

PROGRAMA DE PÓS-GRADUAÇÃO em
Ciências Ambientais

WANESSA MONTEIRO FERNANDES

**NDVI E ÍNDICES DO ESPECTRO VISÍVEL COMO DESCRITORES DE UMA
RANDOM FOREST PARA CLASSIFICAR A COBERTURA DO SOLO NO
ENTORNO DA REPRESA DE ITUPARARANGA – SP POR MEIO DE IMAGENS
DO SATÉLITE CBERS 4A**

Sorocaba

2023

PROGRAMA DE PÓS-GRADUAÇÃO em

ciências
ambientais



unesp
Sorocaba

WANEISSA MONTEIRO FERNANDES

**NDVI E ÍNDICES DO ESPECTRO VISÍVEL COMO DESCRITORES DE UMA
RANDOM FOREST PARA CLASSIFICAR A COBERTURA DO SOLO NO
ENTORNO DA REPRESA DE ITUPARARANGA – SP POR MEIO DE IMAGENS
DO SATÉLITE CBERS 4A**

Dissertação apresentada como requisito para a obtenção do título de Mestre em Ciências Ambientais da Universidade Estadual Paulista “Júlio de Mesquita Filho” na Área de Concentração Geoprocessamento e modelagem matemática ambiental.

Orientador: Prof. Dr. Antonio Cesar Germano Martins.

Coorientador: Prof. Dr. Darllan Collins da Cunha e Silva.

Sorocaba
2023

PROGRAMA DE PÓS-GRADUAÇÃO em

ciências
ambientais



unesp
Sorocaba

F363n	<p>Fernandes, Wanessa Monteiro</p> <p>NDVI e índices do espectro visível como descritores de uma Random Forest para classificar a cobertura do solo no entorno da Represa de Itupararanga - SP por meio de imagens do satélite CBERS4A / Wanessa Monteiro Fernandes. -- Sorocaba, 2023</p> <p>54 p.</p> <p>Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia, Sorocaba</p> <p>Orientador: Antonio Cesar Germano Martins</p> <p>Coorientador: Darllan Collins da Cunha Silva</p> <p>1. CBERS4A. 2. Geoprocessamento. 3. Random Forest. I. Título.</p>
-------	---

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Ciência e Tecnologia, Sorocaba. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO


TÍTULO DA DISSERTAÇÃO: NDVI e índices do espectro visível como descritores de uma Random Forest para classificar a cobertura do solo no entorno da represa de Itupararanga – SP por meio de imagens do satélite CBERS 4A

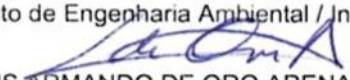
AUTORA: WANESSA MONTEIRO FERNANDES

ORIENTADOR: ANTONIO CESAR GERMANO MARTINS

COORIENTADOR: DARLLAN COLLINS DA CUNHA E SILVA

Aprovada como parte das exigências para obtenção do Título de Mestra em Ciências Ambientais, área: Diagnóstico, Tratamento e Recuperação Ambiental pela Comissão Examinadora:


Prof. Dr. ANTONIO CESAR GERMANO MARTINS (Participação Presencial)
Departamento de Engenharia Ambiental / Instituto de Ciência e Tecnologia UNESP Campus de Sorocaba


Prof. Dr. LUIS ARMANDO DE ORO ARENAS (Participação Presencial)
Departamento de Engenharia Ambiental / Universidade Estadual Paulista - UNESP - Campus de Sorocaba - ICTS

Prof. Dr. WALDEMAR BONVENTI JÚNIOR (Participação Presencial)
Departamento de Engenharia / Faculdade de Tecnologia de Sorocaba - FATEC



Sorocaba, 15 de fevereiro de 2023

Com muito amor e carinho, Josemar Pereira dos Santos, “Tio Neno” (*in memoriam*),
Valdízia Furtado Ferreira da Costa, “Vó Dora” (*in memoriam*)

AGRADECIMENTOS

À Deus, por me sustentar até aqui;

Ao meu Orientador, Prof. Dr. Antonio Cesar Germano Martins, pelos ensinamentos e pela paciência ao longo da elaboração desta pesquisa que foi fundamental para a sua conclusão;

Ao meu coorientador, Prof. Dr. Darllan Collins, pelos auxílios durante a elaboração desta pesquisa;

A toda a minha família, em especial aos meus pais, por sempre acreditarem nos meus sonhos. Aos meus irmãos pela torcida de sempre e ao meu sobrinho Enzo por tornar os meus dias mais felizes;

Ao amigo Jonilson Michel pela grande ajuda na elaboração desta pesquisa e também aos amigos Caick e Lázaro, pelas trocas de ideias, principalmente durante a Pandemia.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

À todos os professores e funcionários da UNESP/Sorocaba.

MUITO OBRIGADA!!

RESUMO

O sensoriamento remoto tem grande destaque na análise e monitoramento da superfície terrestre, especialmente devido a possibilidade de analisar grandes extensões de área, com destaque para os métodos computacionais que utilizam índices baseados nos canais do espectro eletromagnético. Visto isso, o objetivo deste estudo foi utilizar índices do espectro do visível e o NDVI (*Normalized Difference Vegetation Index*) obtidos a partir de imagens do satélite CBERS 4A em conjunto com a aplicação do classificador Random Forest para a análise de regiões no entorno da represa de Itupararanga, localizada na Região Metropolitana de Sorocaba (RMS). Para isso, utilizou-se os softwares QGis e Rstudio, nos quais foram realizados o pré-processamento das imagens, cálculo dos índices, coletas de imagens de treinamento das regiões características e implementação da Random Forest. As classes de cobertura do solo obtidas foram agricultura, área urbanizada, corpo hídrico, solo exposto, vegetação arbórea e vegetação rasteira, permitindo uma análise qualitativa e quantitativa pela Random Forest. Observou-se que o NDVI apresentou um melhor desempenho em relação aos outros índices utilizados na segmentação das diversas regiões de cobertura do solo, além da banda do infravermelho próximo (NIR), sendo os descritores mais relevantes para a classificação das regiões usando a Random Forest.

Palavras-chaves: Índices espectrais; CBERS 4A; Itupararanga; Random Forest.

ABSTRACT

Remote sensing has great prominence in the analysis and monitoring of the Earth's surface, since through sensors it is possible to analyze large areas. One of the methods for this analysis are those that use indices based on the channels from the electromagnetic spectrum. This research aimed to study the use of visible spectrum indices and NDVI (Normalized Difference Vegetation Index) obtained from CBERS 4A satellite images and a Random Forest as a classifier for the analysis of regions around the Itupararanga dam. For this, the softwares QGIS and Rstudio were used, in which the pre-processing of the images, calculation of the indices, collection of training images of the characteristic regions and implementation of the Random Forest were carried out. The following land cover classes were considered in this study: agriculture, urbanized area, water, exposed soil, arboreal vegetation and undergrowth. In addition to the visual qualitative analysis of the index relevancies, the use of Random Forest allowed a quantitative evaluation. It was observed that the NDVI presents an excellent result in the segmentation of the different land cover regions and that, individually, the near infrared band (NIR) and the NDVI are the most relevant descriptors for the classification of the regions using Random Forest.

Keywords: Spectral index; CBERS 4A; Itupararanga; Random Forest.

LISTA DE FIGURAS

Figura 1- Ilustração do satélite CBERS 4A.....	19
Figura 2- Cubo do espaço de cores RGB	22
Figura 3- Exemplo Random Forest	29
Figura 4- Metodologia.....	31
Figura 5 - Área de estudo	31
Figura 6- RGB e composição de banda.....	33
Figura 7 - Amostras exemplificadas.....	33
Figura 8 - Coleta de amostras.....	34
Figura 9 - Resultado da composição da imagem na cor verdadeira	37
Figura 10 - Imagens resultantes dos Índices; (a) NDI, (b) NGRDI, (c) MExG, (d) ExR, (e) VEG, (f) ExG (g) ExGR, (h) CIVE, (i) COM1, (j) COM2, (k) NDVI.....	38
Figura 11- Gráfico do erro OOB em uma função do número de árvores	44
Figura 12- Classificação segundo o grau de importância dos índices.....	45
Figura 13 - - Resultado da classificação com a RF	46

LISTA DE TABELAS

Tabela 1- Principais características da câmera MUX do CBERS 4A.	19
Tabela 2 - Principais características da câmera WFI do CBERS 4A.	20
Tabela 3- Principais características da câmera WPM do CBERS 4A.	20
Tabela 4- Descrição das bandas espectrais do sensor WPAN do CBERS 4A	32
Tabela 5 - Pacotes utilizados Rstudio.....	35

LISTA DE ABREVIATURAS E SIGLAS

B	= Blue
CIVE	= Colour Index of Vegetation Extraction
COM	= Índices combinados
ExG	= Excess Green Index
ExGR	= Excess Green minus Excess Red Index
ExR	= Excess Red Index
G	= Green
INPE	= Instituto Nacional de Pesquisas Espaciais
MExG	= Modified Excess Green Index
NDI	= Normalised difference index
NDVI	= Normalized Difference Vegetation Index
NGRDI	= Normalized Green–Red Difference Index
R	= Red
RF	= Random Forest
VEG	= Vegetative Index

SUMÁRIO

1	INTRODUÇÃO	13
2	JUSTIFICATIVA	16
3	OBJETIVOS.....	17
3.1	OBJETIVOS ESPECÍFICOS	17
4	REVISÃO BIBLIOGRÁFICA	18
4.1	SENSORIAMENTO REMOTO.....	18
4.2	SATÉLITE CBERS 4A	18
4.3	IMAGEM DIGITAL	20
4.4	RGB.....	21
4.5	PROCESSAMENTO DE IMAGEM DIGITAL	22
4.6	ÍNDICES DE VEGETAÇÃO BASEADOS NOS CANAIS DE COR RGB	23
4.7	ÍNDICES DE VEGETAÇÃO POR DIFERENÇA NORMALIZADA (NDVI)	27
4.8	CLASSIFICAÇÃO DE REGIÕES EM IMAGEM DIGITAIS	28
5	METODOLOGIA.....	31
5.1	CARACTERIZAÇÃO DA ÁREA DE ESTUDO	31
5.2	AQUISIÇÃO DE DADOS.....	32
5.3	AMOSTRAGEM DE DADOS E CÁLCULO DE ÍNDICES	32
5.4	IMPLEMENTAÇÃO DA RANDOM FOREST (RSTUDIO)	34
6	RESULTADOS OBTIDOS	37
6.1	QUALITATIVO.....	37
6.2	RESULTADOS DA RANDOM FOREST	44
7	CONCLUSÃO.....	47
	REFÊRENCIAS	48

1 INTRODUÇÃO

A alteração da cobertura do solo pode acarretar diversos problemas para a região onde se está inserida, bem como no seu entorno, devido a modificação do balanço ambiental presente bem como a introdução de processos que podem levar a impactos físicos, químicos e bióticos.

