LSMM, already observed in Table 4.6, is also reflected in the maps. In Figure 4.3, it is possible to notice a better delineation of the Forest areas when using this dataset, even better than in RGB+EVI2 dataset results.

Figure 4.3 - Patches of: A) The WorldView-2 image; B) The reference data; C) Resulting thematic map using RGB+EVI2 dataset; and D) Resulting thematic map using RGB+LSMM datasets. GxS are the misclassified areas between Grassland and Savanna; SxF, between Savanna and Forest; and GxF, between Grassland and Forest.



SOURCE: Author's production.

4.2.2 Hierarchical classification of Forest, Savanna and Grassland

For the hierarchical classification, the input dataset composed by RGB+EVI2 bands was used, since it achieved the best performance in the assessment of the spectral input information for the first level, especially for Savanna and Grassland classes. For

the first level of classification, the accuracy during training was of 97.9%, achieved after 147 epochs. The confusion matrix for the validation step is presented in Table 4.7. The matrix is presented in terms of number of pixels, and the OA was of 92.8%. Forest obtained the highest F1-score of 0.95, and the other two classes achieved F1-scores higher than 0.91. The Grassland Recall (0.89) was the only metric lower than 0.90, as 10.5% of the grassland pixels were misclassified as Savanna.

Table 4.7 - Confusion matrix (in number of pixels), Precision, Recall and F1-score for the first level of classification, using the RGB+EVI2 dataset. OA = 92.8%.

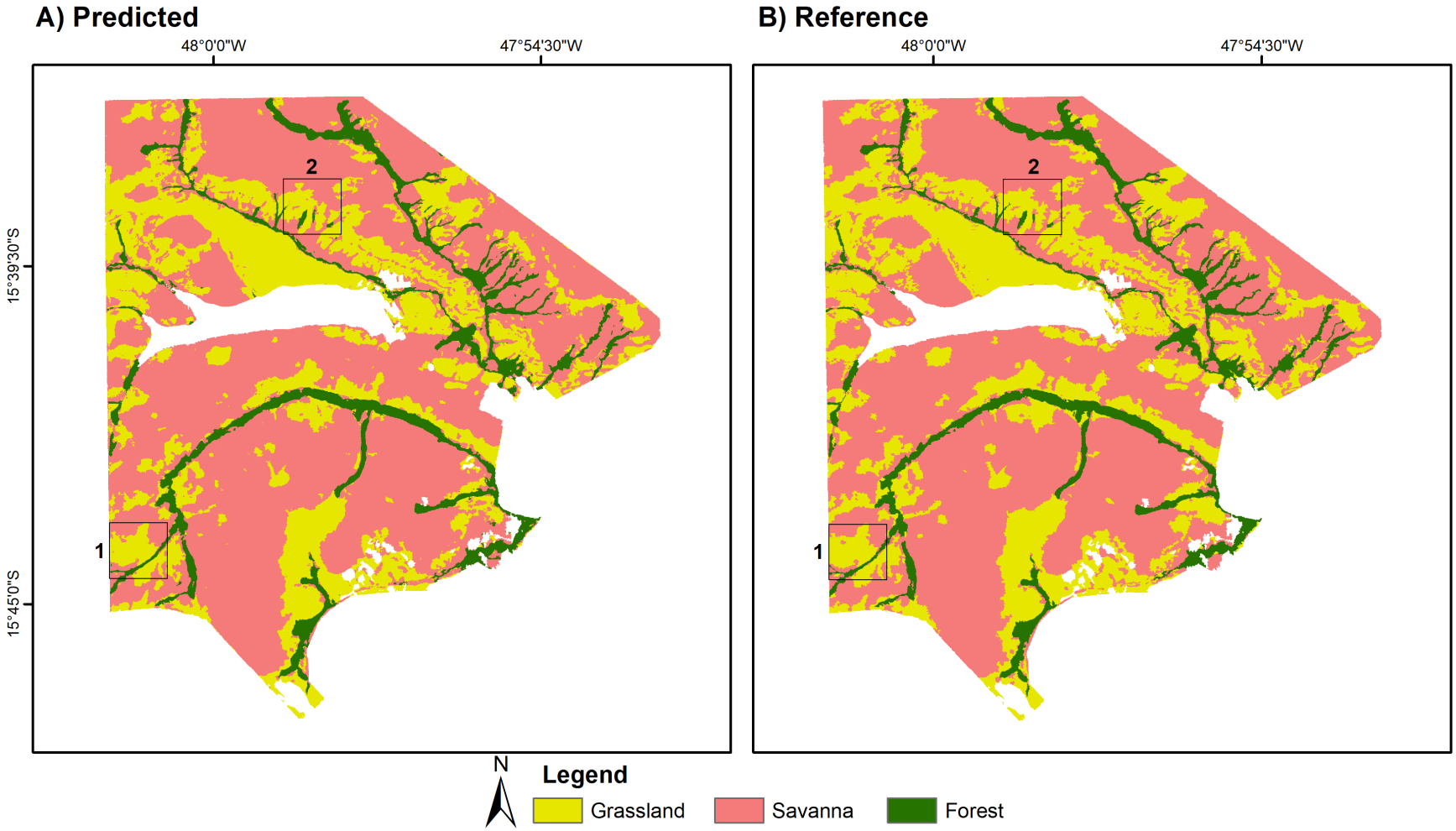
	Reference				
Predicted	**Grassland**	**Savanna**	**Forest**	**Total**	**Precision**
Grassland	5,906,317	315,673	8,084	6,230,074	**0.95**
Savanna	697,394	7,788,065	93,307	8,578,766	**0.91**
Forest	36,929	96,423	2,441,111	2,574,463	**0.95**
Total	6,640,640	8,200,161	2,542,502		
Recall	**0.89**	**0.95**	**0.96**		
F1-score	**0.92**	**0.93**	**0.95**		

SOURCE: Author's production.

Confusion between Grassland and Savannas, which presented the highest error rate, is also the most visually evident one. However, it can be noticed that it occurs mainly along class borders, where it is really hard to define when one class becomes another, even in field campaigns. The classification predicted in the first level is presented in Figure 4.4, where it is possible to note that the misclassified areas occur in the transitions between the classes. Two regions in predicted image and reference were selected to be highlighted.

In order to observe the misclassifications in more detail, the zoom of these two regions and the image representing the difference between predicted image and reference can be seen in Figure 4.5. The majority of misclassified regions, in grey, represent the errors between Grassland and Savanna, which were the most frequent in the confusion matrix (Table 4.7) as well. The errors between Forest and Savanna, in light green, occur mainly in the borders of Gallery Forests. Finally, the misclassifications between Forest and Grassland, in red, were the least frequent in the confusion matrix and, consequently, are rarely observed in the difference image.

Figure 4.4 - Result of the first level of classification, differentiating Forest, Savanna and Grassland. A) Predicted image; and B) Reference. The two boxes in the images contain the regions that will be highlighted in Figure 4.5



SOURCE: Author's production.

4.2.3 Detailed physiognomies mapping

In a hierarchical classification, results of the first level affect directly the performance of the second level. Considering the detailed reference of Savanna physiognomies and the resulting classification of the first level, it is shown that 92.7% of the Woodland Savanna was correctly included in Savanna class, while 6.4% was misclassified as Forest (Table 4.8). The Typical Savanna had the highest percentage of area classified as Savanna (98.2%). While the Rupestrian Savanna showed the highest percentage of misclassification, 88.3% of the area were classified as Savanna, while 10.5% was classified as Grassland. Pixels of Savanna classified as Forest or Grassland were included as errors in the Recall of the second level of classification for Savanna physiognomies (Table 4.9).

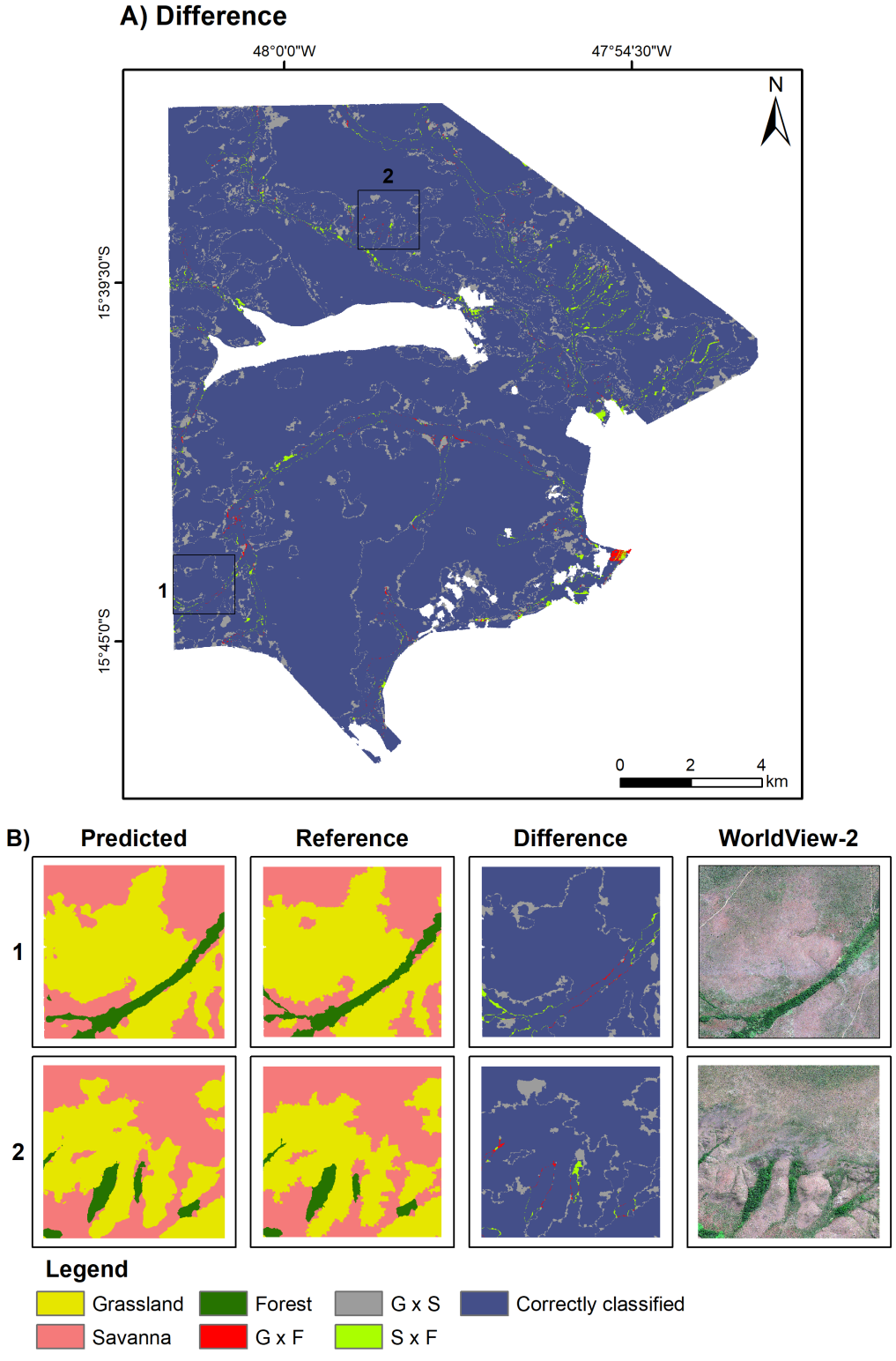
Similar to Table 4.8, Typical Savanna showed the best performance in the classification of the detailed physiognomies, with an F1-score of 0.91 (Table 4.9). The other three Savanna physiognomies achieved F1-scores from 0.84 (Shrub Savanna) to 0.88 (Rupestrian Savanna). Most of the misclassified pixels of Woodland and Shrub Savannas were labeled as Typical Savanna. This was an expected behavior, since Shrub, Typical and Woodland Savanna (in this order) compose an increasing scale of vegetation density and biomass. Regarding Rupestrian Savanna, the network was able to identify its pattern among the Savanna physiognomies, resulting in a Precision of 0.91. However, the confusion of 10.5% of this physiognomy with Grassland in the previous level of classification lead to a Recall of only 0.84. During the training step, the accuracy for Savanna physiognomies was of 96.6% when 190 epochs were achieved. In the validation step, the OA was of 95.6%. However, when errors from the first level of classification are propagated to the second level, then the OA becomes 86.1% (Table 4.9).

