

UNIVERSIDADE FEDERAL DO PAMPA

JULIE DIAS AGUIRRE

**IDENTIFICAÇÃO E CLASSIFICAÇÃO DE CORPOS D'ÁGUA EM IMAGEM
DIGITAL PELA APLICAÇÃO DO MÉTODO RANDOM FOREST**

**Itaqui
2023**

JULIE DIAS AGUIRRE

**IDENTIFICAÇÃO E CLASSIFICAÇÃO DE CORPOS D'ÁGUA EM IMAGEM DIGITAL
PELA APLICAÇÃO DO MÉTODO RANDOM FOREST**

Trabalho de Conclusão de Curso apresentado ao Componente Curricular TCC II do Curso de Engenharia Cartográfica e de Agrimensura da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Bacharel.

Orientador: Cristiano Galafassi

**Itaqui
2023**

Ficha catalográfica elaborada automaticamente com os dados fornecidos
pelo(a) autor(a) através do Módulo de Biblioteca do
Sistema GURI (Gestão Unificada de Recursos Institucionais) .

D284i Dias Aguirre, Julie

IDENTIFICAÇÃO E CLASSIFICAÇÃO DE CORPOS D'ÁGUA EM IMAGEM
DIGITAL PELA APLICAÇÃO DO MÉTODO RANDOM FOREST / Julie Dias
Aguirre.

41 p.

Trabalho de Conclusão de Curso(Graduação)-- Universidade
Federal do Pampa, ENGENHARIA CARTOGRÁFICA E DE AGRIMENSURA,
2023.

"Orientação: Cristiano Galafassi".

1. Inteligência artificial. 2. Landsat 8. 3. Random Forest.
4. Sensoriamento Remoto. I. Título.

JULIE DIAS AGUIRRE

**IDENTIFICAÇÃO E CLASSIFICAÇÃO DE CORPOS D'ÁGUA EM IMAGEM
DIGITAL PELA APLICAÇÃO DO MÉTODO RANDOM FOREST**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia Cartográfica e de Agrimensura da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Bacharel.

Trabalho de Conclusão de Curso defendido e aprovado em: 07 de julho de 2023.

Banca examinadora:

Prof. Dr. Cristiano Galafassi
Orientador
UNIPAMPA – Itaqui/RS

Prof. Dr. Sidnei Luís Bohn Gass
UNIPAMPA – Itaqui/RS

Prof. Dra. Fabiane Flores Penteado Galafassi
UNIPAMPA – Itaqui/RS

Dedico este trabalho ao meu Pai Jorge Moisés Morais Aguirre (in memoriam), motivo este que me permitiu chegar até aqui, na conquista deste Título.

AGRADECIMENTO

Agradeço primeiramente a Deus, que diante de todos os obstáculos foi minha força maior para seguir a realização deste trabalho e a Nossa Senhora Aparecida que incansavelmente nos dias mais difíceis foi o meu amparo.

Agradeço à minha mãe (Ana Maria Dias Aguirre) que em todo o período da graduação foi meu apoio diário para que esse sonho se tornasse possível. Aos meus familiares que de uma forma ou outra estiveram presentes contribuindo nos meus estudos.

Agradeço imensamente ao Prof. Dr. Cristiano Galafassi que aceitou a responsabilidade em me orientar, pelas inúmeras vezes que foi paciente ao sanar dúvidas, pela confiança em meu trabalho, pela dedicação e todos os aprendizados durante esse período.

Aos meus colegas, que assim como eu, se mantiveram firmes nessa jornada acadêmica possibilitando a troca de experiências e criando laços de amizades/profissionais.

LISTA DE FIGURAS

Figura 1: Varredura de um imageador orbital multiespectral.....	15
Figura 2: Processo decisório de múltiplas árvores pelo método de Random Forest.	18
Figura 3: Representação das bacias hidrográficas do RS com destaque a área de estudo.	21
Figura 4: Estrutura do 10-Fold Cross Validation.....	24
Figura 5: Representação do funcionamento da matriz de confusão.	25
Figura 6: Índice NDWI gerado através de imagem do satélite Landsat-8, em 11 de junho de 2017.....	27
Figura 7: Máscara resultante do modelo gerado através da classificação do Random Forest com o conjunto de dados para treinamento referente às Bandas 4, 3 e 2.	29
Figura 8: Composição colorida da Bacia Hidrográfica do Butuí nas Bandas 4, 3 e 2 de imagem do satélite Landsat-8, em 11 de junho de 2017.....	30
Figura 9: Máscara resultante do modelo gerado através da classificação do Random Forest com o conjunto de dados para treinamento referente às Bandas 6, 5 e 4.	32
Figura 10: Composição colorida da Bacia Hidrográfica do Butuí nas Bandas 6, 5 e 4 de imagem do satélite Landsat-8, em 11 de junho de 2017.....	33
Figura 11: Máscara resultante do modelo gerado através da classificação do Random Forest com o conjunto de dados para treinamento referente às 7 Bandas.	35

LISTA DE QUADROS

Quadro 1: Características do imageador LandSat 8.	15
Quadro 2: Matriz de Confusão com a aplicação das Bandas 4 3 2 através do modelo RF.	28
Quadro 3: Matriz de Confusão com a aplicação das Bandas 6, 5 e 4 através do modelo RF.	31
Quadro 4: Matriz de Confusão com a aplicação de Todas as Bandas através do modelo RF.....	34

LISTA DE SIGLAS

RF. – random forest

IA. – Inteligência artificial

SR. – sensoriamento remoto

RGB. – red, green e blue

OLI - *Operational Land Imager*

NDWI - *Normalized Difference Water Index*

CV – *Cross Validation*

RESUMO

Bacias hidrográficas tem um alto índice de suscetibilidade para que aconteçam inundações, pois estão associadas às questões climáticas e ambientais da natureza. Inundação é uma atribuição a áreas que estão submersas a água, consequência dos fenômenos temporais que agem de maneira incontrolável. Para isso são necessárias previsões climáticas que atuem no alerta sobre danos e prejuízos em áreas que, possivelmente, seriam atingidas com as inundações. Existem diversos métodos que executam esse tipo de análise em áreas inundadas, um deles é o Random Forest. O Random Forest funciona como um modelo de classificação e regressão, nesse caso o objetivo do trabalho foi realizar a identificação da massa d'água nas imagens digitais da Bacia Hidrográfica do Butuí, situada na região de São Borja – RS, obtidas pelo satélite artificial Landsat 8.

Palavras-Chave: Inteligência artificial. Imagens multiespectrais. Landsat 8.

ABSTRACT

Watersheds have a high susceptibility index for floods to happen, as they are associated with climatic and environmental issues of nature. Inundation is an attribution to areas that are submerged in water, a consequence of temporal phenomena that act in an uncontrollable way. For this, climate forecasts are necessary to act as a warning about damages and losses in areas that would possibly be affected by floods. There are several methods that perform this type of analysis in flooded areas, one of them is the Random Forest. The Random Forest works as a classification and regression model, in this case the objective of the work is to carry out the identification of the water mass in the digital images of the Butuí Hydrographic Basin, located in the region of São Borja - RS, obtained by the artificial satellite Landsat 8.

Keywords: Artificial intelligence. Multispectral imaging. Landsat 8.

SUMÁRIO

1. INTRODUÇÃO.....	12
2. OBJETIVOS.....	13
3. REFERENCIAL TEÓRICO	13
3.1 SENSORIAMENTO REMOTO.....	13
3.2 RANDOM FOREST.....	17
3.3 TRABALHOS RELACIONADOS.....	19
4. MATERIAIS E MÉTODOS	20
4.1 K – FOLD CROSS VALIDATION E A ACURÁCIA DO MODELO	23
5. RESULTADOS E DISCUSSÕES.....	26
6. CONCLUSÃO	36
7. REFERÊNCIAS	38

1. INTRODUÇÃO

O sensoriamento remoto é uma área que utiliza técnicas para coletar informações sobre a superfície terrestre sem a necessidade de estar fisicamente presente no local. Esse tipo de tecnologia tem sido amplamente utilizado em diversas áreas, como em monitorar desastres naturais, prevenção de desmatamentos, identificação de áreas agrícolas e monitoramento de recursos hídricos.

Apesar de ser uma área em constante expansão, o sensoriamento remoto apresenta desafios. Um dos principais desafios é a identificação de objetos em imagens digitais, o que pode ser difícil e demandar muito tempo de trabalho. No entanto, a adoção de modelos de aprendizado de máquina tem se mostrado eficaz nesse sentido. O aprendizado de máquina pode auxiliar na classificação de objetos em imagens, na detecção de mudanças na superfície terrestre e na identificação de áreas de risco de desastres naturais.

Uma das técnicas de aprendizado de máquina mais utilizadas em sensoriamento remoto é o Random Forest (RF). Esse algoritmo é composto por um conjunto de árvores de decisão geradas a partir de subconjuntos aleatórios de dados. O Random Forest é capaz de lidar com grande quantidade de dados e apresenta alta precisão na classificação de objetos em imagens digitais. Além disso, a utilização de Veículos Aéreos Não Tripulados (VANT) tem facilitado a obtenção de informações sobre a superfície terrestre, reduzindo o tempo e custo de trabalho em campo.

Em resumo, a área de sensoriamento remoto tem se mostrado uma ferramenta poderosa para a obtenção de informações geográficas e monitoramento ambiental. Apesar dos desafios, a adoção de modelos de aprendizado de máquina, como o Random Forest, tem possibilitado a melhoria do processo de análise de dados e a identificação de objetos em imagens digitais. Com o avanço da tecnologia, espera-se que o sensoriamento remoto continue a evoluir, possibilitando a obtenção de informações cada vez mais precisas e abrangentes sobre a superfície terrestre.

2. OBJETIVOS

O objetivo geral deste trabalho é desenvolver e avaliar um modelo baseado em Random Forest para a detecção e classificação automatizada de corpos de água em imagens de satélite, visando contribuir para a identificação precisa e eficiente dessas características geográficas, fornecendo informações valiosas para estudos de recursos hídricos, monitoramento ambiental e planejamento urbano.