Um dos efeitos mais notáveis é a erosão hídrica, que está associada à ocupação irregular de áreas, inexistência de práticas conservacionistas e ao manejo realizado de forma incorreta ocasionando o empobrecimento do solo, poluição e, além disso, o assoreamento de rios; (BERTONI; LOMBARDI NETO, 1999; FERREIRA et al., 2011).

O principal fator desencadeador de processos erosivos é a exposição da superfície do solo ocasionado pela diminuição da cobertura vegetal permitindo a ação da água que resulta no carreamento de materiais e a ocorrência de caminhos que tendem a se aprofundar e alargar com a continuidade do processo.

Um outro processo muito comum é a retirada da cobertura vegetal original para a prática agrícola que leva a um desequilíbrio nos processos físico-químicos do solo além de introduzir no meio produtos que são utilizados como defensivos agrícolas.

A alteração da cobertura do solo pela construção civil também impacta o meio ao alterar o processo de absorção da água pelo solo a partir da colocação de pisos e telhados impermeáveis e a criação de deslocamentos preferencias da água de chuvas que podem gerar alterações nas vazões de corpos hídricos e alagamentos.

O sensoriamento remoto destaca-se como uma importante ferramenta na análise e monitoramento de áreas na superfície da Terra, permitindo avaliar alterações no uso e ocupação do solo que ocasionem degradação ambiental (CUNHA et al., 2009; LOPES et al., 2010; RAMOS et al., 2011; CUNHA et al., 2012; BEZERRA et al., 2014; SALLO et al., 2014).

De uma forma geral, em boas condições, a vegetação verde reflete a radiação infravermelha e, em contrapartida, uma vegetação com taxas de concentração de clorofila menor, absorvem a maior parte dessa radiação

Através dos produtos do sensoriamento remoto é possível calcular índices de vigor da vegetação que permitem supervisionar alterações naturais e antrópicas no uso e cobertura do solo.

Esses índices podem ser calculados através de dados espectrais de duas ou mais bandas que são combinados entre si, com o intuito de associar os seus valores com indicadores biofísicos da vegetação. (PONZONI, 2001).

Proposto por Rouse et al. (1973) o Normalized Difference Vegetation – NDVI, *Índice de Vegetação por Diferença Normalizada* é calculado utilizando dados espectrais e demonstra um bom desempenho para analisar mudanças na cobertura do solo possibilitando avaliar o teor de clorofila da vegetação através das bandas espectrais do infravermelho próximo e vermelho-visível.

Outra possibilidade é a geração de índices utilizando canais de cores do espectro visível como feito por Ohta, Kanade e Sakai (1980) que analisaram índices baseados nos canais R (red), G (green) e B (blue), através das relações $(R + G + B) / 3$, $R-B$ e $(2G-R-B)$ e constataram suas eficiências na segmentação de imagens coloridas.

Pode-se ainda, converter a imagem de RGB para outros espaços de representação de cores (RANGAYYAN; ACHA; SERRANO, 2011) e obter relações entre estes canais, como feito por Cruz (2020).

Hamuda, Glavin e Jones (2016) apresentam um trabalho de revisão que compara os resultados obtidos com diversos índices obtidos dos canais do espectro eletromagnético para avaliar culturas agrícolas.

Existem diversos satélites que fornecem bandas na faixa do visível e do infravermelho. O satélite Sentinel - 2 lançado em 2015 possui treze canais espectrais e dentre eles o RGB e NIR. Da mesma forma o Satélite CBERS 3 e 4, possuem oito canais espectrais do RGB e do NIR. Assim, ambos sensores permitem a extração de recursos para cálculos de índices vegetativos considerando apenas os canais na faixa do visível e associados ao infravermelho próximo.

Após a extração destes índices, deve-se realizar a avaliação dos valores obtidos para que seja feita a classificação da cobertura do solo através de um processo automático.

Devido aos desenvolvimentos e resultados obtidos com os métodos de inteligência artificial (RUSSEL; NORVIG, 2013), os processos de classificação em imagens de satélite tem cada vez mais utilizado abordagens baseadas em Redes Neurais Artificiais, Máquinas de Vetores de Suporte e Florestas Aleatórias (XIE; NICULESCU, 2021; PRASAD et al., 2022; LUCA et al., 2022).

Uma aplicação importante de análise da variação da cobertura do solo está associada a reservatórios de armazenamento de água. Como exemplo têm-se a represa de Itupararanga que está localizada em uma área de preservação ambiental (APA) na Região Metropolitana de Sorocaba e está em um processo gradativo de desmatamento do seu entorno, com a vegetação sendo substituída pela construção de loteamentos e pela agricultura.

Este trabalho tem como objetivo principal, o estudo do uso de índices do espectro visível e do NDVI obtidos de imagens do satélite CBERS 4A e uma Radom Forest como classificador para a análise de regiões no entorno da represa de Itupararanga.

2 JUSTIFICATIVA

O NDVI apresenta um bom desempenho na avaliação de uso e ocupação do solo, pois possibilita a identificação, distribuição e caracterização da vegetação verde, contribuindo com análises da cobertura vegetal de uma determinada região (SILVEIRA et al., 2022). Da mesma forma índices obtidos a partir de canais do espectro visível apresentam bons resultados na segmentação de imagens (HAMUDA; GLAVIN; JONES, 2016).

A comparação de seus desempenhos permite que se escolham aqueles que sejam mais eficientes na caracterização do uso e ocupação do solo para problemas específicos e que tenham assinaturas espectrais características, quando da disponibilização de produtos de imageamento obtidos pelo lançamento de novos satélites, como é o caso do CBERS 4A, cujas características serão apresentadas à frente.

Através do mapeamento das atividades antrópicas no entorno da represa de Itupararanga, pode-se notar ao longo dos anos que tem havido um decréscimo na vegetação original nas áreas de APP e alteração do uso e ocupação do solo principalmente próximo aos afluentes (QUEIROZ; IMAI, 2007; ROSA et al., 2015), o que reflete na qualidade da água do local (MARTINS et al., 2021).

Assim o uso de técnicas de Inteligência Artificial (IA) com descritores que sejam mais apropriados deve melhorar a classificação de uso e ocupação do solo, contribuindo de forma positiva para o acompanhamento desta variação e a tomada de decisão que influencie na melhoria da qualidade da água de reservatórios.

3 OBJETIVOS

Este trabalho tem como objetivo principal, o estudo do uso de índices do espectro visível e do NDVI obtidos de imagens do satélite CBERS 4A e uma Radom Forest como classificador para a análise de regiões no entorno da represa de Itupararanga.

3.1 OBJETIVOS ESPECÍFICOS

- Avaliar a relevância dos índices do espectro visível e do NDVI obtidos de imagens do satélite CBERS 4A para a segmentação de coberturas do solo;
- Contribuir para o uso de técnicas de IA para a análise ambiental;
- Avaliar o uso de Random Forest na análise de imagens do satélite CBERS 4A.

4 REVISÃO BIBLIOGRÁFICA

4.1 SENSORIAMENTO REMOTO

As imagens de sensoriamento remoto (SR) buscam identificar objetos e estruturas na superfície terrestre sem que ocorra um contato físico e podem ajudar a medir e observar estruturas detalhadas mesmo a longas distâncias, sendo o principal objetivo dessas imagens auxiliar a caracterização de regiões para uma melhor análise da distribuição espacial.

De forma bastante abrangente, o SR é um mecanismo de medição e aquisição de dados de um fenômeno ou um objeto através de um dispositivo que pode ser: câmeras digitais, radares, gravímetro, entre outros (LEVIN, 1999).

Para Leite (2011), o SR tem grande destaque já que os dados obtidos através dos sensores servem de suporte para diversas pesquisas, em destaque as que auxiliam em estudos sobre a superfície terrestre, visto que permite a obtenção e análise de dados temporais, levando a uma melhora na quantificação dos dados das imagens geradas.

Para Meneses e Almeida (2012, p.3) o sensoriamento remoto é “Uma ciência que visa o desenvolvimento da obtenção de imagens da superfície terrestre por meio da detecção e medição quantitativa das respostas das interações da radiação eletromagnética com os materiais terrestre”. Dessa forma, dá suporte para diversas pesquisas relacionadas ao assunto.

Ainda nesse sentido, o SR registra objetos através da reflexão de ondas eletromagnéticas pela superfície de qualquer objeto. Dessa forma, essas imagens não devem ser consideradas apenas como aquelas que resultam de satélites orbitais, mas podem ser obtidos de balões, aviões e mais recentemente de veículos aéreos não tripulados (VANTs). Da mesma forma, a radiação eletromagnética pode ter como fonte o Sol ou algum dispositivo de geração dessa radiação, valendo ressaltar que no primeiro caso têm-se o chamado método de coleta de dados passivo e no segundo o ativo (LEVIN, 1999; FIGUEIREDO, 2005).

Mapear a cobertura do solo por SR depende da classificação das regiões presentes na imagem, que podem ser implementadas por meio de abordagens baseadas em pixels, objetos ou cenas (BLASCHKE, 2010; MA et al., 2017; ZHANG et al., 2018).

4.2 SATÉLITE CBERS 4A

Lançado em 20 de Dezembro de 2019, o CBERS 4A, Figura 1, é o sexto satélite do programa CBERS criando em 1988 e que conta com a participação do Instituto Nacional de Pesquisas Espaciais (INPE), da Academia Chinesa de Tecnologia Espacial (CAST) e do Centro de Controle de Satélites de Xian (XSCC), sendo o monitoramento ambiental a principal função deste satélite.

Figura 1- Ilustração do satélite CBERS 4A.



Fonte: MUNDOGEO, 2020.

O CBERS 4A, possui três câmeras imageadoras (MUX, WFI e WPM) descritas a seguir.

- **Câmera Multiespectral (MUX)**

Esta câmera imageadora é a mesma utilizada nas séries anteriores 3 e 4, entretanto, devido a uma redução na altitude da sua órbita, ela possui uma maior resolução que as versões anteriores. A Tabela 1 apresenta as principais características desta câmera.

Tabela 1- Principais características da câmera MUX do CBERS 4A

Característica	Dado
Bandas espectrais	B05: 0,45 - 0,52 μm
	B06: 0,52 - 0,59 μm
	B07: 0,63 - 0,69 μm
	B08: 0,77 - 0,89 μm
Largura da Faixa Imageada	95 km
Resolução Espacial (Nadir)	16,5 m
Visada Lateral de Espelho	Não
Taxa Bruta de Dados	65 Mbps

Fonte: INPE, 2019a.

- **Câmera de Campo Largo (WFI)**

Esta câmera imageadora também possui, de forma geral, as mesmas características que as câmeras das versões 3 e 4, porém, com resolução espacial de 55m. A Tabela 2 apresenta as principais características desta câmera.

Tabela 2 - Principais características da câmera WFI do CBERS 4A

Característica	Dado
Bandas Espectrais	B13: 0,45 - 0,52 μm
	B14: 0,52 - 0,59 μm
	B15: 0,63 - 0,69 μm
	B16: 0,77 - 0,89 μm
Largura da Faixa Imageada	684 km
Resolução Espacial	55 m
Visada Lateral de Espelho	Não
Taxa Bruta de Dados	51 Mbps

Fonte: INPE, 2019b.