Table 4.8 - Analysis of the result of the second level of classification for Savanna physiognomies regarding the first level resulting map (%).

	Reference (Savanna physiognomies)			
Predicted (first level)	**Woodland Savanna**	**Typical Savanna**	**Shrub Savanna**	**Rupestrian Savanna**
Savanna (correct)	92.7	98.2	91.0	88.3
Classif. as Grassland	0.9	0.7	8.4	10.5
Classif. as Forest	6.4	1.1	0.6	1.2

SOURCE: Author's production.

Figure 4.5 - A) Difference between predicted image and reference for the first level of classification. B) Zoom of two regions to show the result in more detail. In the images of Difference, G x S are the misclassified areas between Grassland and Savanna; S x F, between Savanna and Forest; and G x F, between Grassland and Forest.



SOURCE: Author's production.

Table 4.9 - Confusion matrix (in number of pixels) for the Savanna physiognomies, in the second level of classification (OA = 86.1%).

	Reference					
Predicted	**Woodland Savanna**	**Typical Savanna**	**Shrub Savanna**	**Rupestrian Savanna**	**Total**	**Precision**
Woodland Savanna	286,275	40,812	2,348	888	330,323	**0.87**
Typical Savanna	23,794	4,107,171	295,190	6,227	4,432,382	**0.93**
Shrub Savanna	2,258	325,652	2,402,537	4,019	2,734,466	**0.88**
Rupestrian Savanna	232	10,219	11,411	231,146	253,008	**0.91**
Total	337,128	4,566,258	2,980,008	274,599		
Recall	**0.85**	**0.90**	**0.81**	**0.84**		
F1-score	**0.86**	**0.91**	**0.84**	**0.88**		

SOURCE: Author's production.

Considering the reference of the detailed Grassland physiognomies, almost all (99.4%) of Open Grassland was correctly classified as Grassland in the first level (Table 4.10). Rupestrian, Shrub e Humid Open Grassland had 94.1%, 85.5% and 79.4% of their areas classified as Grassland, respectively. For the Shrub and Humid Open Grassland, 14.2% and 18.0%, respectively, were misclassified as Savanna in the first level. In the hierarchical classification system proposed by Ribeiro and Walter (RIBEIRO; WALTER, 2008), Humid Open Grassland is a subtype of Open Grassland. However, it was considered an independent class in this work as it presents a pattern very different from the traditional Open Grassland. Besides that, preliminary experiments showed that separating these two classes increased the Open Grassland's OA by more than 2 percentage points.

Table 4.10 - Analysis of the result of the second level of classification for Grassland physiognomies regarding the first level resulting map (%).

	Reference (Grassland physiognomies)			
Predicted (first level)	**Open Grassland**	**Shrub Grassland**	**Rupestrian Grassland**	**Humid Open Grassland**
Grassland (correct)	99.4	85.5	94.1	79.4
Classif. as Savanna	0.5	14.2	5.5	18.0
Classif. as Forest	0.1	0.3	0.4	2.6

SOURCE: Author's production.

During training step, the OA of the classification of the detailed Grassland physiognomies was 96.3% with 171 epochs. In the validation, the network achieved an OA of 95.6%. After the inclusion of the errors from the first level of the classification, the Grassland physiognomies OA decreased to 85.0%. The resulting confusion matrix is presented in Table 4.11. The F1-scores varied from 0.86 (Humid Open Grassland) to 0.94 (Open Grassland). Shrub and Rupestrian Grassland had F1-scores of 0.89 and 0.93, respectively. The largest amount of misclassified pixels of Shrub Grassland were classified as Open Grassland. For Open, Rupestrian and Humid Open Grassland, the largest amount of misclassified pixels was classified as Shrub Grassland, the majority Grassland physiognomy in the BNP.

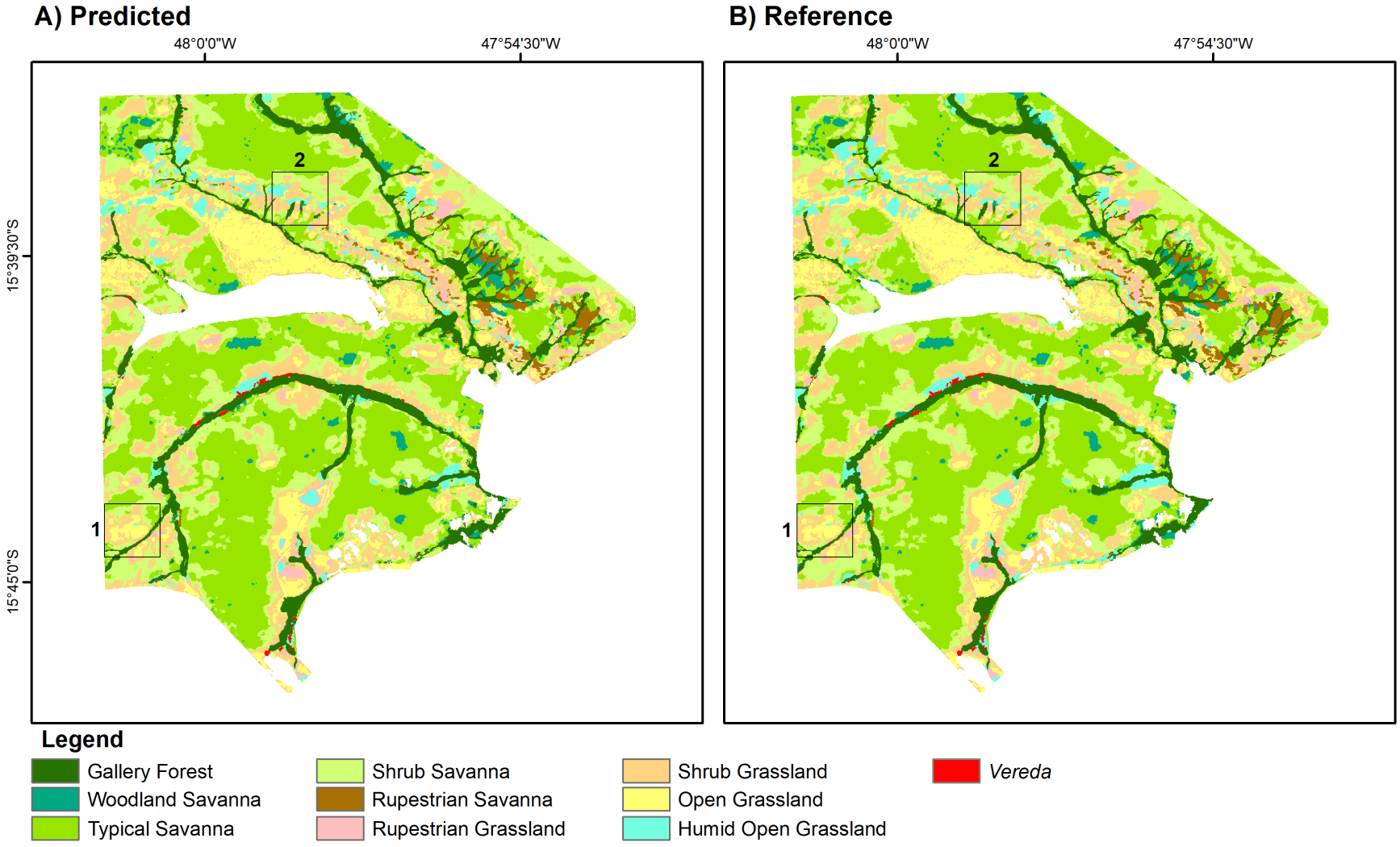
Table 4.11 - Confusion matrix (in number of pixels) for the Grassland physiognomies, in the second level of classification (OA = 85.0%).

