

ANÁLISE ORIENTADA A OBJETOS GEOGRÁFICOS, ATRAVÉS DE ÁRVORE DE DECISÃO, PARA CLASSIFICAÇÃO DA COBERTURA DO SOLO, NO SEMIÁRIDO DA PARAÍBA.

Nathália de Oliveira Silva¹, Ermano Cavalcante Falcão²

^{1,2} Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB - Avenida 1º de Maio, 720 – 58015-430 - João Pessoa-PB, Brasil; nathaliaolisil@gmail.com, ermano@ifpb.edu.br

RESUMO

Este trabalho de pesquisa teve como principal objetivo classificar a cobertura do solo de uma área de proteção ambiental - APA do Cariri, localizada na região semiárida do Estado da Paraíba, utilizando-se análise de imagem baseada em objetos geográficos e técnicas de mineração de dados. Como dados de entrada, foram utilizados imagens do satélite Sentinel-2A e os índices espectrais SAVI e NDWI, processados nos *softwares* ArcGIS e TerraView, com adição do *plugin* GeoDMA. Para a mineração de dados, foram utilizados 143 atributos e 186 amostras de treinamento, empregando-se algoritmo de árvore de decisão. Os resultados indicaram que as classes de maior ocorrência são: “caatinga semidensa” e “caatinga aberta”, ocupando, respectivamente, 37,74% e 26,07%, da área da APA do Cariri. O índice de exatidão global obtido foi de 88,7%, enquanto o índice Kappa foi de 0,866.

Palavras-chave — Semiárido Paraibano, Sensoriamento Remoto, Mineração de Dados, Árvore de Decisão.

ABSTRACT

The main objective of this research was to classify the land cover of an environmental protection area - APA do Cariri, located in the semi-arid region of the State of Paraíba, using geographic object-based image analysis and data mining techniques. As input data, images of the Sentinel-2A satellite and the SAVI and NDWI spectral indexes, processed in the ArcGIS and TerraView software were used, with the addition of the GeoDMA plugin. For the data mining, 147 attributes and 186 training samples were used, using the decision tree algorithm. The results indicated that the highest occurrence classes were: "semi-dense caatinga" and "open caatinga", occupying, respectively, 37.74% and 26.07%, of the APA do Cariri area. The overall accuracy index was 88.7%, while the Kappa index was 0.866.

Key words — Semiarid Paraíba, Remote Sensing, Data Mining, Decision Tree.

1. INTRODUÇÃO

O Bioma Caatinga ocorre exclusivamente no Brasil, e ocupa 11% do território nacional, englobando todos os estados da região Nordeste e o norte de Minas Gerais (ALVES et al.) [1]. Segundo o Ministério do Meio Ambiente - MMA [2], a caatinga está inserida na região denominada Semiárido Brasileiro, caracterizada, principalmente, pela precipitação média anual inferior a 800 mm. A sua cobertura vegetal é classificada como savana estépica, a qual representa a maior extensão fitogeográfica do Nordeste.

De acordo com Myers et al. [3], a Caatinga configura-se como terceiro bioma mais degradado no Brasil. Segundo o Relatório de Monitoramento dos Biomas Brasileiros por Satélite (MMA) [4], até o ano de 2011, a sua porcentagem de área total desmatada foi de 45,82% em relação à área original do bioma, tendo como causa principal o desmatamento da vegetação nativa para produção de lenha, criação de áreas agrícolas e de pastagem.

Com a implantação de unidades objetivando proteger a diversidade biológica, disciplinar o processo de ocupação e assegurar a sustentabilidade do uso dos recursos naturais, e evidenciando a fragilidade da caatinga, torna-se imprescindível o desenvolvimento de estudos técnico-científicos para compreensão e monitoramento desses espaços naturais, incorporando a utilização do potencial tecnológico disponível na atualidade.

Neste aspecto, as geotecnologias, especificamente as técnicas de sensoriamento remoto, têm sido muito utilizadas no estudo e monitoramento da cobertura do solo, tornando possível a extração de informações através da classificação de imagens (COSTA et al.) [5].

Como nova técnica de classificação surge a GEOBIA (*Geographic Object-Based Image Analysis*), a qual, de acordo com Hay e Blaschke [6], é uma subdisciplina da Geoinformática empenhada no desenvolvimento de métodos automáticos para, a partir de imagens de sensoriamento remoto, extrair atributos significativos e particioná-las em objetos, e também analisar suas características através de escalas espaciais, espectrais e temporais.