- **Câmera Multiespectral e Pancromática de Ampla Varredura (WPM)**

A WPM é a principal câmera imageadora deste satélite, já que a resolução panorâmica, projetada neste sensor é de 2m, e a multiespectral é de 8m. Para esta pesquisa foi utilizada esta câmera imageadora, devido sua melhor resolução espacial, o que facilita a visualização das classes. A Tabela 3 apresenta as principais características desta câmera.

Tabela 3- Principais características da câmera WPM do CBERS 4A

Característica	Dado
Bandas espectrais	P: 0,45 - 0,90 μm
	B1: 0,45 - 0,52 μm
	B2: 0,52 - 0,59 μm
	B3: 0,63 - 0,69 μm
	B4: 0,77 - 0,89 μm
Largura da Faixa Imageada	92 km
Resolução Espacial	2 m (pancromática)
Visada Lateral de Espelho	Não
Taxa Bruta de Dados	1800.8 Mbps 450.2 Mbps

Fonte: INPE, 2019c.

4.3 IMAGEM DIGITAL

De acordo com Cruvinel et al (1996), imagens digitais são resultado de estímulos luminoso que são captados pelo olho humano em que numa imagem, real ou fotográfica, cada ponto representa uma informação de cor. Em uma função bidimensional que pode ser representada por “ $i(x,y)$ ”, os valores de “ x ” e “ y ”, correspondem as coordenadas espaciais de um ponto da imagem e “ i ” representa um valor equivalente a intensidade ou cor neste ponto chamado de pixel que representam digitalmente um sinal eletrônico (analógico), gerado por um sensor (HAMILTON, 2010).

Ainda de acordo com Hamilton (2010), três características determinam a qualidade digital da imagem:

- (i) Profundidade de bits: quantidade de bits para representar a intensidade de um pixel;
- (ii) Tamanho da imagem: número de pixels existente na imagem;
- (iii) Densidade de pixels. número de pixels por área da imagem.

Dessa maneira, uma imagem pode ser representada em forma de uma matriz de linhas (x) e colunas (y). Cada pixel da imagem possui um determinado valor digital associado a intensidade de energia eletromagnética que foi refletida ou emitida pelo ponto que representa. Esses valores de intensidade são transformados em uma escala de valores que, normalmente, variam entre 0 a 255, sendo que no caso de imagens apenas com níveis de cinza, representam as cores entre o preto e o branco. (SCHOWENGERDT, 1997).

No caso de imagens coloridas, são usados três canais que, quando apresentados em um dispositivo, transmitem a sensação de cor.

Esses canais podem ser representados por um espaço que mapeia as cores observadas pelo olho humano em pontos de um sistema de coordenadas.

Dentre os diversos espaços de cor, os mais comuns no processamento de imagens são: RGB, CMYK, HSV / HSL, CIE XYZ, $L^*a^*b^*$ e YUV (RANGAYYAN; ACHA; SERRANO, 2011).

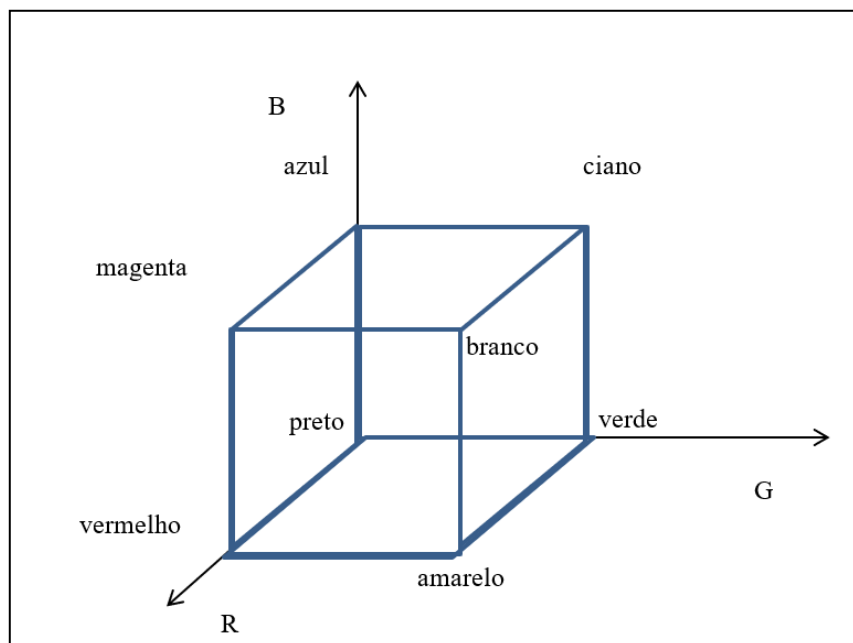
4.4 RGB

O sistema RGB utiliza o modelo de cores aditivas e os canais R, G e B que podem ser representados por um cubo, onde os vértices sobre os eixos são ocupados pelas cores

vermelha, verde e azul e nos demais, pelas cores magenta, ciano e amarelo, conforme mostrado na Figura 2.

Dentre os sistemas de cor, o RGB destaca-se como um dos mais usados, sendo comum que cada canal seja representado por 8bits, permitindo a representação de mais de 16 milhões de cores (GONZALEZ; WOODS, 2000).

Figura 2- Cubo do espaço de cores RGB.



Fonte: Autoria própria.

4.5 PROCESSAMENTO DE IMAGEM DIGITAL

O processamento de imagem digital se refere a manipulação dos pixels presentes nas imagens para melhorar a qualidade visual, ressaltar aspectos importantes presentes, segmentar e classificar objetos e regiões (GONZALEZ; WOODS, 2000).

O tratamento de imagens digitais permite que sejam realizadas várias operações com o intuito de extrair dados que permitem a análise de fenômenos (SILVA, 1996). A primeira etapa do processamento de imagens se baseia na aquisição das mesmas, em que pode ser utilizado qualquer tipo de sensor para digitalizações espectrais, melhorando sua qualidade (MENESES; ALMEIDA, 2012)

De acordo com Gonzales e Woods (2000) as primeiras evidências do uso do Processamento de Imagens Digitais (PDI) foram para melhorar a qualidade das imagens jornalísticas que eram transmitidas através de cabos submarinos entre as cidades de

Londres e Nova York durante a década do 20. Mais tarde o PDI foi utilizado para tratar imagens produzidas por câmeras a bordo da nave espacial Ranger-7.

Computadores de grande porte, microcomputadores (KITTLE et al., 1985) ou pequenos circuitos (LUPPE, 1997) tiveram grande importância no desenvolvimento de diversas aplicações do processamento de imagem. Dois princípios definem processamento digital, o primeiro deles é melhorar o aspecto visual das imagens para melhorar a percepção visual e o segundo é a preparação de imagens para facilitar a identificação de suas características (RUSS, 2007).

Classificar cenas de sensoriamento remoto em um conjunto de classes através de algoritmos que apresentam um resultado satisfatório tem sido um grande desafio. Para Cheng, Xie, Han, Guo e Xia (2020), dentre as dificuldades para esse processo, pode-se destacar:

- I. Grande diversidade intraclasse;
- II. Alta similaridade entre classes;
- III. Grandes variações de escalas de objetos e cenas;
- IV. Coexistência de objetos distintos no solo.

4.6 ÍNDICES DE VEGETAÇÃO BASEADOS NOS CANAIS DE COR RGB

Os índices de vegetação são amplamente utilizados em diversas aplicações agrícolas (MEYER; NETO, 2008; PONTI, 2013; SOONTRANON et al., 2014; HAMUDA; GLAVIN; JONES, 2016) e têm o objetivo de acentuar as características da vegetação, podendo ser calculados a partir de canais do espectro visível ou de outros canais espectrais combinados.

Hamuda, Glavin e Jones (2016) realizaram testes com índices de vegetação em imagens agrícolas com combinações de bandas RGB, que posteriormente foram segmentadas para a diferenciação entre cultura e ervas daninhas e, com base nesses resultados, otimizar o controle da aplicação de herbicidas. Deve-se ressaltar que, além do cálculo de índices, o trabalho analisa o pré-processamento de imagens e posteriormente a isso, discute os estágios da segmentação e compara os algoritmos de extração dos pixels pertencentes as plantas. A seguir serão apresentados os índices comparados pelos autores e que também foram utilizados neste trabalho.

➤ **Normalised difference index (NDI)**

Sugerido por Woebbecke et al. (1992), este índice foi aplicado na separação da cobertura vegetal do solo, sendo calculado pela diferença normalizada entre os canais verde e vermelho:

$$NDI = 128 * \left(\frac{(G-R)}{(G+R)} \right) + 1 \quad (1)$$

Esse índice é de simples implementação computacional e apresenta robustez nos resultados, mesmo quando há diferenças de iluminação.

➤ **Excess Green Index (ExG)**

Woebbecke et al. (1995) analisaram ainda diversos índices obtidos de combinações lineares dos canais de cor RGB normalizados para separar planta verde de solo exposto. Dentre eles, destaca-se o ExG que é obtido por:

$$ExG = 2g - r - b \quad (2)$$

onde:

$$r = \frac{R'}{(R'+G'+B')} \quad g = \frac{G'}{(R'+G'+B')} \quad b = \frac{B'}{(R'+G'+B')}$$

em que R', G' e B', foram normalizados pelos valores máximos em cada canal:

$$R' = \frac{R}{R_{max}}, \quad G' = \frac{G}{G_{max}}, \quad B' = \frac{B}{B_{max}}$$

O ExG é bastante utilizado na separação de plantas dos demais elementos e regiões da imagem (MEYER et al., 1998; LAMM et al., 2002; RIBEIRO et al., 2005; GUERRERO et al., 2012). Vale destacar que em situações de baixa luminosidade, o resultado de seu uso é bastante prejudicado.

➤ **Excess Red Index (ExR)**

Observando que na retina do olho humano existem 4% de receptores azuis, 32% de verde e 64% de vermelhos, Meyer et al. (1998) propuseram o índice ExR para diferenciar pixels de plantas daqueles de outros elementos obtido por:

$$\text{ExR} = 1,3 R - G \quad (3)$$

➤ **Colour Index of Vegetation Extraction (CIVE)**

Proposto por Kataoka et al. (2003), o índice CIVE teve como objetivo inicial avaliar o estágio de crescimento da cultura de beterraba e soja, separando planta de solo. A equação utilizada foi:

$$\text{CIVE} = 0,441R - 0,811G + 0,385B + 18,787 \quad (4)$$

Os proponentes perceberam que o índice CIVE melhora o resultado da segmentação quando comparado com o infravermelho próximo (NIR), já que destaca regiões de área verde. Esse índice possui um bom resultado em imagens retiradas de ambientes ao ar livre, porém a luminosidade é um fator limitante, assim como a presença de sombra na imagem resultante.