	Reference					
Predicted	**Open Grassland**	**Shrub Grassland**	**Rupestrian Grassland**	**Humid Open Grassland**	**Total**	**Precision**
Open Grassland	1,662,547	78,198	4,319	3,426	1,748,490	**0.95**
Shrub Grassland	110,249	2,858,558	9,961	16,827	2,995,595	**0.95**
Rupestrian Grassland	3,898	12,359	379,596	3,283	399,136	**0.95**
Humid Open Grassland	4,170	12,286	938	740,197	757,591	**0.98**
Total	1,791,208	3,461,580	419,588	962,759		
Recall	**0.93**	**0.83**	**0.90**	**0.77**		
F1-score	**0.94**	**0.89**	**0.93**	**0.86**		

SOURCE: Author's production.

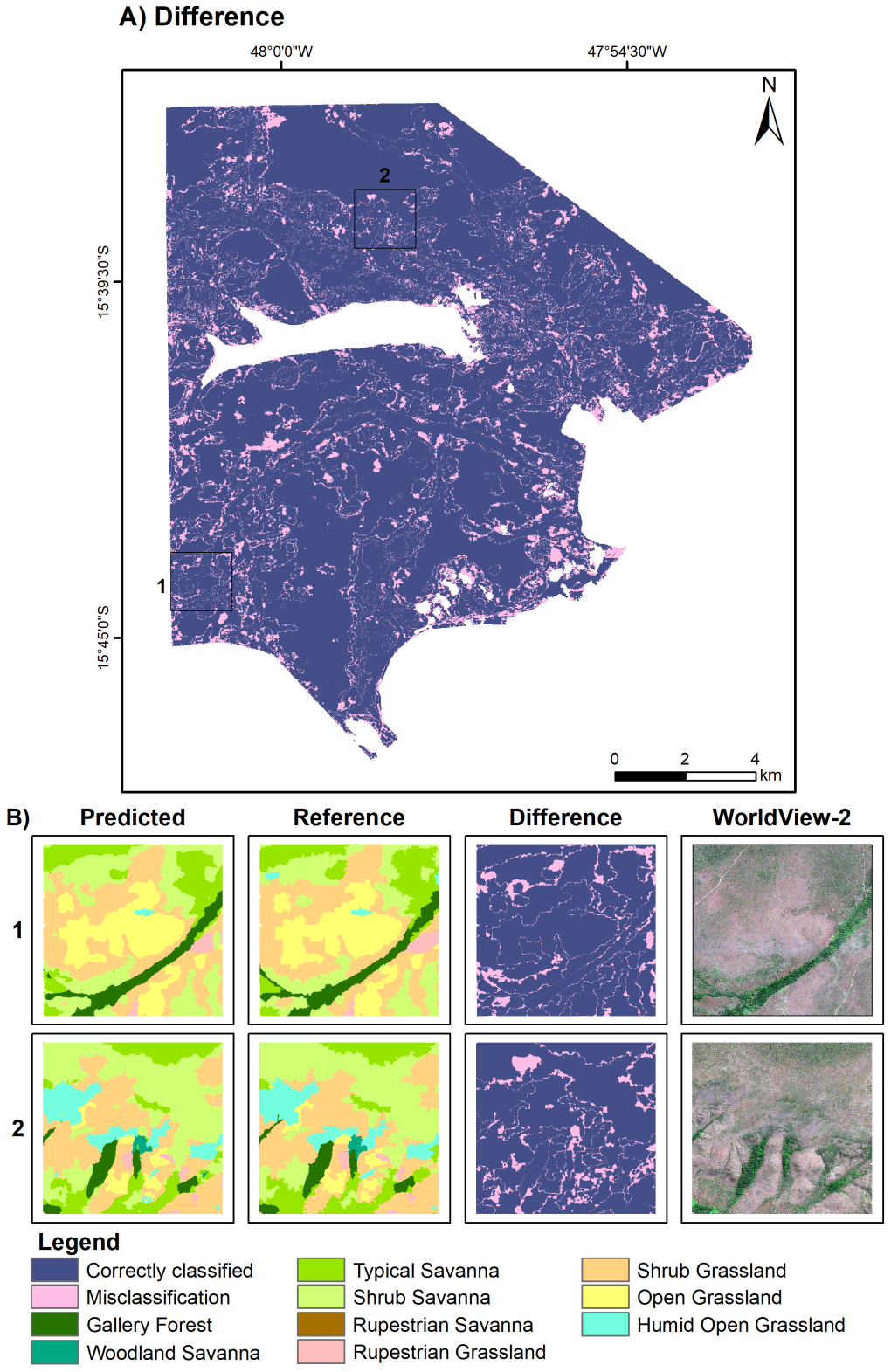
Results of Savanna and Grassland physiognomies is presented in Figure 4.6. This figure includes as well the Gallery Forest, generated in the first level of classification. Therefore, a detailed mapping of all physiognomies in the BNP is created. In addition to the reference and the predicted images, the image representing the difference between both is presented in Figure 4.7 to clearly show correctly classified areas and misclassifications. In order to better visualize the development of the hierarchical classification, the two highlighted regions are the same than the ones presented in Figure 4.5. Just like in the Tables 4.9 and 4.11, errors from the first level of classification are carried over to the second level, so no misclassified area from the first level (grey, light green and red colors in Figure 4.5) can become correctly classified (dark blue in Figure 4.7) in the second level of classification.

Figure 4.6 - Result of the second level of classification, including Savanna and Grassland physiognomies and the Gallery Forest. A) Predicted image; and B) Reference. The two boxes in the images contain the regions that will be highlighted in the next figure.



SOURCE: Author's production.

Figure 4.7 - A) Difference between reference and predicted image for the second level of classification. B) Zoom of two regions to show the result in more detail



SOURCE: Author's production.

5 DISCUSSION

5.1 GEOBIA approach

As observed in the GEOBIA approach, the consideration of the Ribeiro and Walter classification system (RIBEIRO; WALTER, 2008) hierarchy has the potential to improve the methodologies for mapping the Cerrado physiognomies through RS imagery. This improvement using hierarchy was also observed by Ribeiro et al. (2020). From the RS image perspective, the differentiation between some physiognomies (e.g., Shrub, Typical and Woodland Savanna) rely on the identification of a context that shows an increasing vegetation density and height. For this contextual aspect, the use of objects (superpixels, in this study) are essentially more suitable than a per pixel approach.

The GEOBIA results showed that some features are better to characterize some physiognomies than others. The VIs were useful for the discrimination of Grasslands physiognomies. Additionally, the considerable drop in the accuracy values when the texture features were excluded shows that they were essential for the first and the second levels of classification. The GEOBIA approach presented higher accuracy rates when compared to other works that used the same study site, but different approaches (GIROLAMO-NETO, 2018; SCHWIEDER et al., 2016). For Shrub Grassland and Typical Savanna, for instance, the accuracies were higher than Girolamo-Neto (2018). Moreover, Gallery Forest, Typical Savanna, Shrub Grassland and Open Grassland had better accuracies than Schwieder et al. (2016). However, some physiognomies (e.g., Woodland Savanna and Rupestrian Grassland) still presented low accuracies. In the Deep Learning approach, despite generating pixel-wise classifications, the application of convolutions enables the extraction of rich context information (MA et al., 2019). Therefore, this approach can represent the physiognomies context without using texture features as input.

5.2 Deep Learning approach

5.2.1 Assessment of the spectral input information

The majority of research in the optical RS field, that apply Deep Learning techniques, use images of the visible spectrum, which present the three most common channels (spectral bands), the RGB (Red, Green and Blue) (GUIRADO et al., 2017; KATTENBORN et al., 2019). Others also include a NIR channel (NOGUEIRA et al., 2016; JOZDANI et al., 2019). These works rarely use all available satellite bands, sometimes

because of the increase in processing time, or because the network architecture and the algorithms are prepared to use only 3 input layers. In the RS field, several extra spectral information, such as Yellow or Red-edge bands (available in WorldView-2), may be used. Zhang et al. (2018a) tested the efficiency of using datasets containing 4 and 8 bands of the WorldView-2 and WorldView-3 images. The authors achieved better accuracies with the 8-band dataset, although they classified only different urban targets (e.g., buildings and roads). While more bands in general improve the results, in this study, it was observed that the use of some bands, like the Yellow band (present in the 8 Band dataset), does not improve the differentiation of different types of vegetation patterns. Thus, the increase in the number of bands is not necessarily directly related to an increase of OA.

Deep Learning networks learn features from the training data to identify the desired classes. For this reason, it is believed that it is not necessary to give the network supplementary information, commonly called handcrafted features (e.g., vegetation indices, LSMM fractions), in addition to satellite spectral bands (ZHU et al., 2017). However, our results showed the opposite: by combining a handcrafted feature (vegetation index) to original data, we obtained results with the best OA. The EVI2 enhanced the information regarding vegetation biomass in a way that does not occur when using only the Red and the NIR bands. Thus, the OA of the RGB+EVI2 dataset was higher than the OA of other datasets containing Red and NIR bands (8 band, RGB+NIR1+NIR2 and RGB+NIR1+NIR2+RedEdge). A better performance when including vegetation indices was also observed by Kerkech et al. (2018), but the authors used different indices and tested them in a different domain (crop disease detection).

On the other hand, the inclusion of the three LSMM fractions (Vegetation, Soil and Shade) as input did not increase the OA. However, vegetation and shade LSMM fractions were capable to highlight and represent well the conditions of the dense and high vegetation as well as the presence of shades (due to differences in tree heights), improving the Forest delineation. This resulted in the highest F1-score for this class using RGB+LSMM dataset, 0.02 higher than the F1-score achieved using only RGB bands. The extraction of LSMM fractions from high spatial resolution images is still worth it, since the mixture of vegetation targets is still present in 2×2 meter pixels (NICHOL; WONG, 2007; SUN et al., 2021). In a pixel of Woodland Savanna, for instance, the proportions of vegetation and shade fractions are mainly higher than in a pixel of Shrub Savanna. These fractions are widely used to classify vegetation in Cerrado (GIROLAMO-NETO, 2018; FERREIRA et al., 2007; ALENCAR et

al., 2020; MIURA et al., 2003; AMARAL et al., 2015) with traditional Machine Learning techniques. In the Deep Learning field, new methodologies have been tested to generate the LSMM fractions as results of some convolutional neural network (ZHANG et al., 2018b), but, as far as we know, they were never tested as input layers in semantic segmentation using Deep Learning approach.