As técnicas de mineração de dados também vêm sendo utilizadas para o acompanhamento e monitoramento da cobertura do solo, sendo consideradas por Korting [7] um método de descoberta de padrões a partir de dados.

Segundo Breiman et al. [8], dentre as técnicas de mineração de dados, a árvore de decisão possui destaque, tendo como dado de entrada atributos e como valor de saída uma decisão.

Numa árvore de decisão, os nós representam atributos, as ligações de um nó representam os valores desse atributo e as folhas representam as classes a que pertencem os objetos.

Nesse contexto, o principal objetivo deste trabalho consistiu em mapear a cobertura do solo da APA do Cariri-PB, utilizando-se imagens de sensoriamento remoto e técnicas de mineração de dados em auxílio à classificação orientada a objeto.

2. MATERIAIS E MÉTODOS

A Área de Proteção Ambiental do Cariri – APA corresponde a uma Unidade de Conservação (UC) de uso sustentável dos recursos naturais, localizada no semiárido do estado da Paraíba, envolvendo os municípios de Boa Vista, Cabaceiras e São João do Cariri (Figura 1).

As imagens de sensoriamento remoto utilizadas são do sensor *MultiSpectral Instrumet* (MSI) do satélite Sentinel-2A (<https://earth.esa.int/web/sentinel/missions/sentinel-2>), Datum WGS84 - UTM, zona 24S, datadas de 4 de agosto de 2017, bandas 2, 3, 4, 8 e 11.

Foram utilizados os *softwares*: ArcGIS - ESRI (*Environmental Systems Research Institute*), versão 10.3; TerraView – INPE (Instituto Nacional de Pesquisas Espaciais), versão 5.3.1; *plugin* GeoDMA - *Geographic Data Mining Analyst*, versão 2.0.0 alpha 3, que, conforme Korting et al. [9], é uma ferramenta de código aberto, desenvolvida pelo INPE, a qual integra métodos de análise de imagens com técnicas de mineração de dados, dedicado à extração de informações e à descoberta de conhecimento em volumosas bases de dados geométricos.

No *software* ArcGIS, para o cálculo do índice SAVI, foram utilizadas as bandas 4 (*red*) e 8 (*NIR*), e o valor $L = 0,5$, considerando que Ballén et al. [10] identificaram a região da APA do Cariri com vegetação de densidade intermediária.

Para o cálculo do índice NDWI, considerando que o satélite Sentinel-2A não dispõe de canais com comprimentos de onda exatamente iguais aos indicados por Gao [11] - $0,86 \mu\text{m}$ e $1,24 \mu\text{m}$, foram utilizadas as bandas 8 ($0,842 \mu\text{m}$) e 11 ($1,610 \mu\text{m}$).

O algoritmo de segmentação utilizado foi o proposto por Baatz e Schäpe [12], no TerraView. Diversas configurações de parâmetros foram testadas, a fim de se obter uma melhor segmentação. Posteriormente, o arquivo *raster* resultante foi convertido para um arquivo do tipo vetorial.

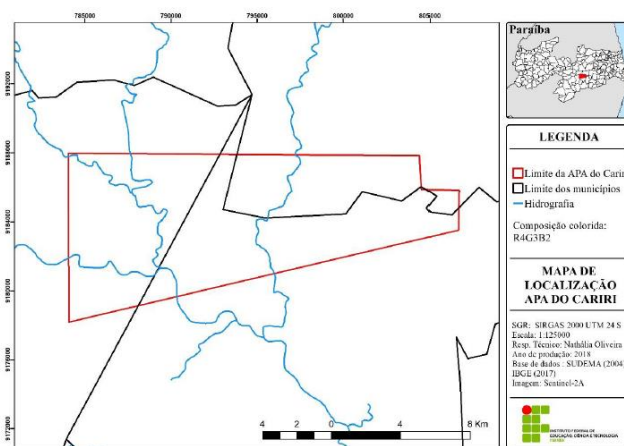


Figura 1 – Mapa de localização da APA do Cariri-PB

Para o cálculo dos atributos, criou-se um agrupamento de arquivos *raster* composto pelas bandas 2, 3, 4, 8, e os *rasters* dos índices SAVI e NDWI.