➤ **Excess Green minus Excess Red Index (ExGR)**

Meyer et al. (2004) combinaram dois índices, o ExG e o ExR, buscando aumentar o poder de segmentação, uma vez que o índice ExG distingue regiões da planta e o ExR atenua elementos verde-vermelho. Dessa forma o ExGR é calculado por:

$$\text{ExGR} = \text{ExG} - \text{ExR} \quad (5)$$

Uma vantagem bastante significativa neste índice é a sua adaptação a imagens tiradas ao ar livre e ao fato da utilização de dois índices, que extrai o verde e elimina o ruído de fundo. Em contrapartida, quando há exposição a luz excessiva ou pouca luz, o resultado não é tão satisfatório.

➤ **Normalized Green–Red Difference Index (NGRDI)**

Proposto por Hunt et al. (2005), o NGRDI, foi inicialmente testado em fotografias digitais de culturas de grãos, como milho e soja para avaliar o teor de biomassa e nutrientes. Esse índice também foi utilizado para monitoramento de fenologia de vegetação (MOTOHKA et al., 2010) e para o manejo de ervas daninhas, bastante prejudiciais para algumas culturas (TORRES-SÁNCHEZ et al., 2013). O NGRDI é definido pela seguinte equação:

$$\text{NGRDI} = \frac{(G-R)}{(G+R)} \quad (6)$$

Esse índice é baseado na reflectância do verde e do vermelho que podem vir das cores reais da imagem e seu intuito principal é diminuir a diferença no resultado de fotografias aéreas.

➤ **Vegetative Index (VEG)**

Proposto por Hague et al. (2006), o índice VEG procura separar pixels de planta de pixels do solo. Os primeiros estudos foram realizados com câmera CCD em condições de iluminação solar. A equação que define esse índice é dada por:

$$\text{VEG} = \frac{G}{R^{\alpha} * B^{(1-\alpha)}} \quad (7)$$

em que α é uma constante de valor igual a 0,667. Vale ressaltar que o uso do VEG apresentou um bom resultado no aumento do contraste entre o solo e a vegetação, conforme reportado por Hague et al. (2006).

➤ **Modified Excess Green Index (MExG)**

Burgos-Artizzu et al. (2011), definiram o índice MExG para a conversão de imagens coloridas em imagens na escala cinza no sentido de facilitar a binarização, obtendo resultados bastante significativos e robusto quanto submetidos a variações de iluminação. A equação que descreve o MExG é dada por:

$$MExG = 1,262G - 0,884R - 0,311B \quad (9)$$

➤ **Índices combinados (COM1 e COM2)**

Sugerido por Guijarro et al. (2011) o COM1 é o resultado da associação de quatro índices, o ExG, o CIVE, o ExGR e o VEG resultando em:

$$COM1 = ExG + CIVE + ExGR + VEG \quad (8)$$

Guijarro et al. (2012) analisaram também plantas de milho através de uma outra combinação de índices bastante semelhante com o COM1, porém, excluíram o ExGR, e introduziram pesos na combinação linear dos demais. A exclusão do ExGR se deve ao fato que, nos testes realizados, os autores notaram que esse índice resultou em valores semelhantes para a vegetação e sombra. A equação que o define esse novo índice, denominado de COM2 é:

$$COM2 = 0,36ExG + 0,47CIVE + 0,17VEG \quad (10)$$

Conforme Hamuda, Glavin e Jones (2016), a combinação desses índices apresentou um melhor resultado do que quando aplicados separadamente.

4.7 ÍNDICES DE VEGETAÇÃO POR DIFERENÇA NORMALIZADA (NDVI)

Considerando as características físico-químicas das plantas que fazem com que reflitam de maneira peculiar os comprimentos de onda da região do infravermelho e vermelho, foi proposto o NDVI que permite identificar o grau de saúde de uma vegetação através da diferença normalizada entre o canal do infravermelho próximo (NIR) e do vermelho (R) de acordo com:

$$NDVI = \frac{NIR-R}{NIR+R} \quad (11)$$

Onde:

NIR – Banda do infravermelho próximos;

R – Banda canal Vermelho.

Esse índice é muito utilizado no monitoramento da dinâmica de vegetações tanto global (NEMANI et al., 2003, ZHU et al., 2016) quanto em regiões menores (PIÃO et al., 2003, WU et al., 2015), pelo fato de ter relação direta com a cobertura vegetal, fotossíntese, área foliar e biomassa (FORKE et al., 2013).

O NDVI foi criado com o intuito de destacar o contraste entre as regiões de vegetação verde e o solo sendo seus resultados independentem de condições de iluminação. (PONZONI; SHIMABUKURO, 2010)

Para Braz et al. (2015), o índice NDVI apresenta de maneira simples regiões de cobertura com vegetação verde viva. Os valores resultantes nesse índice podem variar entre -1 a +1, onde os resultados mais próximos de -1 são áreas em que a vegetação fotossintética está ativa (“vegetação saudável”), já os valores mais próximos de zero, demonstram mais influência do solo, valendo ressaltar que os valores mais próximos de -1 representam regiões de água e sombra.

Destaca-se que os índices espectrais são amplamente utilizados para analisar as mudanças de uso e cobertura do solo, já que permitem caracterizar e quantificar indicadores biofísicos da vegetação, culturas agrônômicas e mudanças no uso do solo através das informações espectrais que foram capturadas pelos satélites. (TUCKER et al. 1985).

4.8 CLASSIFICAÇÃO DE REGIÕES EM IMAGEM DIGITAIS

O processo de classificação de um pixel ou uma região em uma imagem digital consiste em associá-la a um elemento de um dado conjunto de categorias.

Em geoprocessamento, por exemplo, busca-se associar regiões de uma imagem da superfície terrestre a um tipo de uso e ocupação do solo, tais como: construção, vegetação rasteira, solo exposto, vegetação arbórea, entre outras (ZURSSA; MARTINS, 2022).

Para se realizar o processo de classificação, ao invés de se utilizar todos os pixels, utiliza-se um conjunto de descritores que permitam caracterizar as imagens a serem analisadas. Com isto diminui-se a dimensionalidade do problema.

Após a seleção dos descritores, são utilizados classificadores que permitem a avaliação do grau de aderência destes descritores aos elementos do conjunto de categorias.

Como o grande desenvolvimento e resultados alcançados pelas técnicas de Inteligência Artificial, têm-se, cada vez mais, utilizados como classificadores as Redes

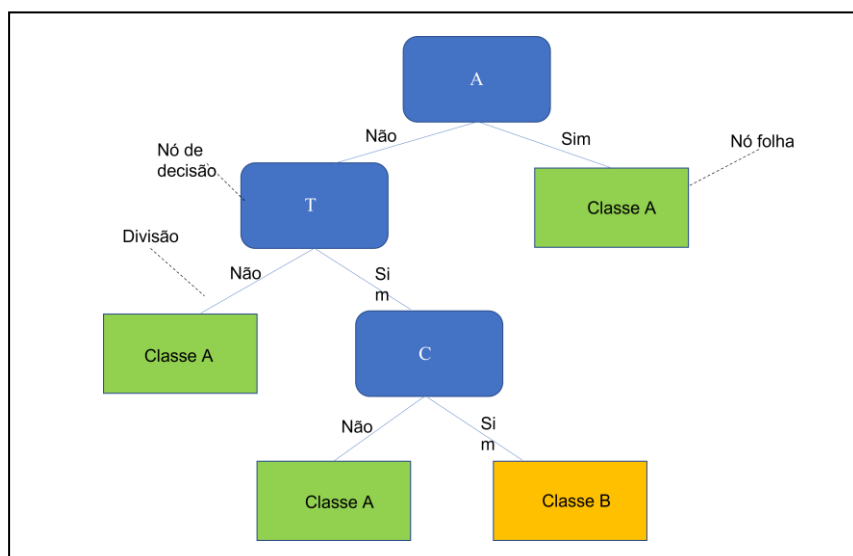
Neurais Artificiais (HAYKIN, 2004), Suporting Vector Machine (MAULIK; CHAKRABORTY, 2017) e Árvores de Decisão (TAN; STEINBACH; KUMAR, 2006).

Uma outra técnica de IA usada para a classificação de um conjunto de descritores é a chamada Random Forest (RF), que foi proposta em 2001 por Leo Breiman, professor da Universidade da Califórnia, Berkeley, EUA como um algoritmo de mineração de dados supervisionado que usa utiliza um extenso número de árvores de decisão. (BREIMAN, 2001).

Essa técnica se baseia no uso de subamostras do conjunto de treinamento para a criação de árvores de decisão que são otimizadas e depois utilizadas para se alcançar um resultado por meio de um processo de votação. Essas subamostras são obtidas de forma aleatória.

Dessa forma, a RF modela conjuntos em que são geradas um grande números de árvores de decisão. Cada árvore é criada com base em uma amostra eventual dos dados de treinamento , Figura 3.

Figura 3- Exemplo Random Forest.



Fonte: Autoria própria.

Para Oshiro (2013, p. 14-15):

Random Forest é outro método ensemble, o qual constrói muitas árvores de decisão que serão usadas para classificar um novo exemplo por meio do voto majoritário. Cada árvore de decisão usa um subconjunto de atributos selecionados aleatoriamente apartir do conjunto original, contendo todos os atributos.

De acordo com Belgiu e Drăguț, (2016) e Breiman, (2001), os modelos de IA que utilizam RF diminuem o overfitting entre os dados, além de ser mais robusto do que outros

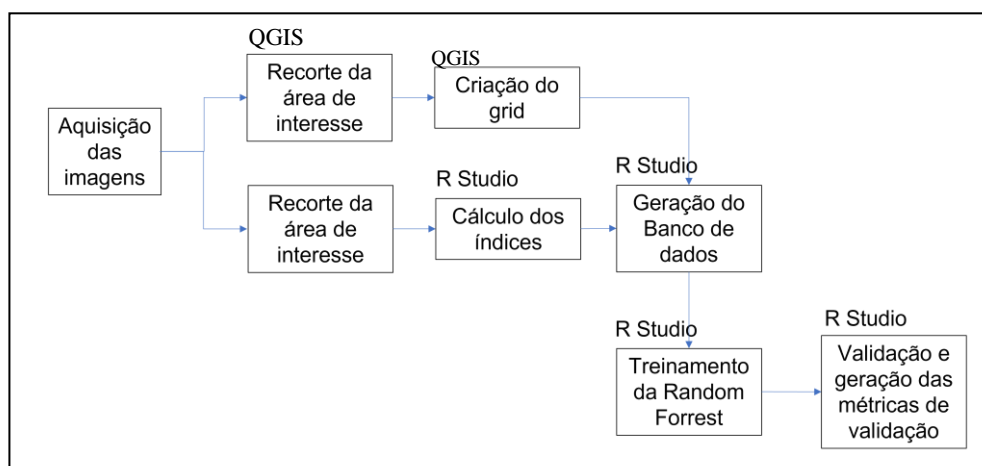
algoritmos preditivos, possuir menos ruídos e apresentarem mais estabilidade nas taxas de erro.