5.2.2 Samples generation

Regarding the generation of samples for the network training and validation, we employed two procedures. The first, used in the analysis of spectral input data, split the image in three parts, two of which generated the training and validation samples and the last one was used to test the network. In the second procedure, used in all steps of the hierarchical classification, the samples were generated based on superpixel centroids. The first procedure is commonly used, since it is able to detect patterns of the same class in different regions of the image (ADARME et al., 2020; PAN et al., 2020; ANDRADE et al., 2020). However, in a case with several classes with different occurrences across the study site, the second procedure may be more appropriate to make sure that the classification represents all classes and their spatial context patterns. In this study, the OA with the first procedure was of 89.3%, raising to 92.8% in the second procedure. Both of them for the first level of classification and using the RGB+EVI2 dataset.

Fu et al. (2018) used an approach similar to the second procedure. They generated the samples using object centroids, but with a multiresolution segmentation algorithm to classify general classes (e.g., water, road and building). The accuracies of this approach with a CNN were 5% to 10% higher than the accuracies achieved with a GEOBIA approach. Superpixel segmentation algorithms, such as the SLIC, creates more uniform objects, while the traditional segmentation algorithms (e.g., multiresolution segmentation) generates objects with very different sizes and shapes. Using this last case to create the network samples could potentially generate patches with very irregular proportions of the classes due to the centroids position (CHEN et al., 2019). In segments with irregular shapes, the centroids are not always located inside them. A supposed elongated and curved segment of Gallery Forest, for example, could generate a centroid outside that class and fail to represent that pattern.

5.2.3 Hierarchical classification

The high OAs achieved in the results are mainly related to two aspects: the Deep Learning methodology and the high spatial resolution of the WorldView-2 imagery.

Using coarser spatial resolution and the same three classes of the first level, previous works (INPE, 2015; BENDINI et al., 2020; ALENCAR et al., 2020) achieved lower OAs using traditional Machine Learning techniques. Although these works intended to identify only three classes, it is valid to notice that each class has a high intraclass variability. Therefore, Grassland class, for instance, contains the pattern (i.e., spectral behavior) of many different types of grasslands (four, in the BNP). Still using coarser spatial resolution (30m), Ferreira et al. (2007) and Schwieder et al. (2016) improved the detail of vegetation classes. The last one had Recalls of 86% and 64% for Gallery Forest and Shrub Grassland, respectively, while we achieved 95% and 83%.

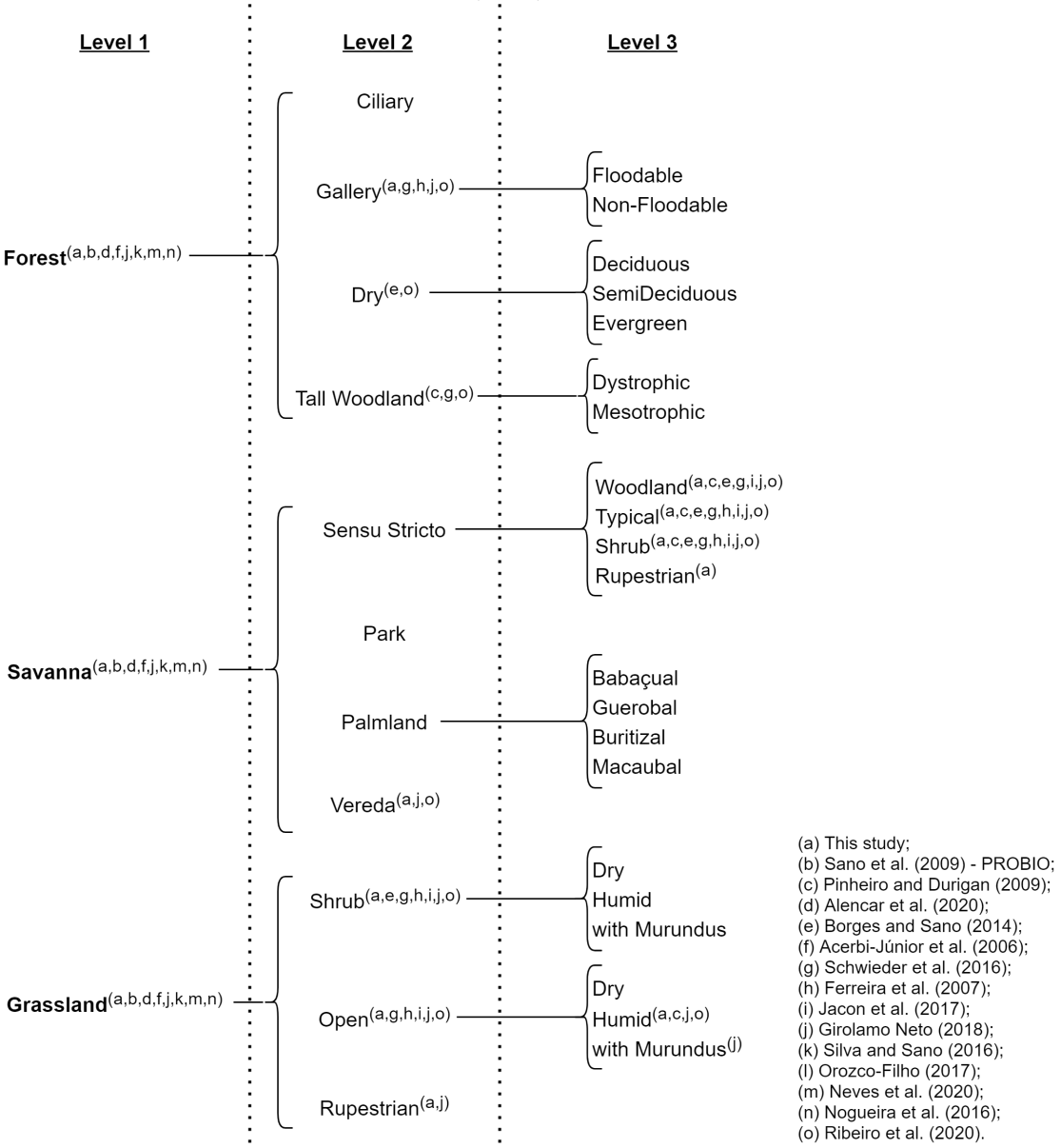
In Ferreira et al. (2007), the classes Shrub Cerrado and Woodland Cerrado are equivalent to Shrub Grassland and Shrub Savanna classes, respectively. While they correctly classified 75% of both, we presented Recalls of 83% and 81% for Shrub Grassland and Shrub Savanna. Thus, probably only the handcrafted features, used by the traditional Machine Learning algorithms, are not enough to acquire all the information needed about the vegetation classes. Improving the spatial resolution of the input imagery is also not enough to solve the problem of classifying the Cerrado vegetation. Keeping the Random Forest algorithm and switching to the same spatial resolution used in this work (2m), Girolamo-Neto (2018) and Neves et al. (2019) obtained 88.9% and 88.2%, respectively, when differentiating the classes Forest, Savanna and Grassland. When dealing with the physiognomies, they achieved Recalls of 39.15% and 39.25% for Shrub Savanna and 63.32% and 72.51% for Shrub Grassland, respectively, while this study achieved 81% and 83% for these two classes.

In Figure 5.1, the three hierarchical levels of the Ribeiro and Walter (2008) classification system are presented, starting with the three major groups of ecosystems and detailing up to 25 physiognomies. In the bottom right corner of this figure, there are the majority of studies described in Section 2.3. Only few studies that used classes not compatible with the Ribeiro and Walter (2008) classification system were not included in the figure. For each physiognomy, the studies that classified it are identified. Therefore, it is possible to notice that most studies tackled the differentiation of the first level classes.

In this study, represented by letter *a* in Figure 5.1, it was possible to identify physiognomies that are not commonly classified, such as *Vereda*, Rupestrian Grassland and Rupestrian Savanna. This last one was not classified in any other study presented. Some physiognomies (e.g., Dry Forest and Tall Woodland) were classified

by other studies, such as in Ribeiro et al. (2020), and were not identified in this study. This occurred because these physiognomies do not occur in the study site. However, this is not an indicative that the proposed methodology can not identify these patterns. The inclusion of other Cerrado regions would be required to test the inclusion of different physiognomies.

Figure 5.1 - Classes identified by this study and by others according to the levels of classes in the Ribeiro and Walter (2008) classification system.



SOURCE: Author's production.

Each detailed physiognomy has its own floristic composition, vegetation density and edaphic factors. The Woodland Savanna, due to its dense vegetation, is often confused with Forests, while the Shrub Savanna, with a more sparse vegetation, is confused with Shrub Grassland, a Grassland physiognomy (JACON et al., 2017; GIROLAMO-NETO, 2018; NEVES et al., 2019). The confusion between Forest and Grasslands is rare. Although such a confusion is surprising, it may be related to the presence of Humid Open Grassland areas in the BNP. This physiognomy is located predominantly close to the Gallery Forests (RIBEIRO; WALTER, 2008) and, consequently, the misclassified areas occur at the boundaries between these two classes.

The physiognomies, according to the classification system used in this thesis (RIBEIRO; WALTER, 2008), present an increasing scale of density and, consequently, biomass. Under these circumstances, the most common errors occur in transition areas between the physiognomies. Although this error is more intense when lower spatial resolutions are used, this is an issue in every mapping of Savanna physiognomies (FERREIRA et al., 2007; JACON et al., 2017; SCHWIEDER et al., 2016). The majority of works that classified detailed physiognomies using traditional Machine Learning techniques performed the validation using independent random points (GIROLAMO-NETO, 2018; NEVES et al., 2019; MIURA et al., 2003; JACON et al., 2017). As a semantic segmentation approach was performed in this study, the validation samples (as well as the training samples) were independent patches (160×160 pixels) entirely classified. Thus, it generates a more robust evaluation of the delineation of the physiognomies and, consequently, the misclassification in transition areas.