Através da ferramenta *Feature Extraction* do GeoDMA, foram computados 21 atributos estatísticos espectrais para cada *raster* e 17 atributos espaciais dos *pixels* do agrupamento de arquivos *raster* contidos nos polígonos do arquivo vetorial gerado no processo de segmentação. Por fim, foi gerada uma tabela geral de atributos do agrupamento de arquivos *rasters*.

As classes de cobertura do solo definidas foram: caatinga densa, caatinga semidensa, caatinga aberta, solo exposto, afloramento rochoso, nuvem e sombra. No total, foram selecionadas 186 amostras.

O algoritmo de árvore de decisão, que estabelece regras de classificação, disponível no GeoDMA, é o algoritmo de Quinlan C5.0, que é executado em duas fases: de treinamento e de classificação. A fase de treinamento consiste em construir uma árvore de decisão baseada nos atributos das amostras de treinamento. Como dados de entrada, utilizaram-se os atributos espectrais e espaciais das amostras. A segunda fase (de classificação) consiste em classificar a imagem segmentada usando a árvore de decisão construída na fase anterior.

Desta forma, tem-se como arquivo de saída um *shapefile*, cujos objetos são classificados de acordo com o modelo de regras da árvore de decisão, gerando uma coluna na tabela de atributos, nomeada como Classe, sendo esta preenchida com as respectivas classes dos objetos.

3. RESULTADOS E DISCUSSÃO

Os parâmetros utilizados na segmentação foram 50 para o tamanho mínimo do segmento, 0,9 para cor, 0,3 para forma e 0,12 para o limiar de similaridade. Como resultado, obteve-se uma imagem segmentada composta de 21.928 polígonos.

Na geração do modelo de regras foram escolhidos como atributos que melhor caracterizam cada classe apenas

atributos espectrais, mostrados na Figura 2, não sendo considerado nenhum atributo espacial.

Esses atributos espectrais foram: o valor mínimo da banda NIR, a mediana da banda SAVI, a razão de bandas (banda do azul), a média do SAVI e o valor mínimo do NDWI.

O algoritmo de mineração de dados posiciona na raiz da árvore de decisão o atributo com a maior quantidade de informação, selecionado por ele. E a cada nó consequente da árvore, tem-se uma partição em conjuntos que vão se tornando cada vez mais homogêneos. A variável posta pelo algoritmo na raiz da árvore foi o valor mínimo da banda NIR. Sendo assim, os objetos que possuem o valor mínimo da banda NIR menor ou igual a 1032 são classificados como sombra, e, para valores superiores a este valor, existem outras regras de classificação.

A próxima divisão é feita pela mediana do SAVI. Os objetos que possuem valores menores ou iguais a 0,621 seguem para a próxima regra de classificação, que é a relação de bandas da banda do azul. Caso a relação de bandas da banda do azul seja menor ou igual a 0,199, segue-se para a próxima regra de classificação, a média do SAVI. Se esta for menor ou igual a 0,398, o objeto é classificado como solo exposto; se for acima deste valor, é classificado como caatinga aberta.

Caso o valor da razão de bandas da banda do azul seja maior que 0,199, segue-se para o atributo valor mínimo do NDWI. Caso o valor deste seja menor ou igual a -0,121, o objeto é classificado como afloramento rochoso; caso seja superior, é classificado como nuvem.

O mapa de cobertura de solo gerado através da classificação da imagem encontra-se na Figura 3.

O Quadro 1 apresenta a área total, em hectares, e a porcentagem de ocupação de cada classe.

Quadro 1 – Quantitativo das classes de cobertura do solo

CLASSE	ÁREA (HA)	PORCENTAGEM
Caatinga aberta	4.044,953	26,07 %
Caatinga semidensa	5.854,866	37,74 %
Caatinga densa	2.649,045	17,07 %
Afloramento rochoso	645,281	4,16 %
Solo exposto	1.920,225	12,38 %
Nuvem	173,377	1,12 %
Sombra	227,105	1,46 %
TOTAL	15.514,852	100 %

Observa-se no Quadro 1 que a cobertura de solo com maior área de ocupação é a caatinga semidensa, abrangendo 37,74 % da área total, e, logo em seguida, tem-se a caatinga aberta com 26,07%.

Com 17,07% de cobertura, as áreas de vegetação mais densa, em sua maior parte, encontram-se nos entornos de afloramentos rochosos, setores que conservam maior umidade devido ao escoamento superficial dos lajedos ocorrente em períodos de chuva.