Para analisar a pureza de um nó de cada árvore que serão utilizadas, o índice Gini tem sido o método mais aplicado por possibilitar a análise do quanto este nó consegue separar classes distintas, dessa maneira é possível verificar o ganho de informação para auxiliar na decisão dessas variáveis. Ao aplicar o índice pode-se diminuir o chamado overfitting através da análise do grau de impureza, já que mede a importância de cada um dos atributos (SHANG et al., 2007).

5 METODOLOGIA

As etapas metodológicas da pesquisa, de forma resumida para uma melhor visualização, podem ser observadas na Figura 4, cada uma das etapas será exemplificada nas subseções seguintes.

Figura 4- Metodologia.

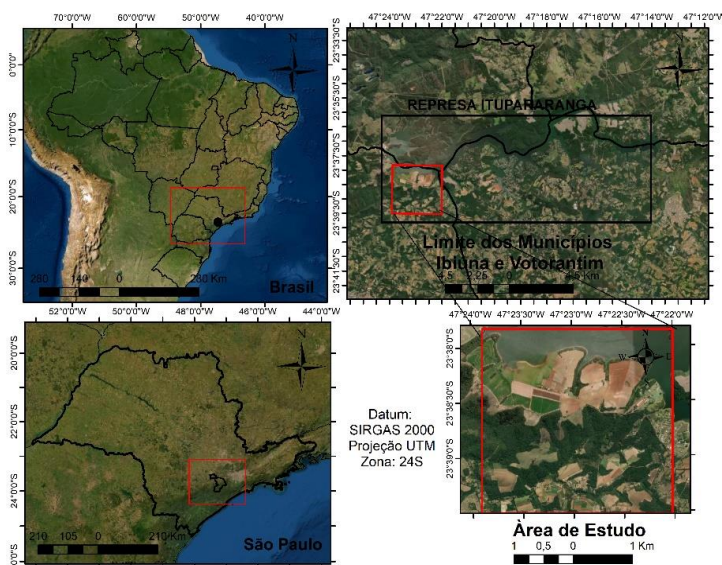


Fonte: Autoria própria.

5.1 CARACTERIZAÇÃO DA ÁREA DE ESTUDO

A área de estudo compreende no entorno da represa de Itupararanga localizada na bacia do Alto Sorocaba criada em 1912 pela Companhia de Energia Elétrica – Light sendo alimentada pelos rios Sorocabaçu, Sorocamirim e Una, Figura 5. Além de abastecimento a represa contribui com o abastecimento de energia elétrica e retém a água do Rio Sorocaba nos meses de cheia.

Figura 5 - Área de estudo.



Fonte: Autoria própria.

O clima da região de acordo com a classificação Koppen é descrito como subtropical úmido (Cwa), sendo a média das temperaturas nos meses mais quentes acima de 22°, uma estação seca e fria e outra quente e chuvosa. A média de pluviosidade anual no Alto Sorocaba é 1493mm, sendo o mês de janeiro o mais chuvoso e agosto o mês mais seco (SALLES et al., 2008).

A barragem foi criada para gerar energia elétrica e, atualmente, é responsável por abastecer as cidades de Votorantim, Sorocaba entre outras (PEREIRA, 2008). O lago da represa possui uma área de 29,9 km² e profundidade média de 7,8 metros (GIRON, 2013).

5.2 AQUISIÇÃO DE DADOS

As imagens utilizadas neste presente trabalho foram obtidas no dia 04 de julho de 2021, por volta das 13:40h pelo Satélites CBERS 4A e baixadas do site <http://www.dgi.inpe.br/catalogo/>, de forma gratuita, possuindo quatro bandas espectrais no sensor Pancromático e Multiespectral (WPAN) de acordo com a Tabela 5. O tempo de revisita para Câmera Pancromática e Multiespectral (PAN) é de 26 dias.

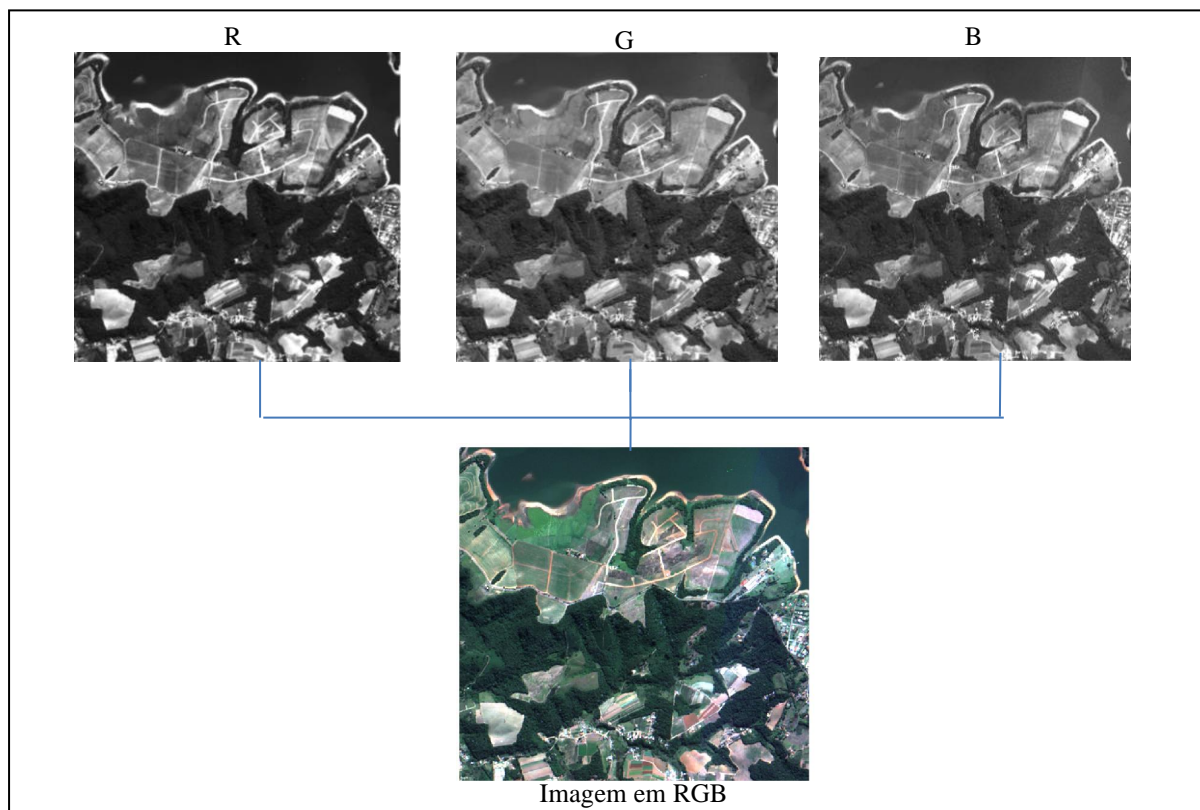
Tabela 4- Descrição das bandas espectrais do sensor WPAN do CBERS 4A

Sensor	Bandas espectrais	Resolução espectral	Resolução espacial
WPAN	PAN	0,51-0,85µm	2m
	G	0,52-0,59µm	
	B	0,45-0,52µm	
	R	0,63-0,69µm	8m
	NIR	0,77-0,89µm	

Fonte: Adaptado de INPE, 2019d.

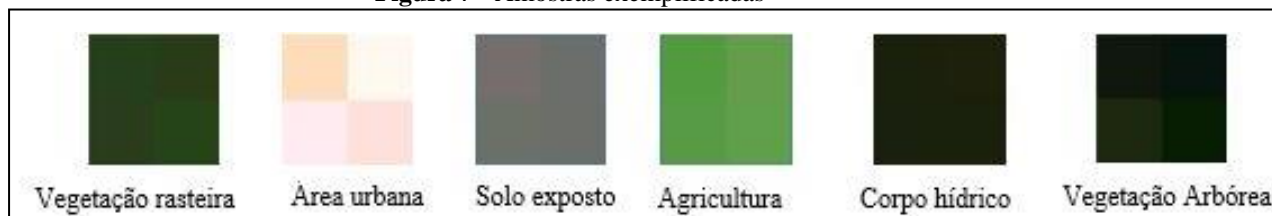
5.3 AMOSTRAGEM DE DADOS E CÁLCULO DE ÍNDICES

A coleta das amostras para treinamento foi realizada no ambiente QGis 3.2.0 onde realizou-se a composição das bandas 1,2,3, respectivamente, Figura 6, fazendo uma composição na cor verdadeira, (cores reais), e coletadas amostras de forma supervisionada através de um gradeamento para que essas possuíssem as mesmas dimensões.

Figura 6- RGB e composição de banda

Fonte: Autoria Própria.

Sendo assim, o banco de dados foi criado com 50 amostras das classes que mais se destacavam na composição de imagem sendo elas: agricultura, solo exposto, corpo hídrico, área urbanizada, vegetação arbórea e vegetação rasteira. Para o treinamento foram coletadas 250 amostras, com mesma dimensão, 2x2, Figura 7.

Figura 7 - Amostras exemplificadas

Fonte: Autoria Própria.

Na Figura 8 é possível observar os locais onde foram coletados os pixels que corresponde a cada uma das classes escolhidas para análise e após isso, salvo em formato shapefile.

Figura 8 - Coleta de amostras

Fonte: Autoria Própria.

O cálculo dos índices e a aplicação da Random Forest, foram realizadas no software Rstudio que possui linguagem e ambiente desenvolvidos para a computação gráfica e estatística, inúmeros meios de utilização e de fácil acesso por ser gratuito, além de possuir vários pacotes operacionais, bibliotecas diversificadas e multidisciplinares, valendo ainda ressaltar que tem sido utilizado em análises espaciais e na geocomputação (GUO et al., 2016).

No primeiro momento foi realizado o cálculo dos índices de acordo com o trabalho de Hamuda, Glavin e Jones (2016) já mencionados anteriormente, produzindo 11 (onze) produtos (imagens) referentes aos valores dos índices espectrais mais o NDVI.

5.4 IMPLEMENTAÇÃO DA RANDOM FOREST (RSTUDIO)

A implementação da RF, foi realizada no software Rstudio 3.6, para isso foram necessárias a utilização de vários pacotes disponibilizados gratuitamente na biblioteca do software “R”, abaixo estão detalhadas as etapas seguidas para a implementação da RF.

1º Etapa

A partir da instalação dos pacotes mencionados na Tabela 5, as bandas espectrais e os shapes files (elaborados no Qgis) foram importadas para o R e realizou-se o recorte da área de interesse.

