

UNIVERSIDADE CANDIDO MENDES – UCAM
PROGRAMA DE PÓS-GRADUAÇÃO EM PESQUISA OPERACIONAL E
INTELIGÊNCIA COMPUTACIONAL
CURSO DE MESTRADO EM PESQUISA OPERACIONAL E
INTELIGÊNCIA COMPUTACIONAL

VITOR DA SILVA GONÇALVES

**ESTUDO DE METODOLOGIAS PARA CLASSIFICAÇÃO DE ÁREA DE USO E
COBERTURA DO SOLO NA REGIÃO NORTE FLUMINENSE**

CAMPOS DOS GOYTACAZES, RJ

2023

UNIVERSIDADE CANDIDO MENDES – UCAM
PROGRAMA DE PÓS-GRADUAÇÃO EM PESQUISA OPERACIONAL E
INTELIGÊNCIA COMPUTACIONAL
CURSO DE MESTRADO EM PESQUISA OPERACIONAL E INTELIGÊNCIA
COMPUTACIONAL

Vitor da Silva Gonçalves

**ESTUDO DE METODOLOGIAS PARA CLASSIFICAÇÃO DE ÁREA DE USO E
COBERTURA DO SOLO NA REGIÃO NORTE FLUMINENSE**

Dissertação apresentada ao Programa de Mestrado em Pesquisa Operacional e Inteligência Computacional da Universidade Candido Mendes – Campos/RJ para obtenção do grau de mestre em Pesquisa Operacional e Inteligência Computacional.

Orientador: Prof. Ítalo de Oliveira Matias, D.Sc.

CAMPOS DOS GOYTACAZES, RJ

2023

Catálogo na fonte

Preparada pela Biblioteca da **UCAM – CAMPOS** 019/2023

Gonçalves, Vitor da Silva.

Estudo de metodologias para classificação de área de uso e cobertura do solo na região Norte Fluminense. / Vitor da Silva Gonçalves. – 2023.
94 f.

Orientador(a): Ítalo de Oliveira Matias.

Dissertação de Mestrado em Pesquisa Operacional e Inteligência Computacional – Universidade Candido Mendes – Campos. Campos dos Goytacazes, RJ, 2023.

Referências: f. 71-76.

1. Classificação do solo. 2. Norte Fluminense. I. Matias, Ítalo de Oliveira, orient. II. Universidade Candido Mendes – Campos. III. Título.

CDU – 631.44(815.3)

Bibliotecária Responsável: Flávia Mastrogirolamo CRB 7ª-6723

Vitor da Silva Gonçalves

**ESTUDO DE METODOLOGIAS PARA CLASSIFICAÇÃO DE ÁREA DE USO E
COBERTURA DO SOLO NA REGIÃO NORTE FLUMINENSE**

Dissertação apresentada ao Programa de Mestrado em Pesquisa Operacional e Inteligência Computacional da Universidade Candido Mendes – Campos/RJ para obtenção do grau de mestre em Pesquisa Operacional e Inteligência Computacional.

Aprovada em: 02 de junho de 2023.

BANCA EXAMINADORA

Prof. Ítalo de Oliveira Matias, D.Sc.
Universidade Candido Mendes

Prof. Aldo Shimoya, D.Sc.
Universidade Candido Mendes

Prof. Milton Erthal Junior, D.Sc.
Instituto Federal Fluminense

CAMPOS DOS GOYTACAZES, RJ

2023

RESUMO

Introdução: Classificações de uso e cobertura do solo auxiliam na compreensão dos fenômenos da superfície terrestre e têm se tornado uma ferramenta importante para diferentes tipos de estudo. Com sua popularização, diversos métodos vêm sendo desenvolvidos para a classificação dos elementos presentes em imagens de sensoriamento remoto. Contudo, observa-se uma carência de informações sobre características dos diferentes tipos de classificadores, que poderiam criar maior embasamento para a discussão e escolha dos mesmos em novos estudos.