Sendo assim, apresentamos os objetivos específicos para o estudo:

- Criar e treinar modelos baseados em random forest, utilizando a biblioteca scikit-learn para segmentação de imagens digitais;
- Avaliar o desempenho do RF em relação a diferentes composições de imagens;
- Aferir a acurácia dos modelos quanto ao seu treinamento e discutir os resultados obtidos classificando a imagem da Bacia Hidrográfica do Butuí.

3. REFERENCIAL TEÓRICO

O estudo da classificação de massas d'água estabelece uma relação direta entre áreas da engenharia, como sensoriamento remoto e computação, que trabalham de forma paralela. Nesse caso, serão abordados, de forma teórica, conceitos que definem um breve contexto do assunto aplicado à identificação de objetos em imagens digitais através da aprendizagem de máquinas.

3.1 Sensoriamento Remoto

A área de sensoriamento remoto teve um avanço significativo nos últimos anos e aprimoraram muitas funções, uma delas é a obtenção de imagens digitais através de sensores que podem ser classificados em passivos ou ativos.

Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE, 2022) O sensoriamento remoto é a técnica de obtenção de informações acerca de um objeto, área ou fenômeno localizado na Terra, sem que haja contato físico com o mesmo. As informações podem ser obtidas através de radiação eletromagnética, gerada por fontes naturais (por meio de sensores passivos) ou por fontes artificiais (por meio de sensores ativos). São apresentadas na forma de imagens, sendo mais utilizadas,

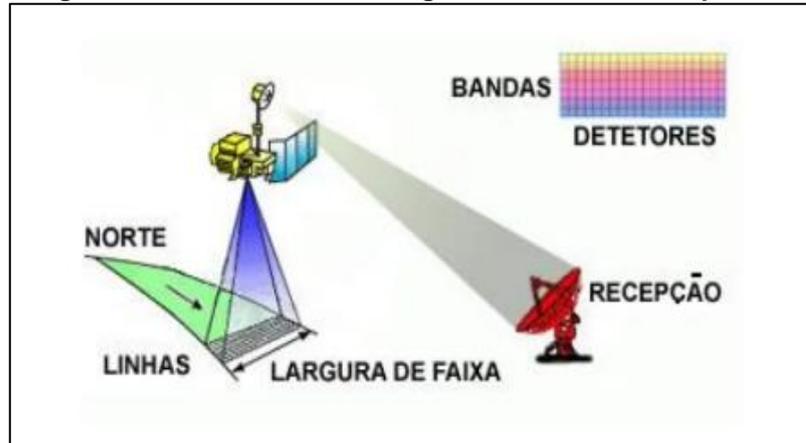
atualmente, aque-las captadas por sensores óticos orbitais localizados em satélites. Os satélites se movimentam em volta da Terra, e possuem um sensor que executa o trabalho de emitir e receber da Terra a energia eletromagnética.

As imagens digitais são representadas através de dados numéricos que compõem as bandas espectrais que o imageador detecta. Essas imagens são constituídas por elementos conhecidos como pixels, que abarcam informações essenciais como cor e posição.

Para a obtenção dessas imagens digitais, são empregados satélites artificiais, construídos pelo engenho humano e lançados em órbita circundando a esfera terrestre. Atribuem-se a tais satélites variadas incumbências, tais como a observação terrestre e a aquisição de dados. Dentre os satélites artificiais, destaca-se o Landsat 8, cujo mecanismo capturador de dados é o sensor OLI (Operational Land Imager), resultando num acréscimo de extrema relevância nas versões do Landsat dentro do âmbito do sensoriamento remoto. O sensor OLI possui a capacidade multiespectral para a captura das imagens, proporcionando resolução espacial de 30 metros em sete faixas espectrais distintas, abrangendo desde o espectro visível até o infravermelho próximo. É importante ressaltar que o Landsat 8 passa diariamente pelo mesmo ponto na superfície da Terra em horários semelhantes, o que assegura uma cobertura coerente e consistente.

Os satélites abrangem áreas terrestres no sentido norte-sul (Figura 1) com o auxílio da iluminação dos raios solares e realizam uma tarefa semelhante à rotação da Terra. Essa tarefa consiste em passar por uma região diferente a cada rotação, e depois de variados dias passa novamente nessa mesma região atualizando suas características. A Figura 1 apresenta de que maneira acontece a varredura de um imageador orbital com multibandas.

Figura 1: Varredura de um imageador orbital multiespectral.



Fonte: HOLLER, (2011).

Esse retorno é referido como "período de revisita", e sua relevância na área é ressaltada por fornecer novas perspectivas sobre os fenômenos temporais e possibilitar análises comparativas entre a região atual e a anterior. O Quadro 1 apresenta características sobre o satélite artificial Landsat 8, como as bandas existentes, o sensor que está acoplado para captura das informações espectrais e o tamanho das resoluções disponíveis.

Quadro 1: Características do imageador Landsat 8.

Sensor	Bandas Espectrais	Resolução Espectral	Resolução Espacial	Resolução Temporal	Área Imageada	Resolução Radiométrica
OLI (Operational Land Imager)	(B1) Azul Costeiro	0.43 - 0.45 μm	30 m	16 dias	170 x 183 km	16 bits
	(B2) AZUL	0.45 - 0.51 μm				
	(B3) VERDE	0.53 - 0.59 μm				
	(B4) VERMELHO	0.64 - 0.67 μm				
	(B5) INFRAVERMELHO PRÓXIMO	0.85 - 0.88 μm				
	(B6) INFRAVERMELHO MÉDIO	1.57 - 1.65 μm				
	(B7) INFRAVERMELHO MÉDIO	2.11 - 2.29 μm				
	(B8) PANCROMÁTICA	0.50 - 0.68 μm	15 m			
	(B9) CIRRUS	1.36 - 1.38 μm	30 m			

Fonte: Empresa EngeSat (2022) adaptada pelo autor.

As características do Landsat 8 resultam em uma boa qualidade na resolução das imagens, o que exige a aplicação de técnicas precisas para processar as informações e garantir uma análise acurada do conjunto de dados. O processamento de imagens ocorre por meio da utilização de algoritmos que visam aprimorar, analisar e extrair informações relevantes. Dentre as técnicas empregadas estão os filtros, a segmentação e o reconhecimento de padrões, entre outros.

Uma das técnicas é a utilização do Índice de Água de Diferença Normalizada (*Normalized Difference Water Index - NDWI*) para detectar corpos d'água dentre as imagens digitais. É um índice indicado para evidenciar corpos d'água de outras coberturas terrestres e pode ser aplicado em diversas plataformas de sensoriamento remoto, como imagens de satélite e imagens obtidas por drones.

Esse índice realiza cálculos através das bandas 3 e 5, nesse caso utilizando o satélite Landsat8, que correspondem respectivamente a banda verde visível (3) e a banda do infravermelho próximo (5).

Para calcular o NDWI, utiliza-se a equação 1:

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (\text{Eq. 1})$$

Onde "Green" é a banda verde (3) no satélite Landsat 8 e o "NIR" corresponde a banda do infravermelho próximo (5).

No item 3.3, serão abordados assuntos relacionados ao algoritmo Random Forest, uma ferramenta essencial para modelar um conjunto de treinamento. Os dados são obtidos através de imagens de satélites e indicam nas classificações os elementos existentes na imagem.

3.2 Random Forest

Breiman (2001) descreve que o Random Forest é uma técnica de aprendizagem de máquina, que se caracteriza por algoritmos que aprendem com dados. A aprendizagem de máquinas possibilita automatizar trabalhos e gerar modelos complexos que, se propostos de forma manual, demandam tempo e poderiam ter baixa eficiência. Dessa forma, a utilização do método Random Forest agiliza grande parte do processo uma vez que ele pode automaticamente classificar dados.

O Random Forest é amplamente reconhecido por sua habilidade de gerar resultados precisos por meio de técnicas de classificação. Especificamente na área do sensoriamento remoto, o método de classificação desempenha um papel crucial ao lidar com imagens de satélite. Esse algoritmo se destaca por sua capacidade de evitar o *overfitting*, ou seja, o ajuste excessivo aos dados de treinamento.

No contexto do sensoriamento remoto, a manipulação de dados frequentemente envolve grandes conjuntos de informações provenientes das imagens. Essa tarefa, embora desafiadora, é executada de maneira eficiente pelo algoritmo. Cho et al. (2021) há pouco tempo, os algoritmos de aprendizagem de máquina vêm se concretizando como atuações promissoras e transformadoras no processamento de dados de sensoriamento remoto.