Tabela 5 - Pacotes utilizados Rstudio

PACOTE	FUNÇÃO
raster	Possibilita a manipulação de dados em formato <i>raster</i> , com objetos da classe <i>RasterLayer</i>
sp	Fornecer classes e métodos que possibilita trabalhar com dados espaciais no R
rgdal	Permite a manipulação de dados espaciais
rgeos	Opera dados vetoriais
rgplot2	Pacote para produção de gráficos complexos.
dplyr	Permite a manipulação de dados especificamente, bastante utilizado em dados de machine learning
reshape2	Possibilita a organização de dados
randomForest	Implementa o algoritmo RF de Breiman
e1071	Fornecer funções para algoritmos estatísticos e probabilísticos.
caret	Desenvolvido para treinamento de modelo de classificação e regressão
caTools	Dentre as várias funções que este pacote apresenta, a utilizada para esta pesquisa foi: leitura e gravação para arquivos binários.
prettymapr	Utilizado para incrementar efeitos gráficos em mapas, como seta de norte e escalas gráficas.

Fonte: Autoria Própria.

2º Etapa

Nesta etapa foram realizados os cálculos dos índices baseados no trabalho de Hamuda, Glavin e Jones (2016) e do NDVI, já mencionados anteriormente.

3º Etapa

Gerou-se o banco de dados. a partir do shape file com as amostras coletadas, foram exportadas e realizado o recorte e atribuídos os rótulos para cada classe: “agricultura”, “area_urbanizada”, “corpo_hidrico”, “solo_exposto”, “vegetacao” e “vegetacao rasteira. A partir disso foi realizado o treino e a validação, onde, 70% treinamento e 30% para validação.

4º Etapa

Nesta etapa implementou -se a RF e realizado os ajustes na quantidade de árvores: $n_{tree} = 100$ e $mtry = 3$. Daí então foram calculadas as métricas de validação: Matriz de confusão, índice Kappa e Descritores de importância GINI.

6 RESULTADOS OBTIDOS

Nas próximas seções serão apresentados os resultados obtidos após a aplicação dos processos metodológicos.

6.1 QUALITATIVO

A Figura 9, apresenta o resultado da composição de banda na cor verdadeira, RGB nas bandas 1, 2 e 3, respectivamente, sendo possível observar as distintas classes de cobertura do solo ao redor da represa de Itupararanga e compará-las com as imagens resultantes dos cálculos de índices.

Figura 9 - Resultado da composição da imagem na cor verdadeira



Fonte: Autoria Própria.

A Figura 10 apresenta os resultados dos cálculos dos índices baseados nos canais de cor RGB e do NDVI, a partir do reescalonamento para o intervalo [0, 255].

Figura 10 - Imagens resultantes dos Índices; (a) NDI, (b) NGRDI, (c) MExG, (d) ExR, (e) VEG, (f) ExG (g) ExGR, (h) CIVE, (i) COM1, (j) COM2, (k) NDVI.

(Continua)

(a)



(b)



Figura 10 - Imagens resultantes dos Índices; (a) NDI, (b) NGRDI, (c) MExG, (d) ExR, (e) VEG, (f) ExG (g) ExGR, (h) CIVE, (i) COM1, (j) COM2, (k) NDVI.

(continuação)

(c)



(d)

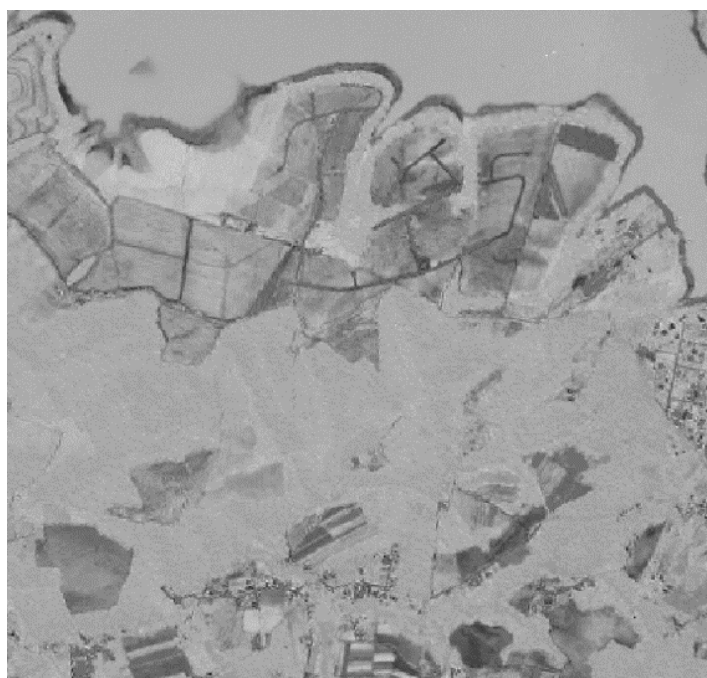


Figura 10 - Imagens resultantes dos Índices; (a) NDI, (b) NGRDI, (c) MExG, (d) ExR, (e) VEG, (f) ExG (g) ExGR, (h) CIVE, (i) COM1, (j) COM2, (k) NDVI.

(continuação)

(e)



(f)



Figura 10 - Imagens resultantes dos Índices; (a) NDI, (b) NGRDI, (c) MExG, (d) ExR, (e) VEG, (f) ExG (g) ExGR, (h) CIVE, (i) COM1, (j) COM2, (k) NDVI.

(g)

(continuação)



(h)



Figura 10 - Imagens resultantes dos Índices; (a) NDI, (b) NGRDI, (c) MExG, (d) ExR, (e) VEG, (f) ExG (g) ExGR, (h) CIVE, (i) COM1, (j) COM2, (k) NDVI.

(i)

(continuação)



(j)

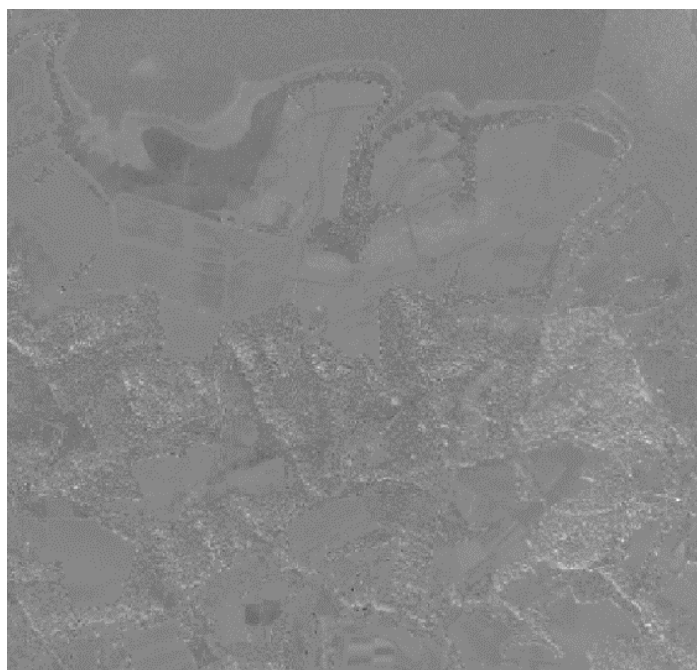


Figura 10 - Imagens resultantes dos Índices; (a) NDI, (b) NGRDI, (c) MExG, (d) ExR, (e) VEG, (f) ExG (g) ExGR, (h) CIVE, (i) COM1, (j) COM2, (k) NDVI.

(conclusão)

(k)



Fonte: Autoria Própria.

Analisando os resultados apresentados na Figura 10, é possível observar que a maioria dos índices possuem proximidade nos valores dos pixels e, por esse motivo, pouco contraste entre classes de cobertura do solo.

Alguns índices destacam regiões de banco de areia no entorno da represa, observados nas Figuras 10(a), 10(b), 10(e) e 10(k), indicando que a represa está no período de estiagem que normalmente ocorre nos meses de maio a julho.

O índice VEG, Figura 10(e), embora apresente valores de pixels bem próximos, permite distinguir vegetação robusta, solo exposto e região hídrica, quando comparada com a composição de imagem Figura 6.

O ExR, Figura 10(d), apresenta um melhor contraste entre as regiões, porém com pequena diferença entre as classes devido a baixa escala dinâmica de tons de cinza na imagem final após o reescalonamento.

Nota-se também que o MExG, Figura 10(c) apresentou um baixo contraste que visualmente dificulta a distinção de classes. O mesmo ocorre com o índice COM1 e COM2, Figura 10(f) e 10(j).

O resultado do NDVI, Figura 10(k), foi o que apresentou melhor contraste e, dessa forma, permite distinguir as classes, destacando bastante bem regiões de vegetação robusta.

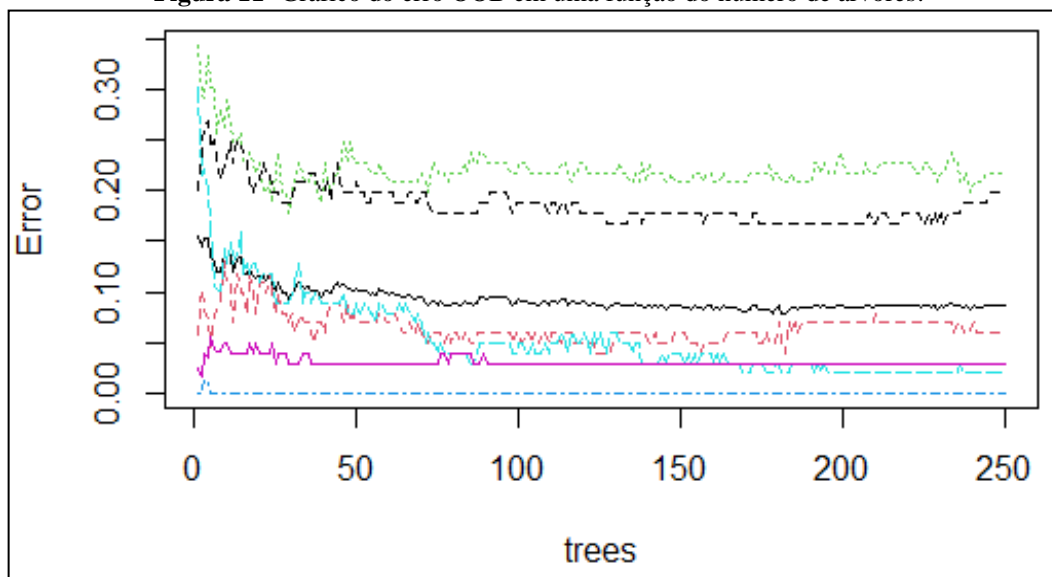
6.2 RESULTADOS DA RANDOM FOREST

➤ Seleção de parâmetros para o modelo

Conforme mostrado no estudo de Breiman (2001), a taxa de erro *out-of-bag* (OOB) é um estimador utilizado para cálculo do erro em função do número de árvores. Sendo assim, OOB como um procedimento imparcial foi usado para selecionar os melhores parâmetros da RF. (GOETZ et al., 2015, TRIGILA et al., 2015).

Conforme apresentado na Figura 11, o erro OOB, para cada uma das classes, é uma função do número de árvores, portanto, reduzido quando mais árvores são adicionadas ao algoritmo de RF. Com base nessa análise, com OOB igual a 0,229, o número de variáveis amostradas aleatoriamente como candidatas em cada divisão e o número de árvores foram obtidos como sendo 3 e 250, respectivamente.