Objetivo: identificar e quantificar as principais metodologias aplicadas nos estudos de classificação de uso e cobertura do solo e suas características, apontando suas vantagens, desvantagens e aplicações mais adequadas. **Metodologia:** foi realizada uma pesquisa bibliométrica na base de artigos científicos Scopus. Os dados obtidos foram tratados por meio de banco de dados SQL para a quantificação da utilização dos métodos empregados nos estudos de classificação de uso e cobertura do solo no período de 2012 a 2022. Em seguida, foram selecionados quatro dos principais classificadores utilizados: máxima verossimilhança (MLE), máquina de vetores de suporte (SVM), Random Forest e redes neurais artificiais (RNA). Foi realizado um comparativo entre os métodos, tomando como área de estudo a região Norte Fluminense e, como dados de referência, o estudo de uso e cobertura do solo do projeto MapBiomias. **Resultados:** Foi identificada uma tendência geral de crescimento nas publicações do tema, tendência de aumento da utilização de métodos baseados em aprendizado de máquina, e estabilização do uso das demais metodologias, sendo os métodos estatísticos também bastante empregados, e os demais, com uso pouco expressivo. Já o comparativo de métodos, este permitiu identificar que a eficiência dos classificadores depende das características espectrais das regiões a serem identificadas, mas com particularidades entre cada método. Em geral, as classes com maiores distinções espectrais foram classificadas com maior acurácia, e as com menores distinções geraram maiores erros. O método MLE apresentou os melhores resultados para classes bastante distintas espectralmente, enquanto o Random Forest demonstrou alguma eficácia ao lidar baixa diferenciação espectral, mas dificuldade para classificação de cobertura não vegetal, que apresentou alta dispersão em relação à média de valores dos seus *pixels*. Os métodos SVM e RNA, não apresentaram eficácia para classes espectralmente similares, mas o SVM apresentou a maior acurácia na classificação de cobertura não vegetal. **Conclusão:** Os resultados obtidos podem permitir a facilitação da escolha de metodologias a serem empregadas em estudos futuros na área de uso e cobertura do solo, principalmente na região Norte Fluminense.

Palavras-chave: uso e cobertura do solo; classificação; SVM; Random Forest; bibliometria.

ABSTRACT

Introduction: Land use and land cover classifications contribute to understanding the earth's surface phenomena, becoming an important tool for different studies. As it gained popularity, various methods have been developed to classify the elements present in remote sensing images. However, there is a lack of information on the use of different methods, as well as a lack of information on the characteristics of each one, which could improve the basis for discussing and choosing them in new studies. **Objective:** identify and quantify the main methodologies applied in land use and land cover classification studies and their characteristics, pointing out their advantages, disadvantages, and most appropriate applications. **Methodology:** a bibliometric research was performed in the Scopus scientific database. The data obtained was processed using an SQL database to quantify the methods employed in land use and land cover classification studies in the period between 2012 and 2022. Next, four of the main methods used in the literature were selected: maximum likelihood (MLE), support vector machine (SVM), random forest, and artificial neural networks (ANN). A comparison was performed between them, using the North region of Rio de Janeiro State as the studied area and, as reference data, the land use and land cover data from the MapBiomias project. **Results:** A general growth trend was identified in publications on the subject, with an increase in the use of machine learning methods and stabilization in the utilization of other methodologies, with statistical methods also widely used and others with little expressive use. The method comparison allowed identifying that the efficiency of classifiers depends on the spectral characteristics of the classes to be determined, but with particularities between each method. In general, more spectrally distinct classes were classified with greater accuracy, while the less distinct generated greater errors. The MLE method presented the best results for classes with high spectral differences, while Random Forest showed some efficacy in a scenario with low spectral similarity, but difficulty in classifying non-vegetative cover, whose class showed high dispersion from the mean of its pixel values. The SVM and RNA methods did not present efficacy in distinguishing spectrally similar classes, but the SVM showed the highest accuracy in non-vegetative cover classification. **Conclusion:** The results obtained can help in choosing a classification methodology for future studies of land use and land cover classification, especially in the North region of Rio de Janeiro state.

Keywords: land cover; land use; classification; SVM; random forest; bibliometry.