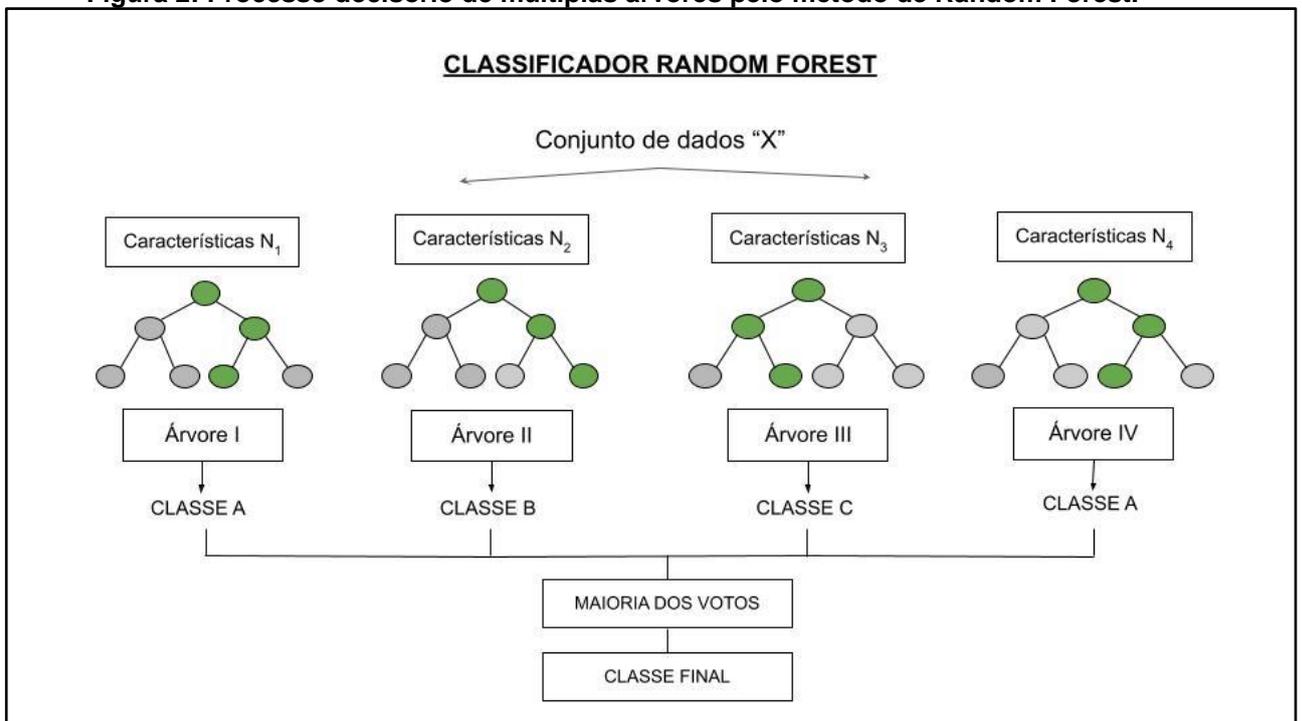
O algoritmo Random Forest gera múltiplas árvores de decisão independentes entre si, combinando seus resultados para obter uma saída final. Utilizando o conjunto de dados, são criadas amostras aleatórias conhecidas como amostras *bootstrap*, que são treinadas em cada árvore. Essas amostras *bootstrap* criam subconjuntos aleatórios a partir do conjunto de dados original, permitindo a ocorrência tanto de repetições quanto de exclusões de dados durante a geração dos novos subconjuntos.

Breiman (2001) propõe que após um grande número de árvores terem sido geradas, ocorra uma votação na classe mais popular. Esse procedimento é denominado florestas aleatórias. Segundo Monard e Baranauskas (2003), uma árvore de decisão apresenta uma base de dados definida como nó folha e nó de decisão, sendo o nó folha um tipo de classe e o de decisão um teste relacionado a um atributo.

Conforme o resultado do teste gera uma página para uma subárvore com uma estrutura semelhante à da árvore.

Na Figura 2 é representado o processo decisório de múltiplas árvores pelo método de random forest, onde um conjunto de dados "X" é distribuído em quatro características denominadas N_x . Dessa forma, é possível visualizar como funciona o treinamento do modelo junto dos dados de treinamento.

Figura 2: Processo decisório de múltiplas árvores pelo método de Random Forest.



Fonte: Elaboração Própria.

Como é possível observar na Figura 2 as características presentes no conjunto de dados "X" compõem as árvores possibilitando que o modelo indique as classes através da maioria dos votos. Após a votação, o algoritmo determina uma classe final com os resultados obtidos. Identifica-se que as características da primeira e última árvores apontam similaridades. Podendo associar que através da maioria dos votos, o algoritmo apresenta a Classe A como classe final.

No item 3.3 são apresentados trabalhos de diferentes autores que utilizaram o modelo Random Forest. Com a ideia de classificar estudos em áreas semelhantes à desse trabalho, apresentando a ampla área de utilização do aprendizado de máquina.

3.3 Trabalhos Relacionados

Os assuntos descritos nesta seção tratam sobre classificação de imagens digitais, aprendizagem de máquina, utilização de software e os diversos métodos, como o índice de água por diferença normalizada (NDWI), o qual será utilizado na comparação com o método Random Forest.

Shiba et al. (2005) descreve um estudo que utiliza a aprendizagem de máquina, para classificar imagens digitais através do método de árvore de decisão, comparado com o método da máxima verossimilhança. Para processamento das imagens digitais, classificação e implementação de bibliotecas, os autores utilizaram o IMG (sistema de domínio público) como ferramenta baseada em algoritmos de árvore de decisão. A conclusão foi satisfatória para o sistema IMG, apesar de apontar deficiência e não recomendar a utilização da imagem gerada, para base de trabalho em campo.

Pereira et al. (2018), utilizou em outro trabalho o método NDWI dividindo-o em dois modelos: o McFeeteres e o Gao. Esses modelos foram calculados de diferentes maneiras, e serviram para demonstrar as áreas úmidas e de inundações que existem no Pantanal. Os resultados obtidos apresentaram eficiência em ambos os modelos, cada um satisfaz suas necessidades. No caso do modelo McFeeteres, a classificação mostrou-se eficiente para a água e a vegetação. O modelo Gao, com o auxílio das bandas espectrais 2 que corresponde ao infravermelho próximo (IVP), e 5, que corresponde ao infravermelho médio (SWIR) referentes ao sensor MODIS, teve maior desempenho no infravermelho próximo otimizando o funcionamento do comportamento espectral em um determinado intervalo.

Paiva et al. (2018), realizou o acompanhamento do comportamento da água, buscando desenvolver um script na linguagem python para fazer a aplicação nos dados. A implementação na linguagem feita pela biblioteca ArcPy, disponível no

software ArcGIS, foi base para os autores realizarem os cálculos pelo método NDWI. As imagens usadas foram extraídas do imageador LandSat 8. Ao obter os resultados, os autores geraram um mapa demonstrando a água existente na área de estudo, e a satisfação com o método NDWI na representação e boa qualidade do produto.

Rodrigues et al. (2017), ao analisar um conjunto de dados em um intervalo de dez anos, e fazer a extração imagens de satélite nesse mesmo período, realizaram o processamento separadamente de ambos os dados. Após o processamento, utilizaram o método NDWI para identificar as áreas de reservatório hídrico em estações de chuva e de seca. O método atuou de forma satisfatória, gerando resultados que facilitaram a visualização das divergências nas áreas, mostrando os locais que existe alagamento e fez uma análise de como essas áreas se comportam em diferentes períodos.

4. MATERIAIS E MÉTODOS

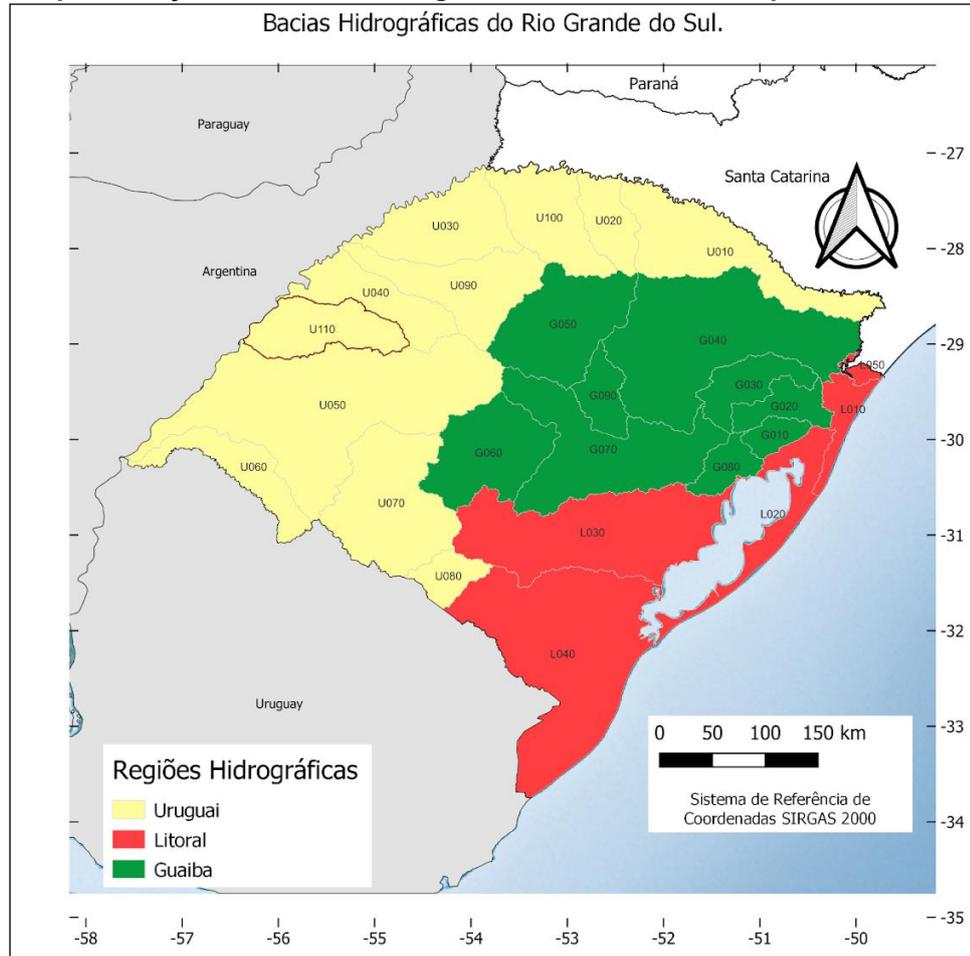
As bacias hidrográficas desempenham um papel crucial como fontes de água doce, recurso fundamental para a vida humana e os ecossistemas. É de extrema importância conscientizar a sociedade sobre a necessidade de manter a qualidade da água para seu uso adequado e responsável.

O estudo dos impactos negativos em uma bacia hidrográfica é essencial para o engajamento e o desenvolvimento de estratégias visando à conservação dos recursos hídricos. Potencializar a busca pela preservação desse recurso, pode qualifica-lo para o uso adequado futuramente.

A Bacia Hidrográfica do Butuí é composta pelos rios Icamaguã, Butuí e Piratinim, recebe o abastecimento do rio Uruguai. O rio Butuí desempenha um papel fundamental ao prover a região com recursos hídricos, sendo utilizado para diversas atividades, como consumo humano, agricultura e indústria.