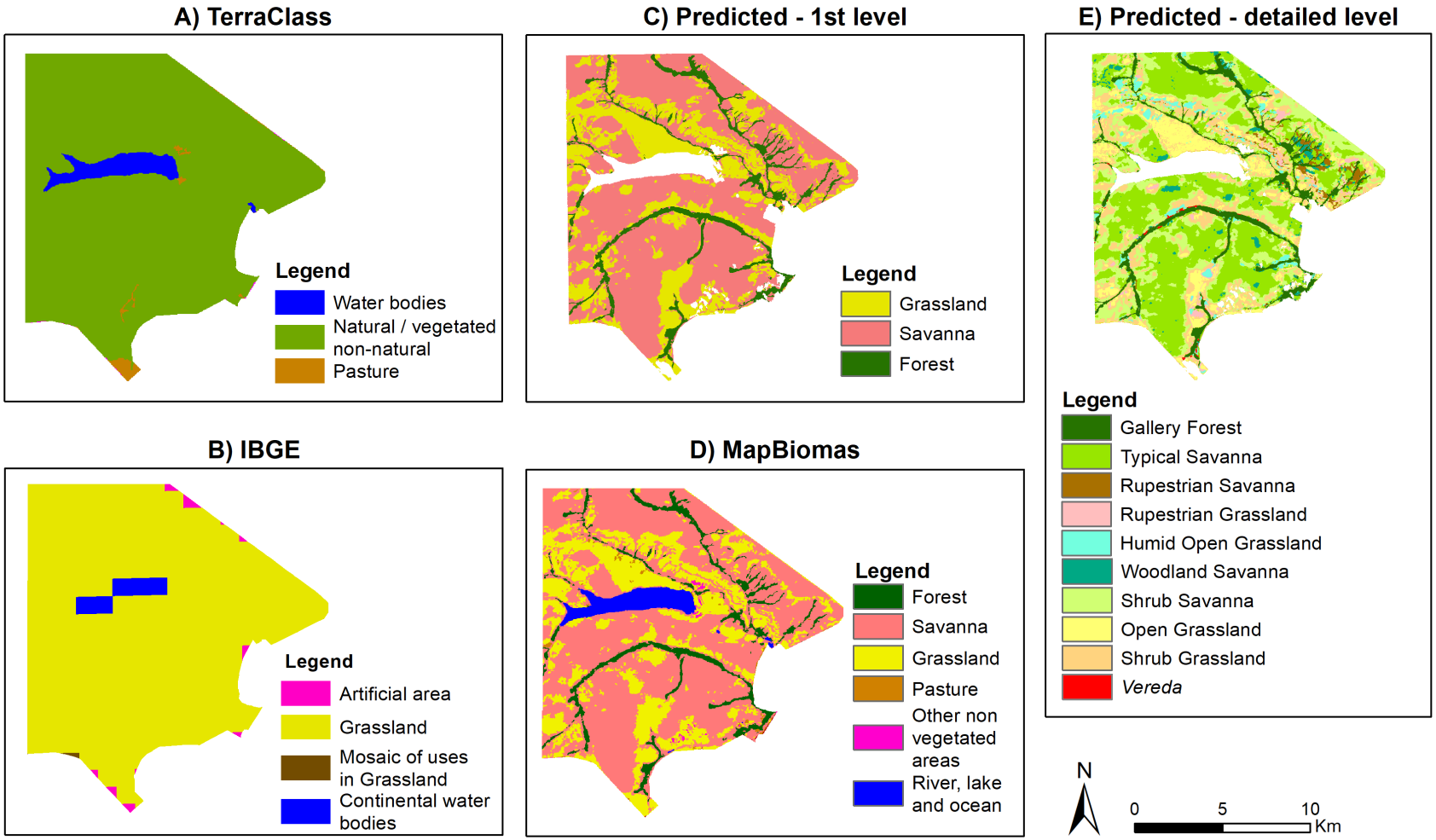
The use of a hierarchical classification approach intended to minimize the confusion between Savanna and Grassland physiognomies or between any of them with Forest patterns. Despite considering the misclassifications of the first level, this approach was efficient to map the detailed physiognomies, since it achieved higher accuracy rates than other works that intended to perform a similar task without using hierarchy (FERREIRA et al., 2007; SCHWIEDER et al., 2016; JACON et al., 2017; GIROLAMO-NETO, 2018). Compared to the other few works that also used hierarchical approaches (NEVES et al., 2019; RIBEIRO et al., 2020), this study presented superior accuracy rates. Thus, we demonstrated the potential of applying Deep Learning techniques to the RS field open problems, such as the classification of targets not yet well delineated with traditional Machine Learning algorithms.

In Figure 5.2, it is possible to evaluate qualitatively the resulting maps (Figure

5.2C and E) with other well-established mappings in the BNP, created by INPE (Figure 5.2A), IBGE (Figure 5.2B) and MapBiomas (Figure 5.2D). IBGE's map has a spatial resolution of 1km and, despite having classified almost the entire area as Grassland, its legend includes several Savanna physiognomies in the Grassland class (IBGE, 2018). TerraClass and MapBiomas maps have 30 meters of spatial resolution and classified the area as natural vegetation or as Forest, Savanna and Grassland, respectively. This comparison clearly shows how the improvements in the spatial resolution and in the methodology can bring refinements in the classes detailing and, consequently, in other products derived from vegetation maps (e.g., biodiversity studies).

It is important to note, though, that in many applications, specially in RS field, one of the main limitations of Deep Learning techniques is the requirement of a robust amount of samples to achieve acceptable results (MA et al., 2019). The amount of samples used in this work (see Sections 3.4.2 and 3.4.3 for more information) was enough to differentiate the physiognomies in the BNP. Despite covering a proportionally small area of Cerrado biome, the BNP is a Preserved Area representative of Cerrado ecossystems, and contains the major physiognomies of the biome (FERREIRA et al., 2007). However, due to the physiognomies heterogeneity across Cerrado biome (SANO et al., 2019), the application of this methodology in the entire biome would require the inclusion of more samples during training phase.

Figure 5.2 - Cerrado vegetation classification in BNP performed by: A) TerraClass Cerrado - 2013; B) IBGE - 2014; C) This study (first level of classes); D) MapBiomas - 2014; and E) This study (detailed level of classes).



SOURCE: Author's production.

6 CONCLUSIONS

This study proposed and evaluated a new methodology based on Deep Learning techniques to hierarchically classify the Cerrado physiognomies, according to the Ribeiro and Walter (2008) classification system. Regarding the hypotheses presented, some considerations can be made: (1) the use of Deep Learning techniques enabled the creation of maps with better detailing and higher accuracies than those other techniques such as GEOBIA and machine learning; (2) although it does not completely prevent the misclassification, the use of a hierarchical approach reduced the confusion between detailed classes, in Deep Learning and in GEOBIA approaches; and (3) testing several datasets as input in the networks showed that, in this study, the best dataset was composed by RGB bands plus EVI2.

Each Cerrado physiognomy has a different amount of biomass above and below the ground and a unique biodiversity. Consequently, the proper identification and delineation of the Cerrado physiognomies are fundamental to truly understand the role of Savanna biomes in the global Carbon cycle. Additionally, Cerrado vegetation maps can be used as a proxy in other studies, such as those related to biodiversity, since a high rate of endemism can be found in its physiognomies.

The greatest limitation when mapping the physiognomies of Cerrado, the second largest biome of a continental-size country like Brazil, is the lack of reference data. The difficulty in differentiating the patterns of physiognomies associated to the biome extension results in few options of vegetation maps, usually available only for three classes (Forest, Savanna and Grassland) and with a spatial resolution of around 30 meters. This limitation is even bigger when dealing with a semantic segmentation approach, since reference points are not enough and entirely classified patches are required as samples for the network training.

Under these circumstances, the proposed methodology to hierarchically classify a relevant Protected Area, the Brasília National Park, using Deep Learning was quite efficient to delineate the three major groups of physiognomies, in a first level, and ten detailed physiognomies in the second level. To attest the superiority of the Deep Learning approach, a GEOBIA approach was performed in the same study site using the same high spatial resolution image. The GEOBIA experiments were fundamental to demonstrate that using a hierarchical structure of classes can improve the accuracies. Besides that, they have also shown that some features are more valuable in the classification than others. The VIs were fundamental to differentiate the Grassland physiognomies, while the texture features proved to be more relevant for the first

level of classification and for the differentiation of the Savanna physiognomies.

Nevertheless, as GEOBIA and Deep Learning approaches work in very different ways, the evaluation of different input datasets was also performed using the Deep Learning approach. The highest accuracies were achieved by the dataset composed by RGB bands plus the EVI2, contradicting the idea that handcrafted features are irrelevant to Deep Learning networks. Although this is a well accepted idea in Deep Learning field, it was established for computer vision experiments and, when it comes to the RS field, more tests are needed. For the first level of classification, two different procedures were used to generate the training/validation samples. The random generation, which includes the use of superpixels as the starting point of the process, proved to be the best method. It not only achieved higher accuracy, but also created more balanced samples regarding the vegetation classes.

The results for the first and second levels of classification achieved superior OAs when compared to other works that performed similar task using other Machine Learning algorithms, GEOBIA approaches, image time series or visual interpretation. The developed method proved to be capable of delineating the physiognomies. As the Cerrado contains several particularities across the biome, the reproduction of the method for the entire biome would require the availability of high spatial resolution images and the production of reference data in the same spatial resolution to generate more samples for training and validation.

6.1 Suggestions for future works

Based on the results achieved in this thesis, some suggestions can be made for future works. First, the inclusion of additional satellite data to consider other aspects of the physiognomies may help in the mapping. A good example of that is the use of satellite image time series to contemplate the physiognomies seasonality in the analysis. However, the use of time series (without considering the spatial context) with well-known techniques have been done already and may not be enough. For this reason, it is also suggested to keep the high spatial resolution and to apply Deep Learning architectures that are appropriate for time series data, such as the Long Short Term Memory (LSTM) networks. The inclusion of active remote sensing data, such as Radar and LiDAR (Light Detection And Ranging) data, can also provide additional information, especially of the vegetation structure, to differentiate the physiognomies patterns.