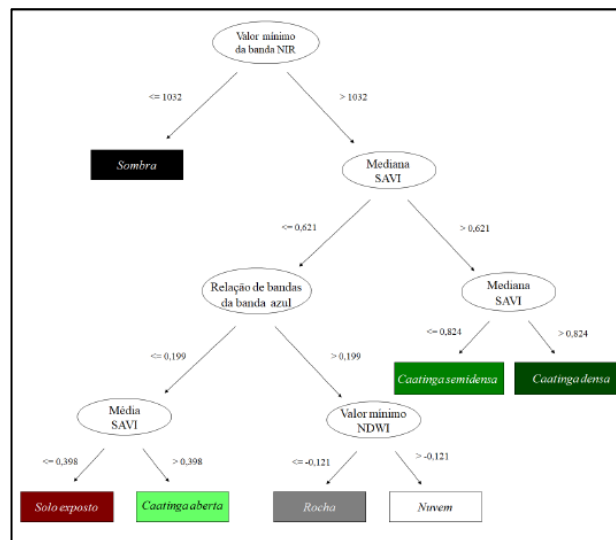


Figura 2 – Árvore de decisão gerada pelo algoritmo

A classe de solo exposto se apresenta em 12,38 % da área total, enquanto a de afloramento rochoso, em 4,16 %.

Em pesquisa realizada na APA, no ano de 2014, Lima [13], utilizando as mesmas classes, encontrou os seguintes resultados: caatinga semidensa ocupando 35,18% da área, seguida por solo exposto (30,39%), caatinga aberta (27,28%), caatinga densa (7,14%) e água (0,01%). No atual trabalho, a classe de cobertura predominante foi, também, a caatinga semidensa. Porém, a classe solo exposto apresentou área inferior a todas as classes de vegetação. Outro ponto de divergência foi a classe caatinga densa, a qual apresentou uma área de cobertura maior do que a do ano de 2014.

Os valores mais divergentes podem ter sido influenciados pelos índices pluviométricos registrados em cada período estudado, pois entre o ano de 2011 até a data da imagem de 2014, a região onde a APA está inserida passou por um período de seca mais acentuada do que o período da imagem deste estudo.

Como resultados da validação da classificação, o índice global obtido foi de 0,887 (88,7%), o qual, segundo Guptill e Morrison [14], encontra-se acima do valor mínimo aceitável de 85% para uma classificação. O índice Kappa para a classificação foi de 0,866.

4. CONCLUSÕES

A combinação da análise de imagens baseada em objetos com o algoritmo árvore de decisão, do campo da mineração de dados, apresentou um bom desempenho na classificação de cobertura do solo da APA do Cariri, considerando o valor alcançado do índice Kappa, de 0,866, indicando uma classificação de qualidade excelente.

A classe de cobertura do solo com maior ocorrência na APA do Cariri-PB é “caatinga semidensa”, ocupando 37,74% de sua área, seguida da classe “caatinga aberta”, que abrange uma extensão correspondente a 26,07% desta área de preservação permanente, no semiárido paraibano.