A Figura 3 apresenta a localização da bacia hidrográfica do Butuí, na região de São Borja – RS. Na figura também estão situadas bacias que fazem divisa com a área de estudo e regiões hidrográficas que compõem o estado do Rio Grande do Sul.

Figura 3: Representação das bacias hidrográficas do RS com destaque a área de estudo.



Fonte: Base cartográfica vetorial (2018) adaptada pelo autor.

Apesar de suas vantagens, a Bacia Hidrográfica do Butuí enfrenta desafios e sofre prejuízos. A degradação ambiental ocorre de diversas formas, sendo a atividade agrícola uma delas. Do ponto de vista ambiental, a bacia hidrográfica desempenha um papel importante na proteção de áreas de preservação permanente, matas ciliares e biodiversidade.

Monitorar constantemente esse ecossistema é fundamental para garantir sua preservação. Dessa forma, o monitoramento pode ser feito através de imagens de satélites que disponibilizam os dados para treinamento. Esses dados são aplicados em técnicas de aprendizagem de máquinas como é o caso do Random Forest, possibilitando a classificação dos elementos existentes na imagem.

A imagem adquirida da região de São Borja apresenta a área em geral. Dentro da área pode ser encontrada a Bacia do Rio Butuí - Icamaquã que está localizada nas coordenadas geográficas na latitude 28° 30' e 29° 15' N e longitude 54° 00' e 56° 30' E. Existe um comitê responsável pela bacia criado pelo Decreto nº 44.401 de 18/04/2006 e é composto por um Presidente e um Vice-Presidente.

Com base nesse contexto, foram empregadas imagens do satélite Landsat 8 tendo órbita 224 e ponto 080 (cena 224/080), capturadas em 11 de junho de 2017. Foi feito o download de 7 das 9 bandas espectrais disponíveis do sensor OLI do satélite, ou seja, as bandas de 1 a 7.

Para desenvolver os modelos de RF foram utilizados algoritmos na linguagem python, com a biblioteca scikit learn e imagens digitais obtidas no site do INPE¹ - Catálogo com dimensões originais de 5608 x 2457 e resolução espacial de 30m. A extração de informações das imagens digitais foi feita no software QGIS, para definir as bandas espectrais, recortes de dados para treinamento e avaliar os resultados esperados do método.

Após feito o recorte dos dados da imagem, o processamento dos dados inicia-se pela construção do conjunto de dados para treinamento. Esses dados relacionam em torno de 200.000 pixels da imagem em “água” ou “não água”, classificados manualmente.

Em seguida, esses dados foram utilizados para treinar os modelos baseados em RF, utilizando o número de estimadores = 100 como valor padrão gerando uma matriz de confusão. Com os resultados obtidos da matriz de confusão é possível calcular a acurácia do modelo com base nos dados de treinamento.

Para implementação dos modelos foi utilizada a plataforma colaborativa do Google (Colab) para inserção de dados no modelo. O colab funciona de forma gratuita

¹ Imagens digitais adquiridas através do site do Instituto Nacional de Pesquisas Espaciais no item Divisão de Geração de Imagens. A imagem utilizada é do período de junho de 2017. Para mais informações acesse: <http://www.dgi.inpe.br/CDSR/>

para práticas de aprendizagem de máquina e IA utilizando a linguagem python, permitindo desenvolver e manipular códigos no ambiente virtual.

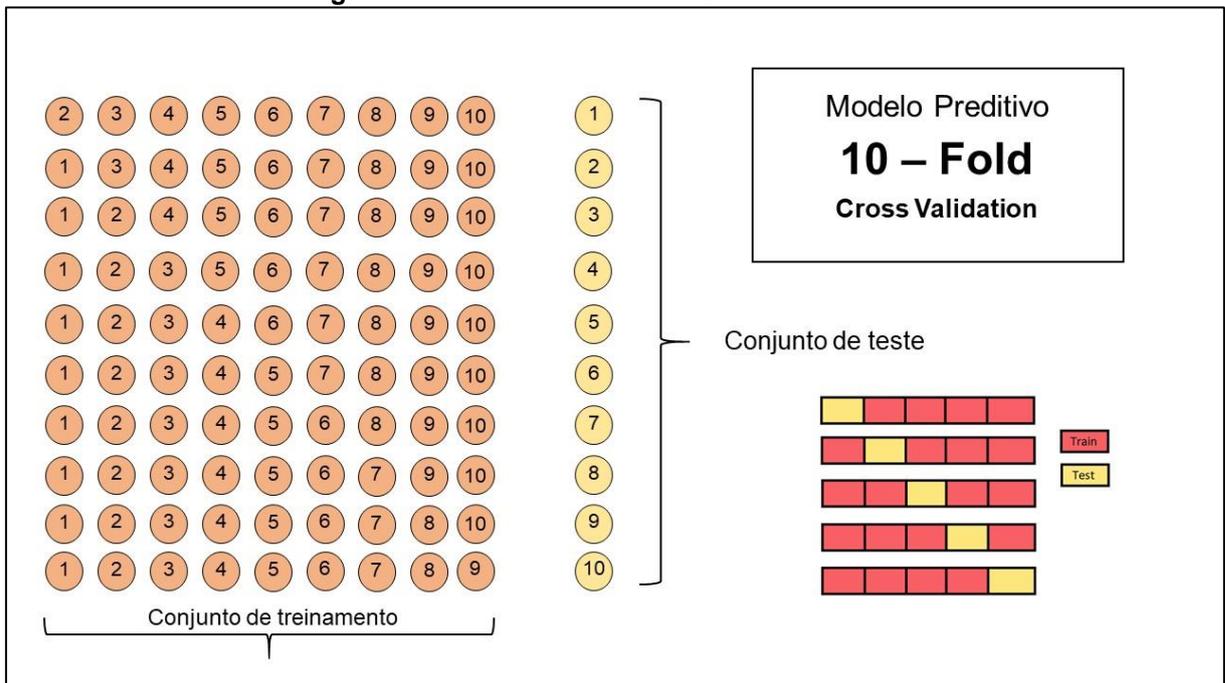
O modelo foi aferido através da técnica k-fold Cross Validation (CV) onde definiu-se $k = 10$ subconjuntos, dos quais nove foram destinados para treinar o modelo e um que foi utilizado para teste, repetindo-se o processo 10 vezes e testando-se sempre com um conjunto diferente. Uma vez que a técnica de K-fold CV é essencial para entender como a acurácia do modelo é aferida, seu funcionamento é melhor explicado na seção 4.1.

4.1 K – Fold Cross Validation e a Acurácia do Modelo

O procedimento k-fold CV é indicado para pequenos conjuntos de dados visando obter resultados precisos e acurados conforme as necessidades do estudo. Essa técnica é amplamente utilizada na área de aprendizado de máquina, e ao avaliar o modelo, a acurácia é calculada usando valores que representam o desempenho desse modelo ao ser aplicado a um determinado banco de dados.

Os autores Kodavi (1995), Borra e Ciaccio (2010) descrevem que utilizar o K-Fold Cross-Validation significa obter uma avaliação robusta da performance de um modelo e a sua habilidade relacionada a diferentes conjuntos de dados. A divisão desse modelo parte de um conjunto de dados dividido em K partes (denominadas “Folds”) em tamanhos aproximadamente iguais. Dessa forma, o treinamento e a avaliação do modelo acontecem repetidamente em K vezes, assim utilizando em cada K um “Fold” diferente gerado através de um conjunto teste. O restante é considerado conjunto de treinamento, e após o processamento o resultado de cada Fold é combinado obtendo uma média geral do desempenho do modelo. É possível visualizar essa estrutura tendo o $k = 10$ na Figura 4.

Figura 4: Estrutura do 10-Fold Cross Validation.



Fonte: Elaboração Própria.

Através da Figura 4 é possível visualizar o funcionamento do CV tendo $k = 10$. Dos 10 fold, 9 são considerados conjunto de treinamento enquanto o conjunto que resta é o conjunto de teste. A média geral do desempenho é expressa por uma matriz de confusão, mostrando nas linhas, os valores esperados e nas colunas os valores classificados, conforme a Figura 5. A Figura 5 apresenta um esquema do funcionamento de uma matriz de confusão.

Figura 5: Representação do funcionamento da matriz de confusão.

		Classe Esperada	
		ÁGUA	NÃO ÁGUA
Classe Prevista	ÁGUA	25 Verdadeiro Positivo	10 Falso Positivo
	NÃO ÁGUA	25 Falso Negativo	40 Verdadeiro Negativo

Fonte: Elaboração Própria.

Na Figura 5 é mostrada a matriz de confusão com a classe prevista e a esperada, as características de “água” e “não água” mostrando os resultados obtidos a partir da classificação de interesse. Para o valor de 25 no verdadeiro positivo, significa que o algoritmo classificou corretamente na imagem o que era “água”.

Já nos 10 em falso positivo o que era “água” ele previu como “não água”. Assim funciona para a classe de “não água”, onde os 25 falsos negativos são previstos incorretamente e os 40 verdadeiros negativos ele previu de forma correta, sendo classificado como “não água” o que na imagem “não água”.

A acurácia é uma métrica que aponta o desempenho do algoritmo de classificação. Para calcular a acurácia do RF, é necessário obter o número de predições corretas das árvores de decisões, conforme mostrado na matriz de confusão (Figura 5).

Os valores encontrados após o processamento dos bancos de dados foram os seguintes:

Para os valores Verdadeiro Positivo encontrou-se: 25

Para os valores Verdadeiro Negativo encontrou-se: 40

Para os valores Falso Positivo encontrou-se: 10

Para os valores Falso Negativo encontrou-se: 25

A partir dessa matriz pode-se calcular a acurácia através da equação 2.

$$acurácia = \frac{VP+VN}{VP+FP+VN+FN} = \frac{\text{predições corretas}}{\text{todas as predições}} \quad (\text{Eq. 2})$$

onde *VP* são os valores verdadeiros positivos, *VN* os valores verdadeiros negativos, *FP* os valores falso positivos e *FN* os valores falso negativos.