Figura 11- Gráfico do erro OOB em uma função do número de árvores.



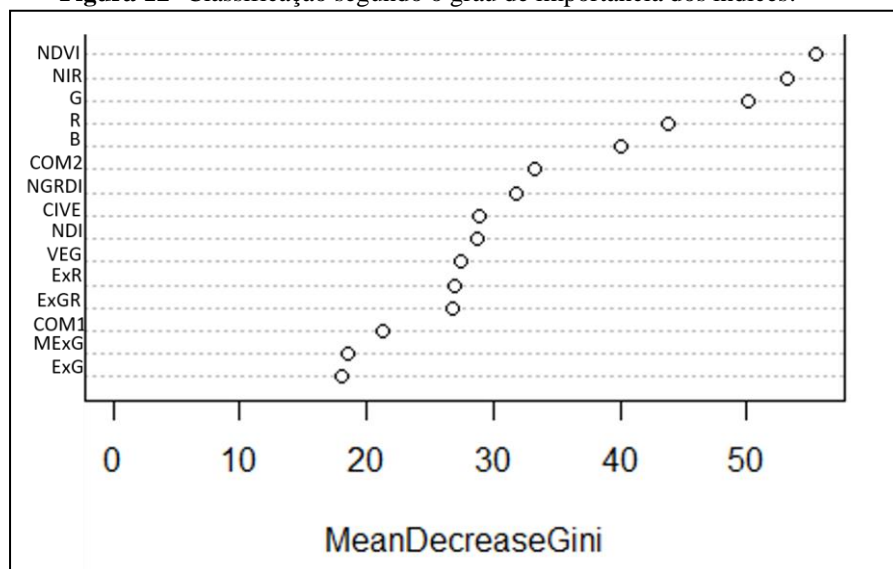
Fonte: Autoria Própria.

Variáveis independentes

A Figura 12 mostra a classificação dos índices por grau de importância baseados no índice GINI, tanto os dos canais visíveis, bem como o NDVI, além canais RGB e NIR, de maneira independente.

É possível observar que o canal NDVI foi o que apresentou um resultado de maior relevância, seguido do NIR. Os índices combinados (COM 1 e COM 2) e o ExG apresentaram menor desempenho quando utilizados na RF.

Figura 12- Classificação segundo o grau de importância dos índices.



Fonte: Autoria Própria.

➤ Classificação com a RF

Os resultados da classificação da RF foram analisados por comparação visual com a composição em cor verdadeira apresentada na Figura 9.

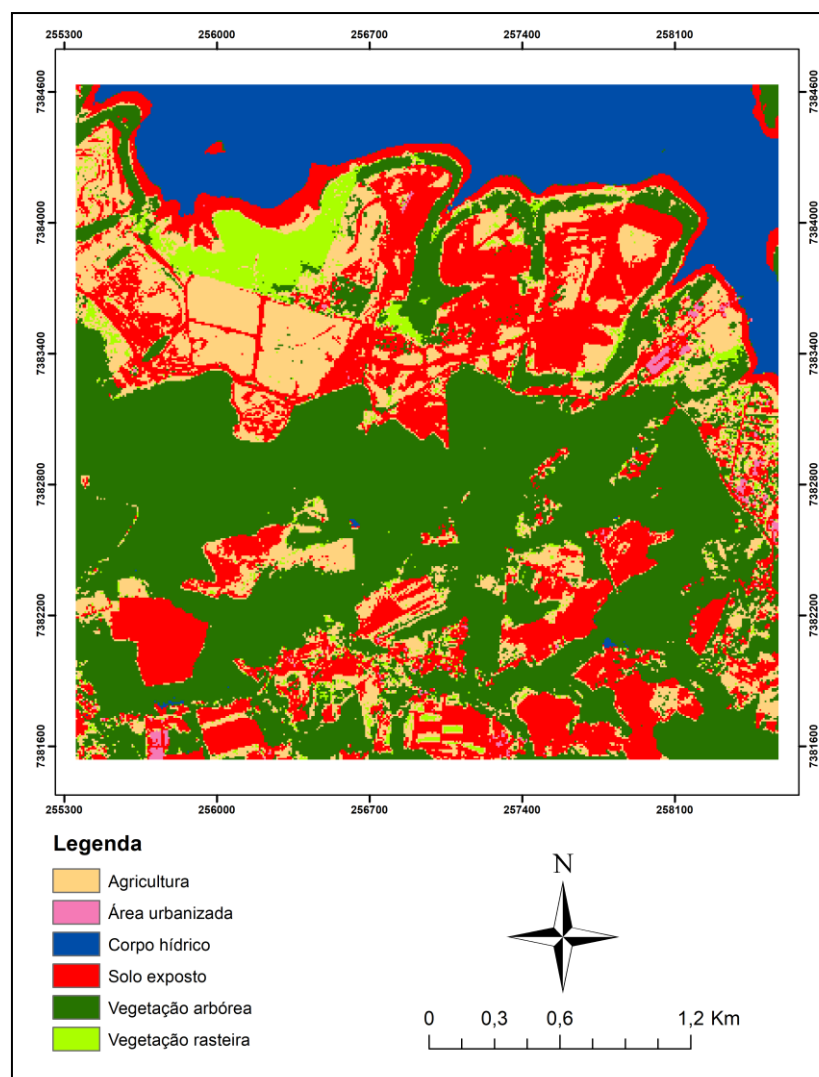
Utilizando o banco de imagens das classes, no processo de treinamento a RF obteve coeficiente Kappa igual a 0.993.

A Figura 13 apresenta os resultados da classificação com a RF após o treinamento com o banco de imagens das amostras das classes sendo consideradas. As regiões de corpo hídrico e vegetação arbórea, são as que mais se destacaram na classificação, assim como as regiões de agricultura.

Já no caso das regiões de área urbana, houve dificuldade na obtenção das amostras de treinamento, visto que, eram áreas muito pequenas e heterogêneas, refletindo no processo de classificação.

Deve-se também salientar que as regiões da classe “agricultura” podem ser classificadas como regiões das classes solo exposto ou vegetação rasteira dependendo do mês em que as imagens são adquiridas, de acordo com o estágio do processo de plantio.

Figura 13 - - Resultado da classificação com a RF.



Fonte: Autoria Própria.

Observa-se ainda da Figura 13 que que áreas da classe vegetação arbórea e corpo hídrico obtiveram ótimos resultados, o que pode ser explicado pelo fato de serem as maiores áreas presentes na imagem o que facilitou no momento de coletas de amostras para o banco de dados de treinamento.

7 CONCLUSÃO

Comparando os resultados dos índices estudados, o NVDI apresenta o melhor contraste no caso de imagens do satélite CBERS 4A para as coberturas do solo presentes na região estudada.

O processo de treinamento da RF com o banco de imagens das classes consideradas resultou em um índice kappa de 0,993, o que possibilitou um bom desempenho no processo de classificação de todos os pixels na imagem completa.

Por comparação visual com a imagem em cor verdadeira, pode-se notar que a RF se apresenta como um bom classificador no caso das imagens do CBERS 4A.

É importante destacar que a dificuldade em se obter amostras de uma dada classe, como ocorreu com as regiões de área urbanizada, leva a um pior desempenho na classificação pela RF. Quando é possível se obter amostras representativas e em bom número, os resultados são melhores, como no caso das classes corpo hidrico e vegetação arborea.

REFÊRENCIAS

BELGIU, M.; DRĂGUȚ, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS journal of photogrammetry and remote sensing**, v. 114, p. 24-31, 2016.

BERTONI, J.; LOMBARDI NETO, F. **Conservação do solo**. 4. ed. São Paulo, Ícone, p.355, 1999.

BEZERRA, J. M.; et al. Parâmetros biofísicos obtidos por sensoriamento remoto em região semiárida do estado do Rio Grande do Norte, Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**. v. 18, n. 1, p. 73 – 84, 2014.

BLASCHKE, T. Object based image analysis for remote sensing. **ISPRS journal of photogrammetry and remote sensing**, v. 65, n. 1, p. 2-16, 2010.

BRAZ, A. M.; ANDRADE, T.; MIRANDOLA, P. H. Análise de índices de vegetação NDVI e SAVI e índice de área foliar (IAF) para a comparação da cobertura vegetal na bacia hidrográfica do Córrego Ribeirãozinho, município de Selvíria–MS. **Revista Percorso**, Maringá, v. 7, n. 2, p. 5-22, 2015.

BREIMAN, L. Bagging predictors. **Machine learning**, v. 24, p. 123-140, 1996.

BREIMAN, L. Random forests. **Machine learning**, v. 45, p. 5-32, 2001.

BURGOS-ARTIZZU, X. P. et al. Real-time image processing for crop/weed discrimination in maize fields. **Computers and Electronics in Agriculture**, v. 75, n. 2, p. 337-346, 2011.

CHENG, G. et al. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 13, p. 3735-3756, 2020.

CRUVINEL, P. E. et al. An automatic method based on image processing for measurements of drop size distribution from agricultural sprinklers. **Anais Do IX Sibigrapi**, v. 3, p. 39-46, 1996.

CRUZ, A. C. **Estimativa da densidade de plantação de cana de açúcar utilizando o espaço de cor cie lab em imagens de alta resolução espacial providas de Vants.** Dissertação (Mestrado em Ciências Ambientais) - Instituto de Ciência e Tecnologia de Sorocaba , Universidade Estadual Paulista, Sorocaba. 2020.

CUNHA, J. E. et al. Dinâmica da cobertura vegetal para a Bacia de São João do Rio do Peixe, PB, utilizando-se sensoriamento remoto. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 16, n. 5, p. 539-548, 2012.

LUCA, G. et al. Integrated use of Sentinel-1 and Sentinel-2 data and open-source machine learning algorithms for land cover mapping in a Mediterranean region. **European Journal of Remote Sensing**, v. 55, n. 1, p. 52-70, 2022.

CUNHA, A. P. M. A. Impactos dos parâmetros biofísicos da vegetação derivados do sensor MODIS/TERRA nas simulações do balanço de energia em área de caatinga. **Ciência e Natura**, Santa Maria, v. 31, p. 109-112, 2009.

FAN, Y. et al. Digital image colorimetry on smartphone for chemical analysis: A review. **Measurement**, v. 171, p. 108829, 2021.

FIGUEIREDO, D. **Conceitos básicos de sensoriamento remoto.** São Paulo, 2005.

FORKEL, M. et al. Trend change detection in NDVI time series: Effects of inter-annual variability and methodology. **Remote Sensing**, v. 5, n. 5, p. 2113-2144, 2013.

FRACETO, L. F.; LOURENÇO, R. **Diagnóstico ambiental e avaliação de uso e ocupação do solo visando a sustentabilidade da represa de Itupararanga, importante área da bacia do médio Tietê.** In: POMPEO, Marcelo; **São Paulo: Institute of Biosciences, University de São Paulo**, p. 212-231, 2015.

GOETZ, J. N. et al. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. **Computers & geosciences**, v. 81, p. 1-11, 2015.