REFERENCES

ABADE, N. A.; CARVALHO-JÚNIOR, O. A.; GUIMARÃES, R. F.; OLIVEIRA, S. N. D. Comparative analysis of modis time-series classification using support vector machines and methods based upon distance and similarity measures in the brazilian cerrado-caatinga boundary. **Remote Sensing**, v. 7, n. 9, p. 12160–12191, 2015. 26, 31

ABADI, M. et al. Tensorflow: large-scale machine learning on heterogeneous systems. **arXiv preprint arXiv:1603.04467**, 2015. 48

ACERBI-JUNIOR, F.; CLEVERS, J.; SCHAEPMAN, M. E. The assessment of multi-sensor image fusion using wavelet transforms for mapping the brazilian savanna. **International Journal of Applied Earth Observation and Geoinformation**, v. 8, n. 4, p. 278–288, 2006. 26, 27, 31

ACHANTA, R.; SHAJI, A.; SMITH, K.; LUCCHI, A.; FUA, P.; SÜSSTRUNK, S. Slic superpixels compared to state-of-the-art superpixel methods. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 34, n. 11, p. 2274–2282, 2012. 39

ADAMI, M. **Estimativa da data de plantio da soja por meio de séries temporais de imagens MODIS**. 2010. 163 p. Tese (Doutorado em Sensoriamento Remoto) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2010. 15

ADARME, M. O.; FEITOSA, R. Q.; HAPP, P. N.; ALMEIDA, C. A.; GOMES, A. R. Evaluation of deep learning techniques for deforestation detection in the brazilian Amazon and cerrado biomes from remote sensing imagery. **Remote Sensing**, v. 12, n. 6, p. 1–29, 2020. 19, 77

ADUAN, R. E.; VILELA, M. d. F.; KLINK, C. A. **Ciclagem de carbono em ecossistemas terrestres: o caso do cerrado brasileiro.** Planaltina, DF: Embrapa Cerrados., 2003. 1

ALENCAR, A.; SHIMBO, J. Z.; LENTI, F.; MARQUES, C. B.; ZIMBRES, B.; ROSA, M.; ARRUDA, V.; CASTRO, I.; RIBEIRO, J. P. F. M.; VARELA, V.; ALENCAR, I.; PIONTEKOWSKI, V.; RIBEIRO, V.; BUSTAMANTE, M. M. C.; SANO, E. E.; BARROSO, M. Mapping three decades of changes in the brazilian savanna native vegetation using landsat data processed in the google earth engine platform. **Remote Sensing**, v. 12, n. 6, p. 1–23, 2020. 25, 26, 31, 76, 77, 78

AMARAL, C. H.; ROBERTS, D. A.; ALMEIDA, T. I.; SOUZA-FILHO, C. R. Mapping invasive species and spectral mixture relationships with neotropical woody formations in southeastern Brazil. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 108, p. 80–93, 2015. 76, 77

ANDRADE, R.; COSTA, G.; MOTA, G.; ORTEGA, M.; FEITOSA, R.; SOTO, P.; HEIPKE, C. Evaluation of semantic segmentation methods for deforestation detection in the Amazon. **The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 43, p. 1497–1505, 2020. 77

BAIG, M. H. A.; ZHANG, L.; SHUAI, T.; TONG, Q. Derivation of a tasselled cap transformation based on landsat 8 at-satellite reflectance. **Remote Sensing Letters**, v. 5, n. 5, p. 423–431, 2014. 17

BALL, J. E.; ANDERSON, D. T.; CHEE, S. C. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. **Journal of Applied Remote Sensing**, v. 11, n. 4, p. 1–54, 2017. 14

BENDINI, H.; FONSECA, L.; SCHWIEDER, M.; RUFIN, P.; KORTING, T.; KOUMROUYAN, A.; HOSTERT, P. Combining environmental and landsat analysis ready data for vegetation mapping: a case study in the brazilian savanna biome. **The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 43, p. 953–960, 2020. 78

BLASCHKE, T. Object based image analysis for remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 65, n. 1, p. 2–16, 2010. 4, 14, 15

BORGES, E. F.; SANO, E. E. Séries temporais de evi do modis para o mapeamento de uso e cobertura vegetal do oeste da Bahia. **Boletim de Ciências Geodésicas**, v. 20, n. 3, p. 526–547, 2014. 26, 27, 31

BRANDT, M. et al. An unexpectedly large count of trees in the west african sahara and sahel. **Nature**, v. 587, n. 7832, p. 78–82, 2020. 30

BRASIL. MINISTÉRIO DO MEIO AMBIENTE - MMA. **Cadastro Nacional de Unidades de Conservação - CNUC**. 2010. Available from: <https://www.mma.gov.br/areas-protegidas/cadastro-nacional-de-ucs>. 1

_____. **PPCerrado – Plano de Ação para Prevenção e Controle do Desmatamento e das Queimadas no Cerrado**. 2014. Available from:

<http://combateaodesmatamento.mma.gov.br/images/conteudo/PPCerrado_2aFase.pdf>. 3

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. 18, 19, 46

CHEN, G.; WENG, Q.; HAY, G. J.; HE, Y. Geographic object-based image analysis (geobia): emerging trends and future opportunities. **GIScience & Remote Sensing**, v. 55, n. 2, p. 159–182, 2018. 14, 15

CHEN, Y.; MING, D.; LV, X. Superpixel based land cover classification of vhr satellite image combining multi-scale cnn and scale parameter estimation. **Earth Science Informatics**, v. 12, n. 3, p. 341–363, 2019. 77

CHOLLET, F. et al. **Keras**. 2015. Available from: https://keras.io. 48

ÇIğLA, C.; ALATAN, A. A. Efficient graph-based image segmentation via speeded-up turbo pixels. In: INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 2010. **Proceedings...** Hong Kong: IEEE, 2010. p. 3013–3016. 38

COSTA, W.; FONSECA, L.; KÖRTING, T. Classifying grasslands and cultivated pastures in the brazilian cerrado using support vector machines, multilayer perceptrons and autoencoders. In: INTERNATIONAL WORKSHOP ON MACHINE LEARNING AND DATA MINING IN PATTERN RECOGNITION, 2015. **Proceedings...** Hamburg: Springer, 2015. p. 187–198. 26, 32

CRIST, E.; KAUTH, R. The tasseled cap de-mystified. **Photogrammetric Engineering and Remote Sensing**, v. 52, n. 1, p. 81–86, 1986. 16, 17

DIGITALGLOBE. **The benefits of the eight spectral bands of WorldView-2**. 2010. Available from: https://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/35/DG-8SPECTRAL-WP_0.pdf. 38, 45

_____. **Information products - standard imagery**. 2021. Available from: https://earth.esa.int/eogateway/documents/20142/37627/DigitalGlobe-Standard-Imagery.pdf. 37

EXELIS VISUAL INFORMATION SOLUTIONS, INC. **ENVI classic tutorial: atmospherically correcting multispectral data using FLAASH**. 2013. Available from: https://www.l3harrisgeospatial.com/portals/0/pdfs/envi/FLAASH_Multispectral.pdf. 37

FERREIRA, L.; YOSHIOKA, H.; HUETE, A.; SANO, E. On the use of the eos–modis vegetation indices for monitoring the cerrado region, Brazil: insights and perspectives. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 10, 2001, São José dos Campos. **Anais...** Foz do Iguaçu: INPE, 2001. p. 20–26. 35

FERREIRA, L. G.; ASNER, G. P.; KNAPP, D. E.; DAVIDSON, E. A.; COE, M.; BUSTAMANTE, M. M.; OLIVEIRA, E. L. de. Equivalent water thickness in savanna ecosystems: modis estimates based on ground and eo-1 hyperion data. **International Journal of Remote Sensing**, v. 32, n. 22, p. 7423–7440, 2011. 1

FERREIRA, M.; FERREIRA, L.; SANO, E.; SHIMABUKURO, Y. Spectral linear mixture modelling approaches for land cover mapping of tropical savanna areas in Brazil. **International Journal of Remote Sensing**, v. 28, n. 2, p. 413–429, 2007. 3, 7, 28, 32, 35, 76, 77, 78, 80, 81

FLOOD, N.; WATSON, F.; COLLETT, L. Using a u-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across Queensland, Australia. **International Journal of Applied Earth Observation and Geoinformation**, v. 82, p. 1–15, 2019. 23

FRANÇOSO, R. D.; BRANDÃO, R.; NOGUEIRA, C. C.; SALMONA, Y. B.; MACHADO, R. B.; COLLI, G. R. Habitat loss and the effectiveness of protected areas in the cerrado biodiversity hotspot. **Natureza & Conservação**, v. 13, n. 1, p. 35–40, 2015. 1

FU, T.; MA, L.; LI, M.; JOHNSON, B. A. Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery. **Journal of Applied Remote Sensing**, v. 12, n. 2, p. 1–21, 2018. 77

GIROLAMO-NETO, C. D. **Identificação de fitofisionomias de Cerrado no Parque Nacional de Brasília utilizando random forest aplicado a imagens de alta e média resoluções espaciais**. 186 p. Tese (Doutorado em Sensoriamento Remoto — National Institute for Space Research (INPE), São José dos Campos, 2018. Available from: <http://urlib.net/rep/8JMKD3MGP3W34R/3RU6Q68>. 4, 29, 32, 35, 39, 75, 76, 77, 78, 80

GIRSHICK, R.; DONAHUE, J.; DARRELL, T.; MALIK, J. Region-based convolutional networks for accurate object detection and segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 38, n. 1, p. 142–158, 2015. 19

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.; BENGIO, Y. **Deep learning**. Cambridge: MIT Press, 2016. 19, 21

GRACE, J.; JOSÉ, J. S.; MEIR, P.; MIRANDA, H. S.; MONTES, R. A. Productivity and carbon fluxes of tropical savannas. **Journal of Biogeography**, v. 33, n. 3, p. 387–400, 2006. 1