Através dos valores da Figura 5, é possível calcular a acurácia do modelo como na equação 2:

$$acurácia = \frac{25+40}{25+10+40+25} = \frac{65}{100} = 0,65$$

Esse resultado obtido no cálculo da acurácia justifica como foi o desempenho do modelo. No caso do exemplo da matriz de confusão, o desempenho do modelo previu corretamente 65% a classificação dos dados. No entanto, a acurácia fornece uma média geral das taxas de acertos do algoritmo.

5. RESULTADOS E DISCUSSÕES

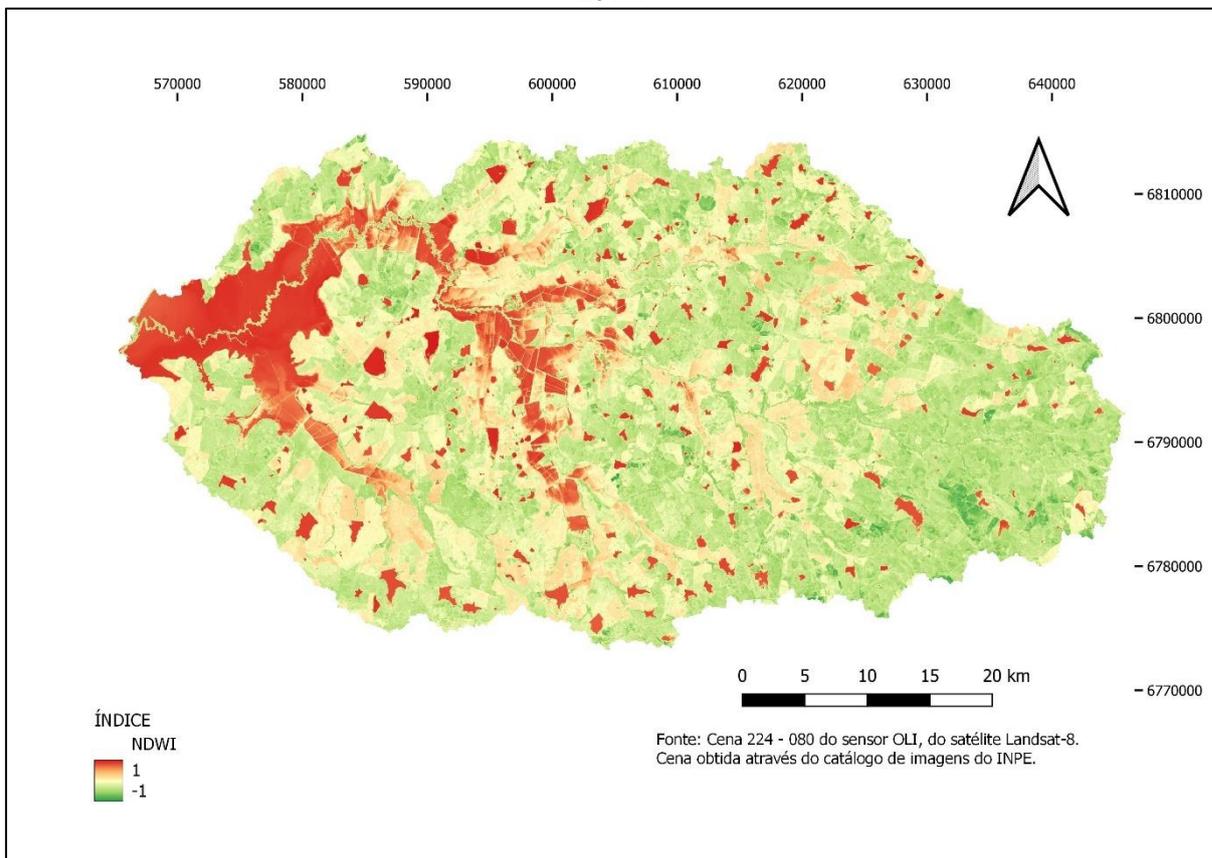
Por meio da segmentação da imagem digital, foi possível extrair um conjunto de dados para realizar o treinamento e a análise direcionada à obtenção dos resultados desejados. O conjunto de dados de treinamento foi classificado manualmente no *software* QGIS.

O procedimento envolveu etapas distintas, nas quais os dados foram treinados separadamente e modelados. Os resultados foram apresentados através de máscaras respectivamente de cada banda e inicialmente, foram utilizadas as imagens das bandas do visível e infravermelho próximo referentes ao espectro eletromagnético.

Após o treinamento dos dados, os resultados obtidos foram apresentados em figuras e quadros para análise e avaliação do desempenho do modelo. Além disso, foram gerados mapas com composições coloridas das bandas 6, 5 e 4, e 4, 3 e 2,

permitindo uma análise visual dos diferentes resultados. Na Figura 5 é possível ver a representação do índice NDWI para fins de comparação das demais imagens.

Figura 6: Índice NDWI gerado através de imagem do satélite Landsat-8, em 11 de junho de 2017.



Fonte: Adaptado pela autora.

Este índice varia de -1 até 1, propondo que valores iguais ou acima de zero determinam a existência de água na área, ao observar a Figura 6 os tons avermelhados identificam a presença de umidade no local. Já nos tons de verde e amarelo claro, como mostrado na legenda, possuem valores negativos remetendo que nesses locais a presença de água ou umidade é nula.

Após análise da Figura 6, o próximo passo foi treinar o modelo com o conjunto de treinamento referente às bandas 4, 3 e 2. Por meio do Quadro 2 apresenta-se a matriz de confusão do modelo de classificação Random Forest referente às bandas 4, 3 e 2.

Quadro 2: Matriz de Confusão com a aplicação das Bandas 4 3 2 através do modelo RF.

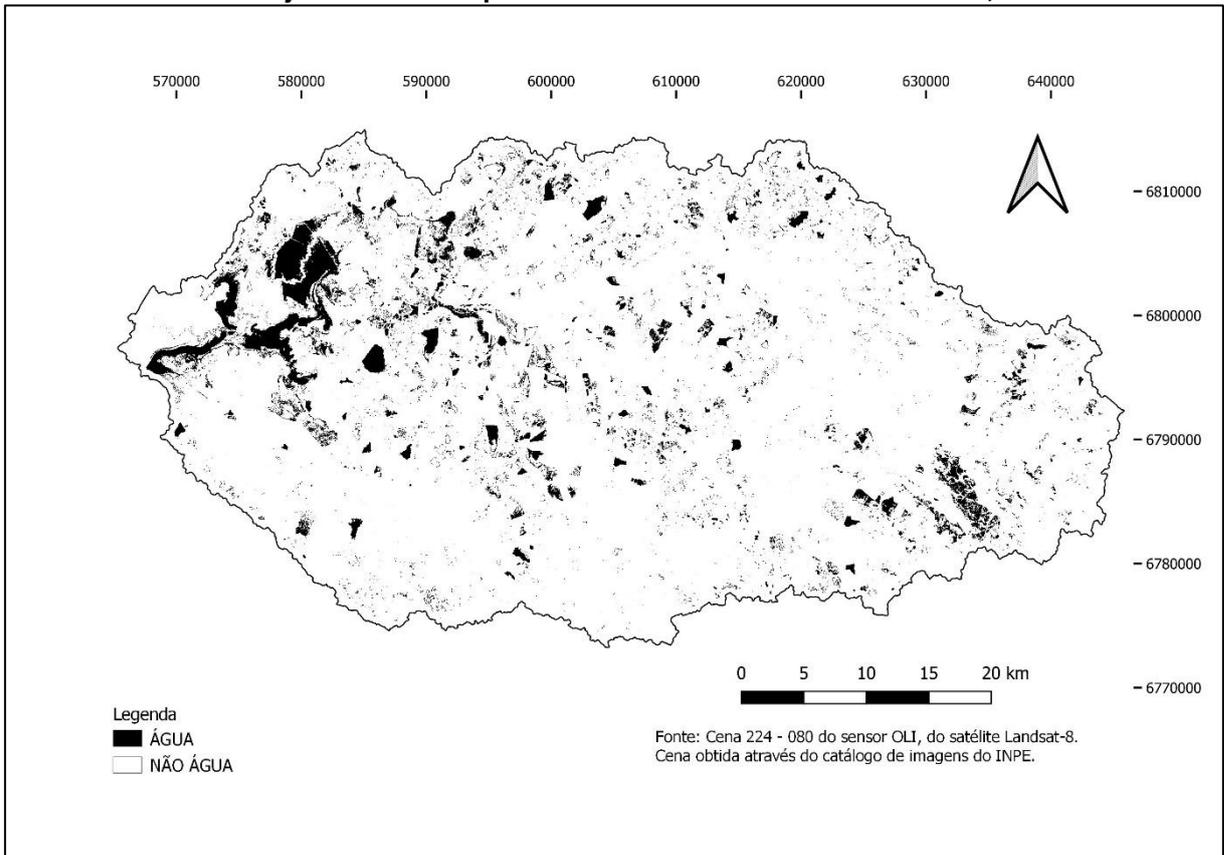
BANDAS 4 3 2		Classe Esperada	
		ÁGUA	NÃO ÁGUA
Classe Prevista	ÁGUA	53606	8832
	NÃO ÁGUA	1815	132076

Fonte: Elaboração Própria.

Através do Quadro 2 observa-se os resultados os quais foram determinados para treinamento do modelo utilizado na classificação. O modelo previu como “água” 53606 dos dados; e na classe “outros”, que nesse caso era irrelevante, o modelo classificou corretamente 132076.

A acurácia obtida foi de 94,5769%, mas a qualidade da máscara mostra um resultado não satisfatório pelo fato de que, algumas características de “água” e “não água” podem estar presentes em comportamentos espectrais semelhantes devido a sobreposição de cores. Através da Figura 7 mostra-se a máscara resultante da classificação das bandas 4, 3 e 2 gerada através do modelo Random Forest.