GONZALEZ, R. C.; WOODS, R. E. **Processamento de imagens digitais.** Editora Blucher, 2000.

GUERRERO, J. M. et al. Support vector machines for crop/weeds identification in maize fields. **Expert Systems with Applications**, v. 39, n. 12, p. 11149-11155, 2012.

GUIJARRO, M. et al. Automatic segmentation of relevant textures in agricultural images. **Computers and Electronics in Agriculture**, v. 75, n. 1, p. 75-83, 2011.

GUO, D.; WESTRA, S.; MAIER, H. R. An R package for modelling actual, potential and reference evapotranspiration. **Environmental Modelling & Software**, v.78, p.216–224, 2016.

HAGUE, T.; TILLET, N. D.; WHEELER, H. Automated crop and weed monitoring in widely spaced cereals. **Precision Agriculture**, v. 7, n. 1, p. 21-32, 2006.

HAMILTON, C. Consumerism, self-creation and prospects for a new ecological consciousness. **Journal of cleaner production**, v. 18, n. 6, p. 571-575, 2010.

HAMUDA, E.; GLAVIN, M.; JONES, E. A survey of image processing techniques for plant extraction and segmentation in the field. **Computers and Electronics in Agriculture**, v. 125, p. 184-199, 2016.

HAYKIN, S. **Neural Networks: a comprehensive foundation**. 2 ed. Upper Saddle River: Prentice-Hall. 2004.

HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote sensing of environment**, v. 25, n. 3, p. 295-309, 1988.

HUNT, E. R. et al. Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. **Precision Agriculture**, v. 6, p. 359-378, 2005.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Câmeras Imageadoras CBERS 04A**. 2019a. Disponível em:
<http://www.cbers.inpe.br/sobre/cameras/cbers04a.php>. Acesso em: 13 abr. 21.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Câmeras Imageadoras CBERS 04A**. 2019b. Disponível em:
<http://www.cbers.inpe.br/sobre/cameras/cbers04a.php>. Acesso em: 10 mar. 2020.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Câmeras Imageadoras CBERS 04A**. 2019c. Disponível em:
<http://www.cbers.inpe.br/sobre/cameras/cbers04a.php>. Acesso em: 10 mar. 2020.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Câmeras Imageadoras CBERS 04A**. 2019d. Disponível em:
<http://www.cbers.inpe.br/sobre/cameras/cbers04a.php>. Acesso em: 10 mar. 2020.

KATAOKA, T. et al. Crop growth estimation system using machine vision.
 In: **Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)**. IEEE, v.2, p. b1079-b1083 2003.

LAMM, R. D.; SLAUGHTER, D. C.; GILES, D. K. Precision weed control system for cotton. **Transactions of the ASAE**, v. 45, n. 1, p. 231, 2002.

LEITE, M. R. **Sensoriamento remoto aplicado à análise temporal da relação uso da terra / temperatura e albedo de superfície na bacia do rio vieira no norte de minas gerais**. Dissertação (Mestrado em Geografia) - Universidade Federal de Uberlândia. Uberlândia – MG. 111p .2011.

LEVIN, N. **Fundamentals of remote sensing. 1st Hydrographic Data Management Course**, IMO—International Maritime Academy, Trieste, v. 76, 1999.

LOPES, H. L. et al. Parâmetros biofísicos na detecção de mudanças na cobertura e uso do solo em bacias hidrográficas. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 14, n. 11, p. 1210-1219, 2010.

LUPPE, W. Kallimachos, Aitien-Prolog V. 7-12. **Zeitschrift für Papyrologie und Epigraphik**, p. 50-54, 1997.

MA, L. et al. A review of supervised object-based land-cover image classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 130, p. 277-293, 2017.

MARTINS, T. F. G. et al. Uso e ocupação do solo, heterogeneidade espacial de matéria orgânica, frações granulométricas e complexação de metais em sedimentos de reservatório. **Acta Limnologica Brasiliensia**, v. 33, 2021.

MAULIK, U.; CHAKRABORTY, D. Remote Sensing Image Classification: A survey of support-vector-machine-based advanced techniques. **IEEE Geoscience and Remote Sensing Magazine**, v. 5, n. 1, p. 33-52, 2017.

MENESES, P. R.; ALMEIDA, T. de. Introdução ao processamento de imagens de sensoriamento remoto. **Universidade de Brasília, Brasília**, 2012.

MEYER, G. E.; NETO, J. C. Verification of color vegetation indices for automated crop imaging applications. **Computers and electronics in agriculture**, v. 63, n. 2, p. 282-293, 2008.

MOTOHKA, T. et al. Applicability of green-red vegetation index for remote sensing of vegetation phenology. **Remote Sensing**, v. 2, n. 10, p. 2369-2387, 2010.

MUNDOGEO. **Satélite de sensoriamento remoto CBERS 4A é lançado com sucesso**. 2020. Disponível em: <https://mundogeo.com/2019/12/20/satelite-de-sensoriamento-remoto-cbers-4a-e-lancado-com-sucesso/>. Acesso em: 25 maio 2021.

NEMANI, R. R. et al. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. **science**, v. 300, n. 5625, p. 1560-1563, 2003.

OHTA, Y.; KANADE, T.; SAKAI, T. Color information for region segmentation. **Computer graphics and image processing**, v. 13, n. 3, p. 222-241, 1980.

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**. Dissertação (Mestrado em Bioinformática) - Universidade de São Paulo, Ribeirão Preto, 2013

PEREIRA, A. C. F. Desenvolvimento de método para inferência de características físicas da água associadas às variações espectrais. **Caso de Estudo: Reservatório de Itupararanga/SP**. 2008.

PIAO, S. et al. Interannual variations of monthly and seasonal normalized difference vegetation index (NDVI) in China from 1982 to 1999. **Journal of Geophysical Research: Atmospheres**, v. 108, n. D14, 2003.

PONZONI, F. J.; DISPERATI, A. A. Comportamento espectral da vegetação. **Sensoriamento remoto: reflectância de alvos naturais**, v. 1, p. 157-199, 2001.

PONZONI, F.J.; SHIMABUKURO, Y.E. A Imagem como fonte de dados radiométricos (abordagem quantitativa). In: PONZONI, F.J.; SHIMABUKURO, Y.E. (Org.). **Sensoriamento Remoto no estudo da vegetação**. São José dos Campos: Parêntese, cap.3, p.67-99, 2010.

PRASAD, P.j et al. Evaluation and comparison of the earth observing sensors in land cover/land use studies using machine learning algorithms. **Ecological Informatics**, v. 68, p. 101522, 2022.

QUEIROZ, R. P.; IMAI, N. N. Mapeamento das atividades antrópicas na área de entorno do reservatório de Itupararanga–SP: uma abordagem baseada na diminuição gradativa do grau de complexidade da cena interpretada. **Simpósio Brasileiro de Sensoriamento Remoto. Florianópolis: INPE**, p. 1034-1045, 2007.

RAMOS, R. R. D. et al. Determinação de variáveis biofísicas na bacia do rio Pajeú – Pernambuco, Brasil. **Anais XV Simpósio Brasileiro de Sensoriamento Remoto – SBSR**, Curitiba, PR, Brasil, 2011.

RANGAYYAN, R.M.; ACHA, B.; SERRANO, C. **Color image processing with biomedical applications**. Bellingham: SPIE Press, 2011.

RIBEIRO, A. et al. **Development of an image analysis system for estimation of weed pressure**. **Precision agriculture**, v. 5, p. 169-174, 2005.

RUSSELL, S.; NORVIG, P. **Inteligência artificial**. 3 ed, Rio de Janeiro: Elsevier, 2013

SALLES, M. H. D. et al. Avaliação simplificada de impactos ambientais na bacia do Alto Sorocaba (SP). **Revista de estudos ambientais**, v. 10, n. 1, p. 6-20, 2008.

SALLO, F. S. et al. Estimativa de componentes do balanço de radiação em diferentes tipos de uso e cobertura do solo. **Revista Ambiente & Água**, v. 9, p. 347-358, 2014.

SHANG, W. et al. A novel feature selection algorithm for text categorization. **Expert systems with applications**, v. 33, n. 1, p. 1-5, 2007.

SILVA, F. D. A. **Aplicação de técnicas de processamento de imagens na caracterização de materiais**. Tese (Doutorado em Engenharia Mecânica) - Universidade Estadual de Campinas, Campinas, 1996

SILVEIRA, N. T et al. Aplicação do Índice de Vegetação por Diferença Normalizada (NDVI) para a análise espaço-temporal da Bacia Hidrográfica do rio Terra Nova, Pernambuco, Brasil. **Revista Brasileira de Sensoriamento Remoto**, v. 3, n. 3, 2022.

SOONTRANON, N; SRESTASATHIERN, P; RAKWATIN, P. Rice growing stage monitoring in small-scale region using ExG vegetation index. In: **2014 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications And Information Technology (Ecti-Con)**. IEEE, p. 1-5, 2014.

TORRES-SÁNCHEZ, J. et al. Configuration and specifications of an unmanned aerial vehicle (UAV) for early site specific weed management. **PloS one**, v. 8, n. 3, p. e58210, 2013.

TRIGILA, A. et al. Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). **Geomorphology**, v. 249, p. 119-136, 2015.

TAN, P.; S., Michael; KUMAR, V. **Data mining introduction**. Pequim: Pearson, 2006. 796 p.

TUCKER, C. J.; TOWNSHEND, J. RG; GOFF, T. E. African land-cover classification using satellite data. **Science**, v. 227, n. 4685, p. 369-375, 1985.

WOEBBECKE, D.M. et al. Color indices for weed identification under various soil, residue, and lighting conditions. **Transactions of the ASAE**, v. 38, n. 1, p. 259-269, 1995.

WOEBBECKE, D. M. et al. Plant species identification, size, and enumeration using machine vision techniques on near-binary images. In: **Optics in Agriculture and Forestry**. International Society for Optics and Photonics, p. 208-219, 1993.

WU, D. et al. Time-lag effects of global vegetation responses to climate change. **Global change biology**, v. 21, n. 9, p. 3520-3531, 2015.

XIE, G.; NICULESCU, S. Mapping and Monitoring of Land Cover/Land Use (LCLU) Changes in the Crozon Peninsula (Brittany, France) from 2007 to 2018 by Machine Learning Algorithms (Support Vector Machine, Random Forest, and Convolutional Neural Network) and by Post-classification Comparison (PCC). **Remote Sensing**, v. 13, n. 19, p. 3899, 2021.

ZHANG, C. et al. An object-based convolutional neural network (OCNN) for urban land use classification. **Remote sensing of environment**, v. 216, p. 57-70, 2018.

ZHU, Zaichun et al. Greening of the Earth and its drivers. **Nature climate change**, v. 6, n. 8, p. 791-795, 2016.

ZURSSA, L. R.; MARTINS, A. C. G. Análise Automática do Uso do Solo no Entorno de Rodovias Usando uma Abordagem Fuzzy. **Trends in Computational and Applied Mathematics**, v. 23, p. 383-399, 2022.