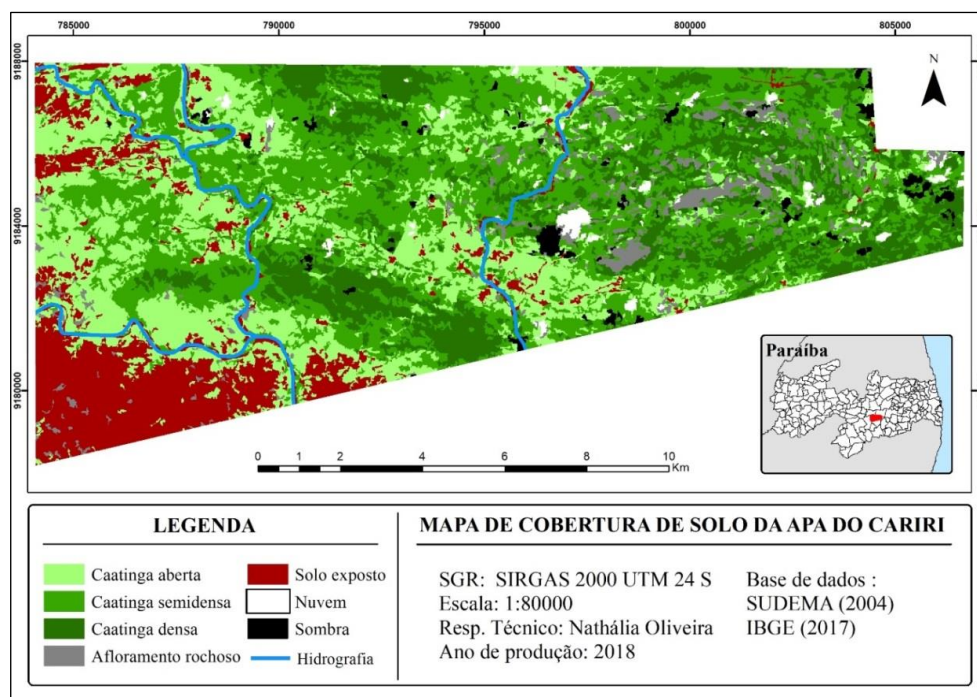


Figura 3 – Mapa de cobertura do solo da APA do Cariri-PB.

5. REFERÊNCIAS

- [1] ALVES, J. J. A.; ARAÚJO, M. A. DE; NASCIMENTO, S. S. DO. Degradação da caatinga: uma investigação ecogeográfica. *Revista Caatinga*, Mossoró, v. 22, n. 3, p. 126-135. 2009.
- [12] BAATZ, M.; SCHÄPE, A. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *Journal of Photogrammetry and Remote Sensing*, Heidelberg, v. 58, n. 3-4, p. 12-23, 2000.
- [10] BALLÉN, L. A. C.; SOUZA, B. I.; LIMA, E. R. V. Análise espaço-temporal da cobertura vegetal na área de proteção ambiental do Cariri, Paraíba, Brasil. *Boletim Goiano de Geografia*, Goiânia, v.36, p. 555-571, 2016.
- [4] BRASIL. Ministério do Meio Ambiente. Monitoramento do desmatamento dos Biomas Brasileiros por Satélite: acordo de Cooperação Técnica MMA/IBAMA. Monitoramento do Bioma Caatinga 2010 a 2011. MMA: Brasília, 2011. 32 p.
- [2] BRASIL. Ministério do Meio Ambiente. Subsídios para a Elaboração do Plano de Ação para a Prevenção e Controle do Desmatamento na Caatinga. MMA: Brasília, 2010. 128 p.
- [8] BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R.A.; STONE, C. J. Classification and regression trees (The Wadsworth statistics/probability series). Belmont: Wadsworth International, 1984. 368 p.
- [5] COSTA, G.A.O.P.; PINHO, C.M.D.; FEITOSA, R.Q.; ALMEIDA, C.M.; KUX, H.J.H.; FONSECA, L.M.G.; OLIVEIRA, D. InterIMAGE: Uma plataforma cognitiva open source para a interpretação automática de imagens digitais. *Revista Brasileira de Cartografia – RBC*, n. 60/4, p.331-337, dez. 2008.
- [11] GAO, B. C. NDWI – A Normalized Difference Water Index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, v. 58, p. 257-266, 1996.
- [14] GUPTILL, S. C; MORRISON, J.L. Elements of Spatial Data Quality. 2. ed. Elsevier Sci., U.K. 78p. 1995.
- [6] HAY, G.; BLASCHKE, T. Object-based image analysis: strengths, weaknesses, opportunities and threats (SWOT). *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, v. XXXVI, n. 4/C42, Salzburg, Austria, 2006.
- [7] KORTING, T.S. GeoDMA: a toolbox integrating data mining with object-based and multi-temporal analysis of satellite remotely sensed imagery. 2012. 123f. Tese (Doutorado em Sensoriamento Remoto) – INPE, São José dos Campos, 2012.
- [9] KORTING, T. S.; FONSECA, L. M. G.; ESCADA, M. I. S.; CÂMARA, G. GeoDMA – Um sistema para mineração de dados de sensoriamento remoto. In: *Simpósio Brasileiro de Sensoriamento Remoto*, 14., Natal. Anais... Natal: INPE 2009. p. 7813 - 7820.
- [13] LIMA, F. S. Análise espaço-temporal da cobertura vegetal na área de proteção ambiental (APA) do Cariri - PB. 2016. 85 p. Monografia (Bacharel em Ecologia) - Universidade Federal da Paraíba, RioTinto, 2016.
- [3] MYERS, N.; MITTERMEIER, R. A.; MITTERMEIER, C. G.; FONSECA, G. A. B.; KENT, J. Biodiversity hotspots for conservation priorities. *Nature*, n. 403, p.853-859, 2000.