Figura 7: Máscara resultante do modelo gerado através da classificação do Random Forest com o conjunto de dados para treinamento referente às Bandas 4, 3 e 2.

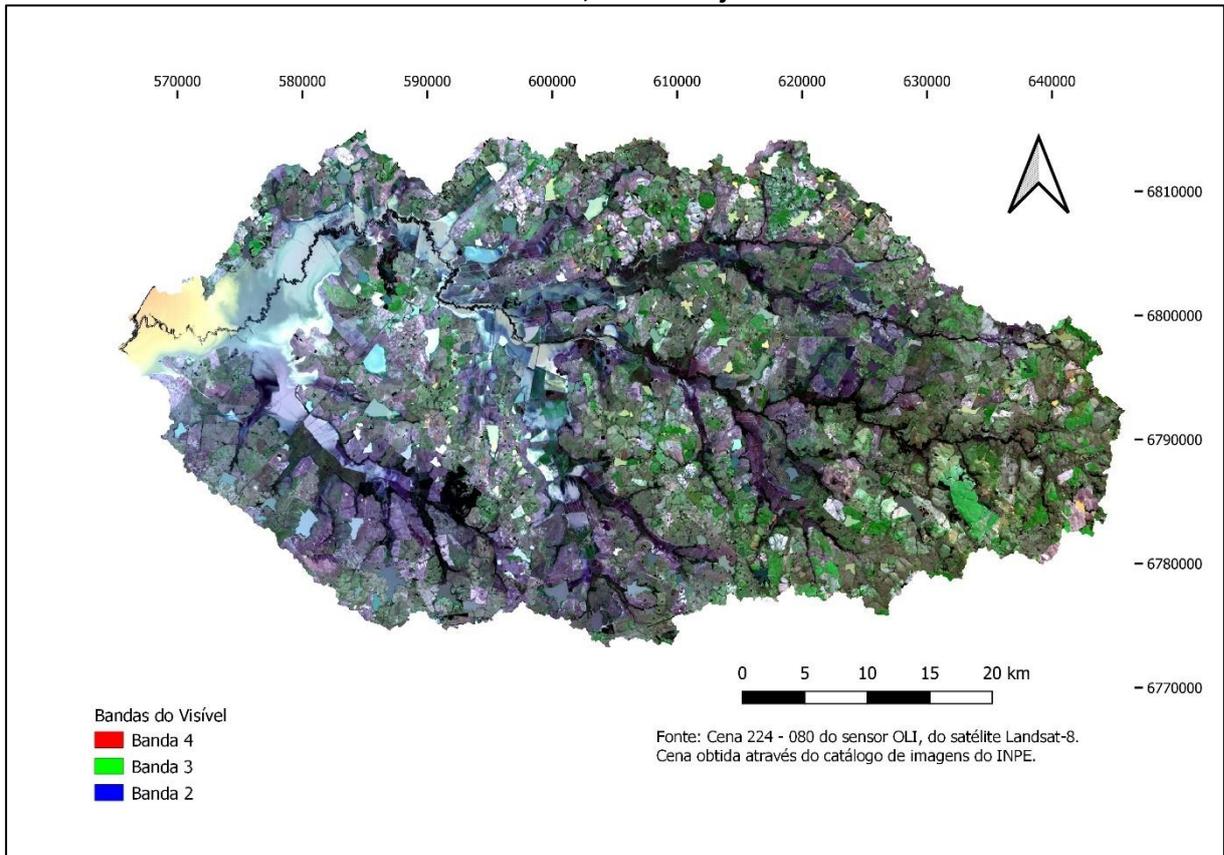


Fonte: Adaptado pela autora.

A máscara apresentada na Figura 7 resultou do treinamento do modelo através do conjunto de dados. O modelo generaliza de forma adequada os dados, porém a composição colorida de cores naturais (RGB – bandas 4, 3 e 2) consiste em informações complexas considerando a existência de sobreposição de cores.

A Figura 8 apresenta a área da Bacia Hidrográfica do Butuí na composição colorida das bandas 4, 3 e 2. A região da luz do visível, no espectro eletromagnético, remete a cores primárias (vermelho, verde e azul) que realizam a composição da imagem em cores naturais.

Figura 8: Composição colorida da Bacia Hidrográfica do Butuí nas Bandas 4, 3 e 2 de imagem do satélite Landsat-8, em 11 de junho de 2017.



Fonte: Adaptado pela autora.

A Figura 8 mostra as características da Bacia Hidrográfica do Butuí e sua região através da composição colorida nas bandas 4, 3 e 2, ou seja, cores naturais visíveis à olho nu. Apesar de boa qualidade, existem regiões na imagem que dificultam o reconhecimento dos elementos presentes na imagem.

Em relação a alta quantidade de erros tanto para a classe “água” como a “não água”, o modelo responsável por essa classificação encontrou dificuldades para identificação dos objetos. Isso ocorre por conta das informações contidas nas bandas 4, 3 e 2 como é possível analisar na Figura 8.

É desafiador identificar alguns elementos na imagem digital, portanto é possível que o modelo também identifique complexidade na análise dos dados. Assim como tem elementos que são facilmente identificáveis, seja pela cor, forma ou localização na imagem.

Seguindo o treinamento dos dados, um outro modelo foi treinado para análise das informações provindas das bandas 6, 5 e 4. Observando o Quadro 3, apresenta-se a matriz de confusão obtida através do modelo treinado com dados referente às bandas 6, 5 e 4.

Quadro 3: Matriz de Confusão com a aplicação das Bandas 6, 5 e 4 através do modelo RF.

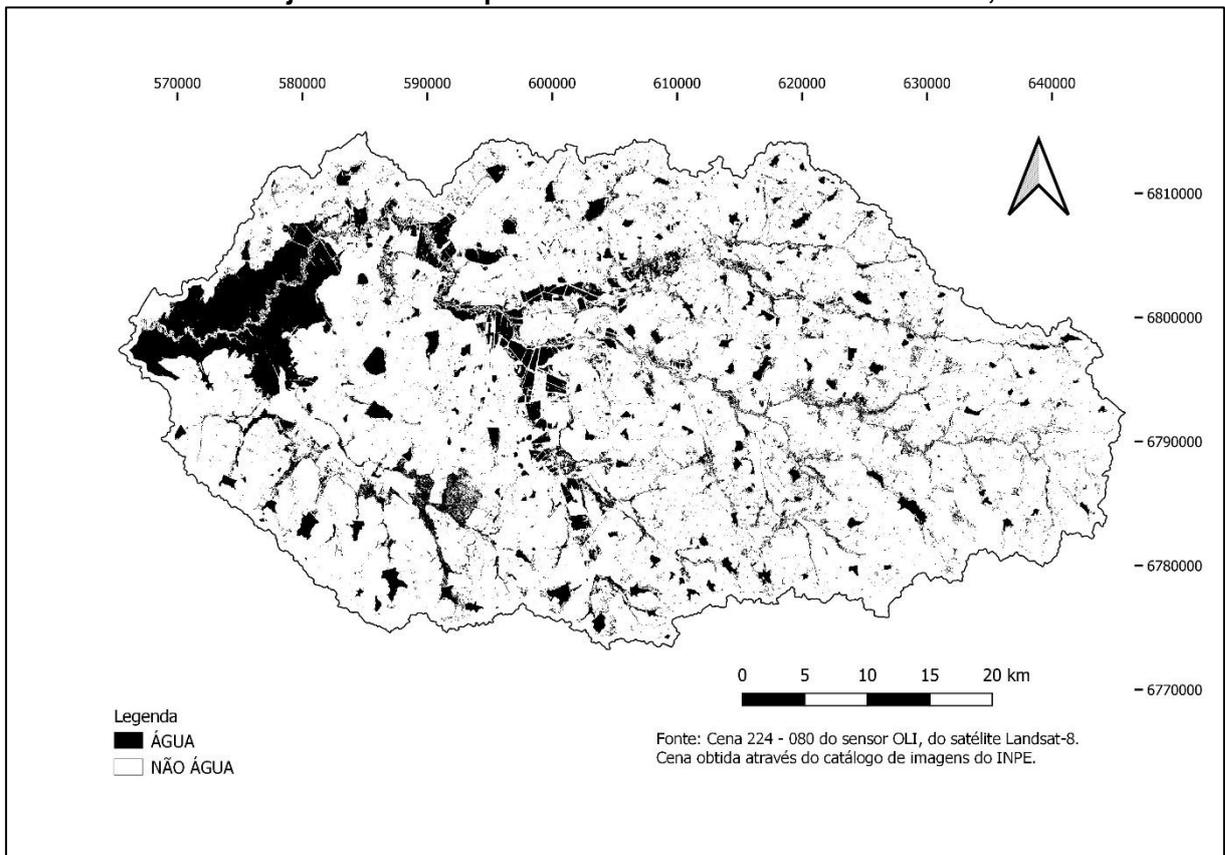
BANDAS 654		Classe Esperada	
		ÁGUA	NÃO ÁGUA
Classe Prevista	ÁGUA	62432	6
	NÃO ÁGUA	3	133888

Fonte: Elaboração Própria.

O modelo classificou 62432 valores corretos para a classe “água” e 133888 para classe “não água”. Os erros foram consideravelmente baixos, sendo 3 para classe que era “água” e o modelo classificou “não água”, e 6 erros para a classe que era “não água” e ele classificou como sendo “água”.

Nota-se que a acurácia ao ser calculada através dos resultados da matriz de confusão obteve o valor de 99,9954%. Esse resultado indica como foi o desempenho do modelo conforme os dados de treinamento e através da Figura 9 mostra-se a máscara resultante da classificação das bandas 6, 5 e 4 gerada através do modelo Random Forest.

Figura 9: Máscara resultante do modelo gerado através da classificação do Random Forest com o conjunto de dados para treinamento referente às Bandas 6, 5 e 4.



Fonte: Adaptado pela autora.