GRECCHI, R. C.; GWYN, Q. H. J.; BÉNIÉ, G. B.; FORMAGGIO, A. R. Assessing the spatio-temporal rates and patterns of land-use and land-cover changes in the cerrados of southeastern mato grosso, brazil. **International Journal of Remote Sensing**, v. 34, n. 15, p. 5369–5392, 2013. 4, 27, 29, 32

GUIRADO, E.; TABIK, S.; ALCARAZ-SEGURA, D.; CABELLO, J.; HERRERA, F. Deep-learning versus obia for scattered shrub detection with google earth imagery: ziziphus lotus as case study. **Remote Sensing**, v. 9, n. 12, p. 1–22, 2017. 4, 5, 19, 75

GUO, Y.; LIU, Y.; BAKKER, E. M.; GUO, Y.; LEW, M. S. Cnn-rnn: a large-scale hierarchical image classification framework. **Multimedia Tools and Applications**, v. 77, n. 8, p. 10251–10271, 2018. 5

GWENZI, D.; LEFSKY, M. A. Modeling canopy height in a savanna ecosystem using spaceborne lidar waveforms. **Remote Sensing of Environment**, v. 154, p. 338–344, 2014. 1

HALL-BEYER, M. GLCM texture: a tutorial v. 3.0 march 2017. University of Calgary. Available at `https://prism.ucalgary.ca/handle/1880/51900`, 2017. 17, 18, 44

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The weka data mining software: an update. **ACM SIGKDD Explorations Newsletter**, v. 11, n. 1, p. 10–18, 2009. 46

HARALICK, R. M.; SHANMUGAM, K.; DINSTEIN, I. Textural features for image classification. **IEEE Transactions on Systems, Man, and Cybernetics**, n. 6, p. 610–621, 1973. 17, 44

HAY, G. J.; CASTILLA, G. Geographic object-based image analysis (geobia): a new name for a new discipline. In: BLASCHKE, T.; LANG, S.; HAY, G. (Ed.). **Object-based image analysis**. Berlim: Springer, 2008. p. 75–89. 14, 18

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the modis vegetation indices. **Remote Sensing of Environment**, v. 83, n. 1-2, p. 195–213, 2002. 15

HUETE, A.; LIU, H.; BATCHILY, K.; LEEUWEN, W. V. A comparison of vegetation indices over a global set of tm images for eos-modis. **Remote Sensing of Environment**, v. 59, n. 3, p. 440–451, 1997. 42

HUETE, A. R. A soil-adjusted vegetation index (savi). **Remote Sensing of Environment**, v. 25, n. 3, p. 295–309, 1988. 42

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Manual técnico da vegetação brasileira**. 1992. Available from: <https://biblioteca.ibge.gov.br/visualizacao/monografias/GEBIS%20-%20RJ/ManuaisdeGeociencias/Manual%20Tecnico%20da%20Vegetacao%20Brasileira%20n.1.pdf>. Access in: 21 ago. 2020. 7

_____. _____. 2012. Available from: <https://www.terrabrasilis.org.br/ecotecadigital/pdf/manual-tecnico-da-vegetacao-brasileira.pdf>. Access in: 21 ago. 2020. 7, 25

_____. **Monitoramento da cobertura e uso da terra do Brasil 2014 - 2016**. 2018. Available from: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101625.pdf>. Access in: 11 jan. 2021. 81

_____. **Biomas e sistema costeiro-marinho do Brasil: compatível com a escala 1:250 000**. 2019. Available from: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101676.pdf>. Access in: 11 ago. 2020. 1

INSTITUTO CHICO MENDES DE CONSERVAÇÃO DA BIODIVERSIDADE - ICMBIO. **Plano de manejo do Parque Nacional de Brasília**. 1998. Available from: <http://www.icmbio.gov.br/portal/images/stories/imgs-unidades-coservacao/PARNA%20Brasilia.pdf>. 35

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **Projeto TerraClass Cerrado - mapeamento do uso e cobertura vegetal do Cerrado**. 2015. Available from: <http://www.dpi.inpe.br/tccerrado/Metodologia_TCCerrado_2013.pdf>. 1, 2, 25, 32, 78

_____. **Annual deforestation in Brazilian Savannah**. INPE - National Institute for Space Research, 2020. Available from: <http://www.obt.inpe.br/cerrado>. 1

JACON, A. D.; GALVÃO, L. S.; SANTOS, J. R. dos; SANO, E. E. Seasonal characterization and discrimination of savannah physiognomies in brazil using hyperspectral metrics from hyperion/eo-1. **International Journal of Remote Sensing**, v. 38, n. 15, p. 4494–4516, 2017. 3, 4, 7, 28, 33, 80

JARDIM BOTÂNICO DO RIO DE JANEIRO - JBRJ. **Reflora: lista de espécies da flora do Brasil**. 2015. Available from: <http://floradobrasil.jbrj.gov.br/>. 1

JIANG, Z.; HUETE, A. R.; DIDAN, K.; MIURA, T. Development of a two-band enhanced vegetation index without a blue band. **Remote Sensing of Environment**, v. 112, n. 10, p. 3833–3845, 2008. 42

JOHANSEN, K.; ARROYO, L. A.; PHINN, S.; WITTE, C. Comparison of geo-object based and pixel-based change detection of riparian environments using high spatial resolution multi-spectral imagery. **Photogrammetric Engineering & Remote Sensing**, v. 76, n. 2, p. 123–136, 2010. 15

JOZDANI, S. E.; JOHNSON, B. A.; CHEN, D. Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification. **Remote Sensing**, v. 11, n. 14, p. 1–24, 2019. 5, 75

KATTENBORN, T.; EICHEL, J.; FASSNACHT, F. E. Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution uav imagery. **Scientific Reports**, v. 9, n. 1, p. 1–9, 2019. 5, 75

KAUTH, R. J.; THOMAS, G. The tasselled cap – a graphic description of the spectral-temporal development of agricultural crops as seen by landsat. In: LARS SYMPOSIA, 1976. **Proceedings...** West Lafayette, 1976. p. 41–51. 16

KERKECH, M.; HAFIANE, A.; CANALS, R. Deep leaning approach with colorimetric spaces and vegetation indices for vine diseases detection in uav images. **Computers and Electronics in Agriculture**, v. 155, p. 237–243, 2018. 76

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Advances in Neural Information Processing Systems**, v. 25, n. 2, p. 1097–1105, 2012. 23

KUMAR, S. **Deep U-net for satellite image segmentation**. GitHub, 2018. Available from: <https://github.com/reachsumit/deep-unet-for-satellite-image-segmentation/>. 46, 48

KUSSUL, N.; LAVRENIUK, M.; SKAKUN, S.; SHELESTOV, A. Deep learning classification of land cover and crop types using remote sensing data. **IEEE Geoscience and Remote Sensing Letters**, v. 14, n. 5, p. 778–782, 2017. 4

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. 4, 19, 20, 21

LECUN, Y.; BOSER, B. E.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W. E.; JACKEL, L. D. Handwritten digit recognition with a back-propagation network. In: TOURETZKY, D. S. (Ed.). **Advances in neural information processing systems**. San Francisco: Morgan Kaufmann, 1990. p. 396–404. 23

LI, F. F. **CS231n: convolutional neural networks for visual recognition. Course notes**. 2018. Available from: <https://cs231n.github.io/convolutional-networks/>. 22

LITJENS, G.; KOOI, T.; BEJNORDI, B. E.; SETIO, A. A. A.; CIOMPI, F.; GHAFOORIAN, M.; LAAK, J. A. V. D.; GINNEKEN, B. V.; SÁNCHEZ, C. I. A survey on deep learning in medical image analysis. **Medical Image Analysis**, v. 42, p. 60–88, 2017. 19

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2015. **Proceedings...** Boston: IEEE, 2015. p. 3431–3440. 23

MA, L.; LI, M.; MA, X.; CHENG, L.; DU, P.; LIU, Y. A review of supervised object-based land-cover image classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 130, p. 277–293, 2017. 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 Journal of Photogrammetry and Remote Sensing**, v. 152, p. 166–177, 2019. 4, 75, 81

MACQUEEN, J. et al. Some methods for classification and analysis of multivariate observations. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 1967. **Proceedings...** Oakland, 1967. v. 14, n. 14, p. 281–297. 38

MIURA, T.; HUETE, A. R.; FERREIRA, L. G.; SANO, E. E. Discrimination and biophysical characterization of cerrado physiognomies with eo-1 hyperspectral hyperion. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 11, 2003. **Anais...** São José dos Campos: INPE, 2003. p. 1077–1082. 76, 77, 80

MÜLLER, H.; RUFIN, P.; GRIFFITHS, P.; SIQUEIRA, A. J. B.; HOSTERT, P. Mining dense landsat time series for separating cropland and pasture in a heterogeneous brazilian savanna landscape. **Remote Sensing of Environment**, v. 156, p. 490–499, 2015. 27, 33

NEVES, A.; KÖRTING, T.; FONSECA, L.; GIROLAMO-NETO, C.; WITTICH, D.; COSTA, G.; HEIPKE, C. Semantic segmentation of brazilian savanna vegetation using high spatial resolution satellite data and u-net. **ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences**, v. 5, n. 3, 2020. 14, 19, 23, 30, 33, 57

NEVES, A. K.; KÖRTING, T. S.; GIROLAMO-NETO, C. D.; SOARES, A. R.; FONSECA, L. M. G. Hierarchical classification of brazilian savanna physiognomies using very high spatial resolution image, superpixel and geobia. In: IGARSS INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2019. **Proceedings...** Yokohama: IEEE, 2019. p. 3716–3719. 29, 33, 57, 78, 80

NICHOL, J.; WONG, M. S. Remote sensing of urban vegetation life form by spectral mixture analysis of high-resolution ikonos satellite images. **International Journal of Remote Sensing**, v. 28, n. 5, p. 985–1000, 2007. 76

NOGUEIRA, K.; SANTOS, J. A. D.; FORNAZARI, T.; SILVA, T. S. F.; MORELLATO, L. P.; TORRES, R. d. S. Towards vegetation species discrimination by using data-driven descriptors. In: IAPR WORKSHOP ON PATTERN RECOGNITON IN REMOTE SENSING, 9, 2016. **Proceedings...** Cancun: IEEE, 2016. p. 1–6. 5, 14, 19, 30, 33, 75