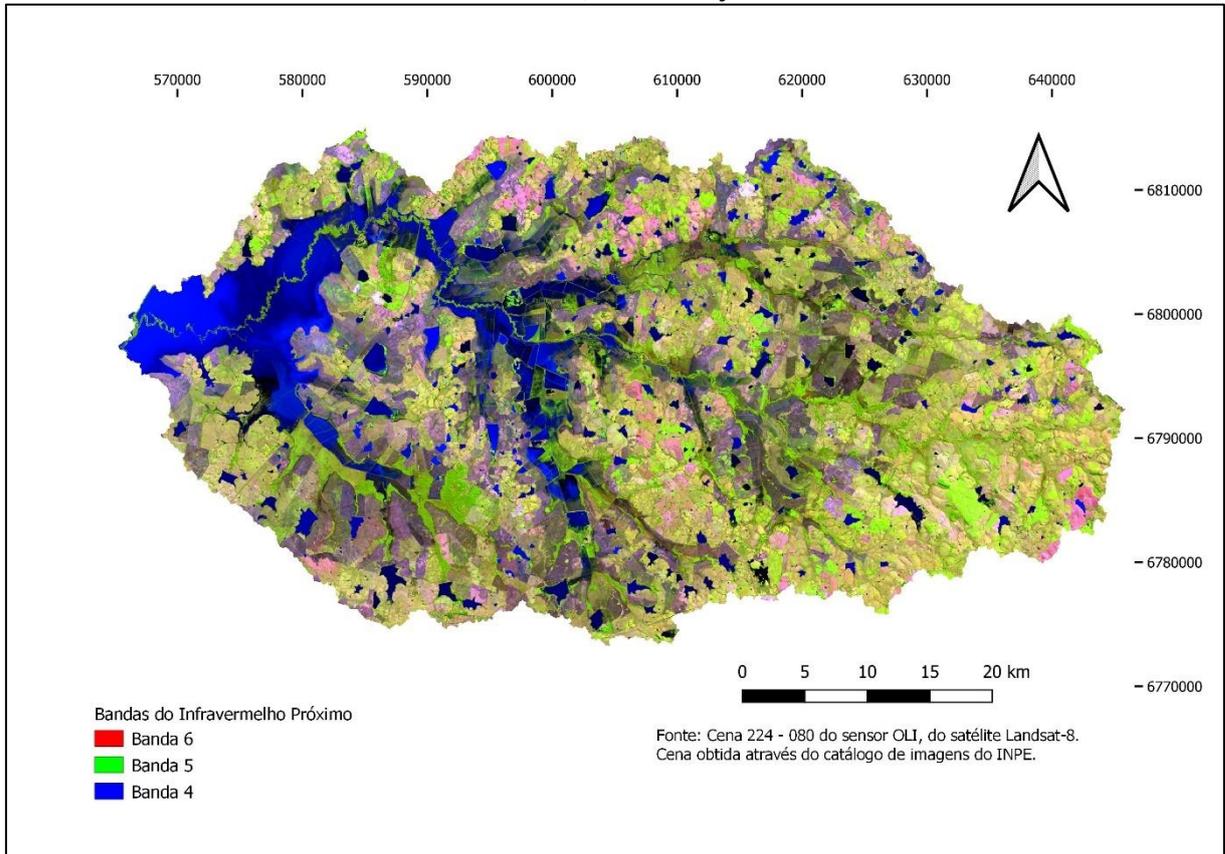
Na Figura 9 a máscara gerada através das bandas 6, 5 e 4 visando a boa qualidade, instiga-se que o conjunto de treinamento pode ser disjunto, ou seja, tem características bem específicas em relação as classes “água” e “não água”. Entretanto não possui pixels que afirmem incerteza, dificultando a generalização dos dados.

A predominância do preto existente na máscara mostra a presença de corpos hídricos ou até mesmo umidade, assim como a sinuosidade de elementos que provavelmente sejam cursos d’água, provindos de objetos semelhantes à afluentes da Bacia Hidrográfica do Butuí. Na imagem de satélite o que dá a entender é que sejam cursos d’água que realizam a drenagem da bacia, mantendo a mesma em seu nível.

As bandas 6, 5 e 4 ocupam parte da região do infravermelho no espectro eletromagnético. Isso implica na alta absorção de luz e baixa reflectância. Nesse caso, as classes “água” e “não água” são enfatizadas na obtenção das informações pelo

sensor e por meio da Figura 10 apresenta-se a área da Bacia Hidrográfica na composição colorida referente às bandas 6, 5 e 4.

Figura 10: Composição colorida da Bacia Hidrográfica do Butuí nas Bandas 6, 5 e 4 de imagem do satélite Landsat-8, em 11 de junho de 2017.



Fonte: Adaptado pela autora.

Na Figura 10 pode-se visualizar a composição colorida das bandas 6, 5 e 4 identificando os tons de azul como a existência de corpos d'água no local. As demais colorações indicam outras classes como por exemplo, vegetação, solo exposto, etc.

Há regiões que o modelo identifica legivelmente a presença de água através do tom de azul predominante a Oeste da imagem, o que na Figura 9 não apresenta a mesma quantidade de água. Alguns traços também são mostrados do que provavelmente sejam córregos vindos de áreas onde há armazenamento de água.

Assim, as características existentes na Figura 10 apresentam de forma clara algumas áreas com água ou umidade e as áreas que não eram água, possibilitando analisar facilmente a região da Bacia Hidrográfica do Butuí. O próximo passo foi

mostrar o modelo treinado com os dados de treinamento referente às 7 bandas através do Quadro 4 que apresenta os resultados obtidos por meio do Random Forest ao treinar os dados e gerar a matriz de confusão.

Quadro 4: Matriz de Confusão com a aplicação de Todas as Bandas através do modelo RF.

TODAS AS BANDAS		Classe Esperada	
		ÁGUA	NÃO ÁGUA
Classe Prevista	ÁGUA	62431	7
	NÃO ÁGUA	9	133883

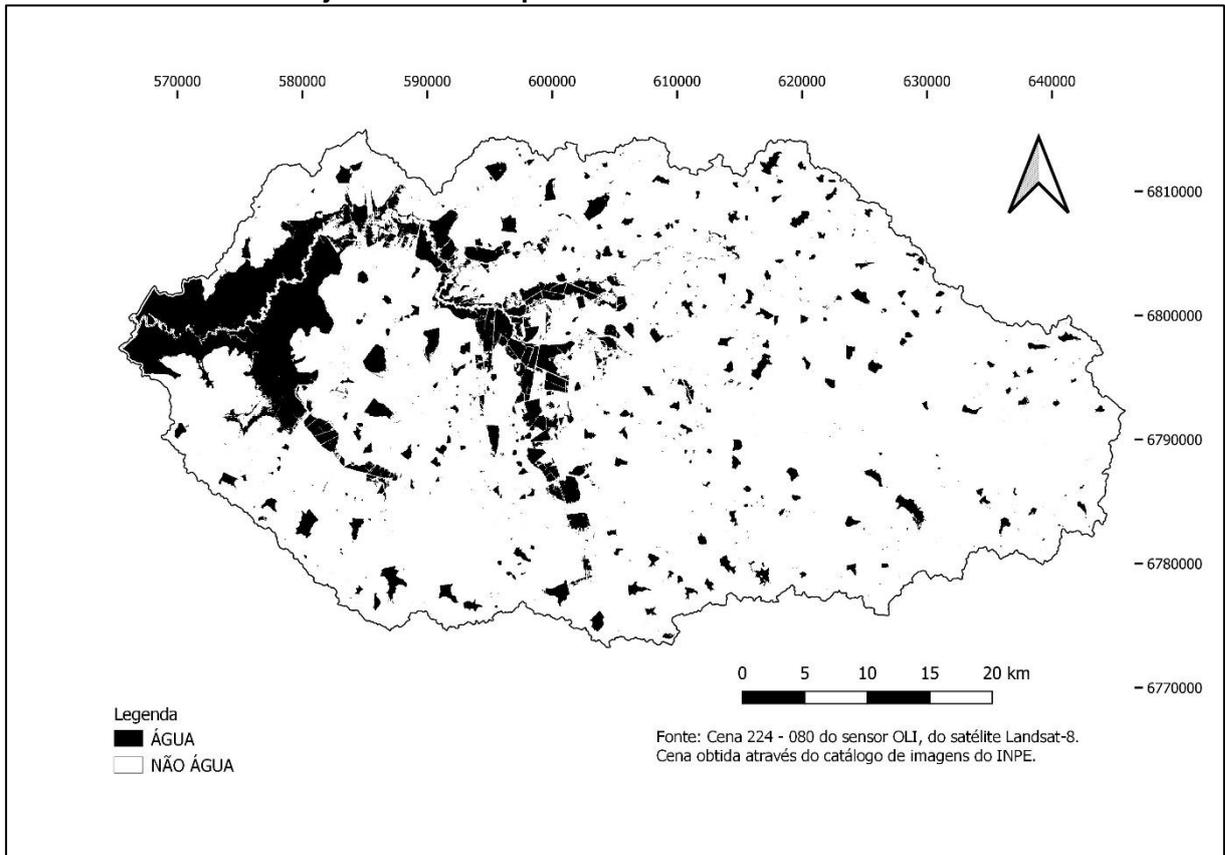
Fonte: Elaboração Própria.

O modelo classificou corretamente 62431 como “água” e 133883 como “não água”. Previu incorretamente 9 como “não água” sendo que eram “água” e 7 como “água” onde o esperado era “não água”.

Se comparar essa classificação com as demais, foi a que teve maior qualidade, pois de forma adequada generalizou o modelo através dos dados de treinamento e a máscara supriu as características existentes na imagem original. Ao calcular a acurácia da máscara contendo às 7 bandas, obteve-se 99,9918% de previsões corretas e do excelente desempenho do modelo para o treinamento de dados.

Recapitulando que o conjunto de treinamento pode apresentar um viés, conforme descrito nos resultados das bandas 6, 5 e 4. Para avaliação da qualidade do treinamento do modelo, na Figura 11 mostra-se a máscara resultante da classificação referente às 7 bandas através do modelo Random Forest.

Figura 11: Máscara resultante do modelo gerado através da classificação do Random Forest com o conjunto de dados para treinamento referente às 7 Bandas.



Fonte: Adaptado pela autora.

A máscara referente às 7 bandas após o treinamento de dados gerou resultados com boa precisão. Isso porque o conjunto de treinamento foi modelado incluindo todas as bandas, ou seja, bandas 1 a 7 referentes ao comportamento espectral do satélite Landsat8, ocasionando o bom desempenho do modelo.

Como é possível analisar a Figura 11 onde a água está determinada pela cor preta e a cor branca determina áreas da classe não água. Observando com maior precisão a máscara, existem áreas que o modelo não conseguiu detectar a umidade que poderia ser representada através da cor preta em traços menores.

Com relação a qualidade da máscara apresentada na Figura 11, ela evidenciou outras características diferentes das bandas 6, 5 e 4 e apresentou melhor qualidade que a máscara classificada com as bandas 4, 3 e 2. Os resultados das máscaras referentes às 7 bandas e bandas 6, 5 e 4 se mostraram satisfatórios.

A classificação apresentada na Figura 9 foi mais adequada do que é “água” e “não água”, porém possui diferenças na classificação de objetos que, supostamente sejam traços similares a afluentes, se comparada com a Figura 11. Em consideração a máscara resultante mostrada na Figura 11 generaliza melhor a água existente em outra área (mais a Oeste) da imagem.

Posteriormente ao analisar os resultados obtidos após o processamento dos dados, é possível observar que o comportamento da água varia nas diferentes faixas do espectro eletromagnético. Isso explica os resultados encontrados nas bandas RGB (4, 3, 2) e em todas as bandas.

Como mencionado anteriormente, as bandas usadas para calcular o índice NDWI e as bandas do infravermelho próximo desempenham um papel significativo. Pois auxiliam em enfatizar a presença da água na imagem e facilitam o reconhecimento pelo modelo.

Assim, a acurácia de 94,5769%, calculada com base na matriz de confusão gerada pelo modelo usando as bandas 4, 3, 2, representa como foi o desempenho do modelo, ou seja, satisfatório. Conforme evidenciado nos Quadros 2 e 4, a composição com todas as bandas apresentou um percentual maior (99,9918%), enquanto o RGB teve um percentual menor (94,5769%), devido à classificação incorreta de dados ser mais prevalente nesse caso.

6. CONCLUSÃO

Conclui-se que o método Random Forest demonstrou uma notável capacidade de identificar e classificar corpos d'água em imagens de satélite, mesmo quando as máscaras geradas apresentavam qualidade inferior. Os resultados obtidos sugerem que o Random Forest é um modelo robusto e adaptável, capaz de superar as limitações da qualidade das máscaras e obter resultados precisos na identificação e classificação de corpos d'água.

Foi mostrada uma imagem do NDWI gerada no software QGIS para servir de linha base de comparação. Foram gerados 3 modelos de classificação (bandas 4, 3 e 2, bandas 6, 5 e 4 e as 7 bandas).

O modelo referente às bandas 4, 3 e 2 obteve a acurácia de 94,5769%, o 6, 5 e 4 resultou em uma acurácia de 99,9954% e o modelo com as 7 bandas obteve acurácia de 99,9918%. O modelo das 7 bandas alcançou resultados aproximados do modelo das bandas 6, 5 e 4, levando em consideração a identificação de pixels que o modelo 6, 5 e 4 foi incapaz.

O modelo referente às bandas 4, 3 e 2 em relação aos demais modelo gerou uma máscara de qualidade satisfatória, pois o modelo generalizou adequadamente os dados. Por conta da sobreposição de cores, teve acurácia baixa conforme as dificuldades em caracterizar os pixels.

O modelo referente às bandas 6, 5 e 4 gerou uma máscara satisfatória e a acurácia maior que os outros modelos. Essa máscara apresentou maiores informações comparada com a máscara do modelo 4, 3 e 2 e vale salientar que o conjunto de treinamento provavelmente é disjunto o que justifica a excelente acurácia, mas a precisão da máscara inferior.

O modelo referente às 7 bandas resultou em uma máscara com características distintas dos demais modelos e a acurácia muito próxima do modelo 6, 5 e 4. As divergências ocorrem na indicação de presença de água ou não, comparando com a máscara resultante das bandas 6, 5 e 4.

Há regiões na imagem onde existem incertezas, onde o modelo não generaliza bem os dados. Aparentemente tendo mais dados (7 bandas) o modelo parece ter generalizado melhor os dados da imagem original e apresenta uma máscara de classificação similar ou mais similar ao NDWI em comparação com as demais.

Deve-se enfatizar que o modelo com a aplicação nas bandas 432, ou seja, composição natural da imagem da Bacia Hidrográfica do Butuí, foi o modelo menos adequado tendo como base o valor da sua acurácia. A vasta complexidade das

informações contidas na composição colorida implica na generalização do modelo, onde a coloração igual ou similar pode ser classificada tanto “água” como “não água”.

Por fim, como trabalhos futuros, sugere-se a análise das árvores de decisão geradas em cada modelo. Na intenção de aferir, quais características o algoritmo considera mais relevante na classificação dos dados.

7. REFERÊNCIAS

Borra e Ciaccio (2010) Simone Borra e Agostino Di Ciaccio. Measuring the prediction error. a comparison of cross-validation, bootstrap and covariance penalty methods. *Computational Statistics and Data Analysis*, 54:2976–2989. Citado na pág. 1, 2, 5, 6, 10, 11, 19

CHO, David Fernando; SCHWAIDA, Samuel Fernando; ENNES CICERELLI, Rejane; ALMEIDA, Tati; MARQUES RAMOS, Ana Paula; EYJI SANO, Edson. Desempenho do Algoritmo de Classificação de Imagens Random Forest para Mapeamento do Uso e Cobertura do Solo no Cerrado Brasileiro. **Anuário do Instituto de Geociências**, Rio de Janeiro, v. 44, p. 1-11, 2021. DOI https://doi.org/10.11137/1982-3908_2021_44_37979. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/227879/1/Sano-Desempenho-de-algoritmo-de-classificacao.pdf>. Acesso em: 5 jun. 2023.

EngeSat. IMAGEM de Satélite: LANDSAT 8. Paraná, 2013. Disponível em: <LANDSAT 8 - EngeSat - Imagens de satélite e geoprocessamento>. Acesso em: 02 ago. 2022.

HOLLER, Wilson. Embrapa Monitoramento por Satélite: Gestão Territorial Estratégica. *In*: VOIGT, Maicon. **Sensoriamento Remoto**. [S. l.], 2011.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, IBGE. Atlas Geográfico Escolar na Internet. *In: O que é cartografia?* sensoriamento remoto. [S. l.], 2022.

Kohavi (1995) Ron Kohavi. Measuring the prediction error. a comparison of c. *Computational Statistics and Data Analysis*, 1:2976–2989. Citado na pág. 9, 10

MARTINS, Cinthia Mara; DA SILVA, Benedito Cláudio; PONS, Nívea Adriana. Estimativa de cheias em bacias hidrográficas com base em previsões de precipitação por conjunto. **Revista Brasileira de Geografia Física**, [s. l.], v. 12, n. 5, p. 1713-1729, 2019.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Indução de Regras e Árvores de Decisão. *In: SISTEMAS Inteligentes para Engenharia*. [S. l.: s. n.], 2003. cap. 5, p. 57-74.

NDWI: Normalized Difference Water Index. *In: PRODUCT FACT SHEET: NDWI – EUROPE*. 1. 1. ed. Europe, 2011. Disponível em: https://edo.jrc.ec.europa.eu/documents/factsheets/factsheet_ndwi.pdf. Acesso em: 14 jun. 2023.

PAIVA, Magaywer; GALVANIN, Edinéia Aparecida; SILVA, Gustavo; COSTA, Angélica Aparecida; NEVES, Sandra Mara. Índice de água por diferença normalizada com a linguagem de programação Python aplicado ao Pantanal de Cáceres – Mato Grosso. **Anais 7º Simpósio de Geotecnologias no Pantanal**, Mato Grosso, ano 2018, p. 799-807, 24 out. 2018.

PEREIRA, Luciana; AMORIM, Gustavo; GRIGIO, Alfredo Marcelo; PARANHOS, Antônio. Análise Comparativa entre Métodos de Índice de Água por Diferença Normalizada (NDWI) em Área Úmida Continental. **Anuário do Instituto de Geociências - UFRJ**, Mato Grosso

do Sul, ano 2018, p. 654-662, 17 maio de 2018. DOI https://doi.org/10.11137/2018_2_654_662.

QGIS Development Team, <YEAR>. QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>

RODRIGUES, Wendel; MORAIS, Fernando; PASCHOAL, Leticia. ÍNDICE DE DIFERENÇA NORMALIZADA DA ÁGUA (NDWI) CALCULADO PARA ESTAÇÕES CHUVOSAS E SECAS NA BACIA DO CÓRREGO BARREIRO, LAGOA DA CONFUSÃO – TO. **XVII Simpósio Brasileiro de Geografia Física Aplicada**, Campinas - SP, ano 2017, p. 254-263, 28 jun. 2017. DOI 10.20396/sbgfa.v1i2017.1794.

SHIBA, Marcelo; SANTOS, Rosangela; QUINTANILHA, José Alberto; KIM, Hae. Classificação de imagens de sensoriamento remoto pela aprendizagem por árvore de decisão: uma avaliação de desempenho. **Anais XII Simpósio Brasileiro de Sensoriamento Remoto**, [S. l.], p. 4319-4326, 16 abr. 2005.

ZUCUNI PRINA, Bruno; TRENTIN, Romario. GMC: Geração de Matriz de Confusão a partir de uma classificação digital de imagem do ArcGIS®. **Anais XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, [s. l.], p. 1-9, 2015. Disponível em: <http://www.dsr.inpe.br/sbsr2015/files/p0031.pdf>. Acesso em: 22 maio 2023.