OROZCO-FILHO, J. C. **Avaliação do uso da abordagem orientada-objeto com imagens de alta resolução RapidEye na classificação das fitofisionomias do Cerrado**. 2017. 44 p. Dissertação (Mestrado em Geografia)

— Instituto de Ciências Humanas, Universidade de Brasília (UNB), Brasília - DF, 2017. 4, 29, 33

OSHIRO, T. M. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. 2013. 90 p. Dissertação (Mestrado em Bioinformática) — Universidade de São Paulo (USP), São Paulo, 2013. 18

PAN, Z.; XU, J.; GUO, Y.; HU, Y.; WANG, G. Deep learning segmentation and classification for urban village using a worldview satellite image based on u-net. **Remote Sensing**, v. 12, n. 10, p. 1–17, 2020. 77

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011. 38

PERKINS, T.; ADLER-GOLDEN, S.; MATTHEW, M.; BERK, A.; ANDERSON, G.; GARDNER, J.; FELDE, G. Retrieval of atmospheric properties from hyper and multispectral imagery with the flaash atmospheric correction algorithm. In: REMOTE SENSING OF CLOUDS AND THE ATMOSPHERE, 10, 2005. **Proceedings...** Bruges: Spie, 2005. v. 5979. 37

PINHEIRO, E. D. S.; DURIGAN, G. Dinâmica espaço-temporal (1962-2006) das fitofisionomias em unidade de conservação do cerrado no sudeste do brasil. **Brazilian Journal of Botany**, v. 32, n. 3, p. 441–454, 2009. 25, 34

PONTI, M. A.; RIBEIRO, L. S. F.; NAZARE, T. S.; BUI, T.; COLLOMOSSE, J. Everything you wanted to know about deep learning for computer vision but were afraid to ask. In: SIBGRAPI CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES TUTORIALS,30, 2017. **Proceedings...** Niterói: IEEE, 2017. p. 17–41. 20, 21, 23

PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento remoto no estudo da vegetação**. São José dos Campos: Parêntese, 2010. 15, 42

QI, J.; CHEHBOUNI, A.; HUETE, A. R.; KERR, Y. H.; SOROOSHIAN, S. A modified soil adjusted vegetation index. **Remote Sensing of Environment**, v. 48, n. 2, p. 119–126, 1994. 42

RESENDE, F. M.; CIMON-MORIN, J.; POULIN, M.; MEYER, L.; LOYOLA, R. Consequences of delaying actions for safeguarding ecosystem services in the brazilian cerrado. **Biological Conservation**, v. 234, p. 90–99, 2019. 1

RIBEIRO, F. F.; ROBERTS, D. A.; HESS, L. L.; DAVIS, F. W.; CAYLOR, K. K.; DALDEGAN, G. A. Geographic object-based image analysis framework for mapping vegetation physiognomic types at fine scales in neotropical savannas. **Remote Sensing**, v. 12, n. 11, p. 1–27, 2020. 7, 29, 30, 34, 75, 79, 80

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do bioma cerrado. **Cerrado: Ecologia e Flora**, v. 1, p. 151–212, 2008. xv, xvii, 3, 4, 5, 7, 9, 13, 14, 25, 27, 29, 35, 37, 70, 75, 78, 79, 80, 83

RIBEIRO, S. C.; FEHRMANN, L.; SOARES, C. P. B.; JACOVINE, L. A. G.; KLEINN, C.; GASPAR, R. de O. Above-and belowground biomass in a brazilian cerrado. **Forest Ecology and Management**, v. 262, n. 3, p. 491–499, 2011. 1, 4

ROBERTS, D. A.; BATISTA, G. T.; PEREIRA, J. L. G.; WALLER, E.; NELSON, B. W. Change identification using multitemporal spectral mixture analysis: applications in eastern Amazônia. In: LUNETTA, R. S.; ELVIDGE, C. (Ed.). **Remote Sensing Change Detection: Environmental Monitoring Applications**. Chelsea: Ann Arbor Press, 1998. p. 137–161. 16

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: convolutional networks for biomedical image segmentation. In: INTERNATIONAL CONFERENCE ON MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION, 2015. **Proceedings...** Munich: Springer, 2015. p. 234–241. 23, 24, 46, 47

ROUSE, J.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the great plains with erts. **NASA Special Publication**, v. 351, n. 1974, p. 309–317, 1974. 42

SAINATH, T. N.; MOHAMED, A.-r.; KINGSBURY, B.; RAMABHADRAN, B. Deep convolutional neural networks for lvcsr. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 2013. **Proceedings...** Vancouver: IEEE, 2013. p. 8614–8618. 19

SANO, E. E.; BEZERRA, H. d. S.; LOPES, T. d. S. **Mapeamento da cobertura vegetal do bioma cerrado.** Planaltina: Embrapa Cerrados, 2009. 25, 34

SANO, E. E.; RODRIGUES, A. A.; MARTINS, E. S.; BETTIOL, G. M.; BUSTAMANTE, M. M.; BEZERRA, A. S.; COUTO-JUNIOR, A. F.; VASCONCELOS, V.; SCHÜLER, J.; BOLFE, E. L. Cerrado ecoregions: a spatial framework to assess and prioritize brazilian savanna environmental diversity for conservation. **Journal of environmental management**, v. 232, p. 818–828, 2019. 81

SCHOWENGERDT, R. A. **Remote sensing: models and methods for image processing**. 3. ed. Amsterdam: Elsevier, 2006. 14

SCHWIEDER, M.; LEITÃO, P. J.; BUSTAMANTE, M. M. da C.; FERREIRA, L. G.; RABE, A.; HOSTERT, P. Mapping brazilian savanna vegetation gradients with landsat time series. **International Journal of Applied Earth Observation and Geoinformation**, v. 52, p. 361–370, 2016. 7, 27, 28, 34, 35, 75, 78, 80

SHIMABUKURO, Y. E.; SMITH, J. A. The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 29, n. 1, p. 16–20, 1991. 16, 42

SILVA, L. R.; SANO, E. E. Análise das imagens do satélite rapideye para discriminação da cobertura vegetal do bioma cerrado. **Revista Brasileira de Cartografia**, v. 68, n. 7, p. 1269–1283, 2016. 29, 34

SOUZA, C. M. et al. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. **Remote Sensing**, v. 12, n. 17, p. 1–27, 2020. 26

STRASSBURG, B. B. et al. Moment of truth for the cerrado hotspot. **Nature Ecology & Evolution**, v. 1, n. 4, p. 1–3, 2017. 1

SUN, X.; WU, W.; LI, X.; XU, X.; LI, J. Vegetation abundance and health mapping over southwestern antarctica based on worldview-2 data and a modified spectral mixture analysis. **Remote Sensing**, v. 13, n. 2, p. 1–26, 2021. 76

THE WORLD BANK. **Development of systems to prevent forest fires and monitor vegetation cover in the Brazilian Cerrado project**. 2016. Available from: <http://fip.mma.gov.br/wp-content/uploads/2018/11/03-Projeto-FIP-FM-Project-paper-Ingl%C3%AAs.pdf>. 3

TORRES, D.; FEITOSA, R.; ROSA, L. L.; HAPP, P.; MARCATO, J.; GONÇALVES, W.; MARTINS, J.; LIESENBERG, V. Semantic segmentation of endangered tree species in brazilian savanna using deeplabv3+ variants. In: LATIN AMERICAN GRSS & ISPRS REMOTE SENSING CONFERENCE, 2020. **Proceedings...** Santiago: IEEE, 2020. p. 515–520. 30

VELOSO, H. P.; GÓES-FILHO, L. **Fitogeografia brasileira: classificação fisionômico-ecológica da vegetação neotropical**. Salvador: Projeto RADAMBRASIL, 1982. Available in: https://biblioteca.ibge.gov.br/visualizacao/livros/liv92051.pdf. 7

WAGNER, F. H.; SANCHEZ, A.; TARABALKA, Y.; LOTTE, R. G.; FERREIRA, M. P.; AIDAR, M. P.; GLOOR, E.; PHILLIPS, O. L.; ARAGAO, L. E. Using the u-net convolutional network to map forest types and disturbance in the atlantic rainforest with very high resolution images. **Remote Sensing in Ecology and Conservation**, v. 5, n. 4, p. 360–375, 2019. 23

WHITESIDE, T. G.; BOGGS, G. S.; MAIER, S. W. Comparing object-based and pixel-based classifications for mapping savannas. **International Journal of Applied Earth Observation and Geoinformation**, v. 13, n. 6, p. 884–893, 2011. 15

YANG, C.; ROTTENSTEINER, F.; HEIPKE, C. Exploring semantic relationships for hierarchical land use classification based on convolutional neural networks. **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 2, p. 599–607, 2020. 5

YARBROUGH, L. D.; NAVULUR, K.; RAVI, R. Presentation of the kauth–thomas transform for worldview-2 reflectance data. **Remote Sensing Letters**, v. 5, n. 2, p. 131–138, 2014. 17, 43

ZHANG, L.; ZHANG, L.; DU, B. Deep learning for remote sensing data: a technical tutorial on the state of the art. **IEEE Geoscience and Remote Sensing Magazine**, v. 4, n. 2, p. 22–40, 2016. 14, 20

ZHANG, P.; KE, Y.; ZHANG, Z.; WANG, M.; LI, P.; ZHANG, S. Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery. **Sensors**, v. 18, n. 11, p. 1–21, 2018. 76

ZHANG, X.; SUN, Y.; ZHANG, J.; WU, P.; JIAO, L. Hyperspectral unmixing via deep convolutional neural networks. **IEEE Geoscience and Remote Sensing Letters**, v. 15, n. 11, p. 1755–1759, 2018. 77

ZHU, X. X.; TUIA, D.; MOU, L.; XIA, G.-S.; ZHANG, L.; XU, F.; FRAUNDORFER, F. Deep learning in remote sensing: a comprehensive review and list of resources. **IEEE Geoscience and Remote Sensing Magazine**, v. 5, n. 4, p. 8–36, 2017. 76