LISTA DE FIGURAS

Figura 1 - Número de artigos publicados no período de 2012 a 2022 sobre o tema “classificação de uso e cobertura do solo” na base Scopus.....	20
Figura 2 - Quantidade de artigos publicados para o tema para cada 10.000 artigos publicados para todos os temas.....	21
Figura 3 - Percentual de utilização das diferentes técnicas no período de 2012 a 2022	22
Figura 4 - Percentual de utilização de diferentes técnicas de aprendizado de máquina	25
Figura 5 - Percentual de artigos que aplicaram tipos de técnicas não categorizadas	26
Figura 6 - Imagem da região de estudo obtida a partir do Sentinel-2 (exibindo bandas vermelho, verde e azul)	43
Figura 7 - Imagem da região de estudo obtida a partir do projeto Mapbiomas e utilizada como referência	44
Figura 8 - Plotagem espectral para os dados de treinamento	45
Figura 9 - Imagem classificada pelo método da máxima verossimilhança (MLE)	47
Figura 10 - Pixels classificados erroneamente pelo método da máxima verossimilhança (MLE).....	48
Figura 11 - Imagem classificada pelo método máquina de vetores de suporte (SVM)	51
Figura 12 - Pixels classificados erroneamente pelo método máquina de vetores de suporte (SVM)	52
Figura 13 - Imagem classificada pelo método Random Forest	55
Figura 14 - Pixels classificados erroneamente pelo método Random Forest.....	56
Figura 15 - Imagem classificada pelo método redes neurais artificiais (RNA)	59
Figura 16 - Pixels classificados erroneamente pelo método redes neurais artificiais (RNA)	60

LISTA DE TABELAS

Tabela 1 - Relação de palavras-chave pesquisadas e seus métodos associados....	18
Tabela 2 - Resultados das consultas por palavra-chave.....	24
Tabela 3 - Coeficiente de similaridade Bray-Curtis para cada par de classe	46
Tabela 4 - Matriz de confusão para o método da máxima verossimilhança (MLE) ...	49
Tabela 5 - Coeficiente <i>kappa</i> de cada classe para o método da máxima verossimilhança (MLE).....	50
Tabela 6 - Matriz de confusão para o método máquina de vetores de suporte (SVM)	53
Tabela 7 - Coeficiente <i>kappa</i> de cada classe para o método da máxima verossimilhança (SVM)	54
Tabela 8 - Matriz de confusão para o método Random Forest	57
Tabela 9 - Coeficiente <i>kappa</i> de cada classe para o método Random Forest	58
Tabela 10 - Matriz de confusão para o método redes neurais artificiais (RNA).....	61
Tabela 11 - Coeficiente <i>kappa</i> de cada classe para o método da máxima verossimilhança (RNA).....	62
Tabela 12 - Coeficiente de concordância kappa e acurácias do usuário e produtor para os quatro métodos	62

LISTA DE ABREVIATURAS E SIGLAS

CE	Commission Error (erro de comissão)
CSV	Comma-Separated Values (valores separados por vírgula)
GIS	Geographic Information System
IBGE	Instituto Brasileiro de Geografia e Estatística
INPE	Instituto Nacional de Pesquisas Espaciais
MLE	Maximum Likelihood Estimation (estimação de máxima verossimilhança)
NIR	Near-Infrared (infravermelho próximo)
OE	Omission Error (erro de omissão)
PA	Producer's Accuracy (acurácia do usuário)
QGIS	Quantum GIS
RNA	Redes Neurais Artificiais
SIG	Sistemas de Informação Geográfica
SVM	Support Vector Machine (máquina de vetores de suporte)
SWIR	Short-wave infrared (infravermelho de onda curta)
TWDTW	Time-Weighted Dynamic Time Warping
UA	User's Accuracy (acurácia do usuário)

SUMÁRIO

1	INTRODUÇÃO GERAL	9
1.1	OBJETIVOS	11
1.1.1	Objetivos específicos	11
1.2	JUSTIFICATIVA	11
1.3	DELIMITAÇÃO DO PROBLEMA	12
1.4	ESTRUTURA DA PESQUISA	12
2	ARTIGO A – O ESTADO DA ARTE DOS CLASSIFICADORES DE USO E COBERTURA DO SOLO	13
2.1	INTRODUÇÃO	13
2.2	REVISÃO DA LITERATURA	14
2.3	METODOLOGIA	17
2.4	RESULTADOS E DISCUSSÃO	19
2.5	CONSIDERAÇÕES FINAIS	28
	REFERÊNCIAS	29
3	ARTIGO B – ESTUDO COMPARATIVO DE CLASSIFICADORES DE USO E COBERTURA DO SOLO	32
3.1	INTRODUÇÃO	32
3.2	REVISÃO DA LITERATURA	34
3.2.1	Classificação de uso e cobertura do solo	34
3.2.2	Métodos de classificação supervisionada para uso e cobertura do solo	38
3.2.3	Ferramentas populares para classificação de uso e cobertura do solo ..	39
3.3	METODOLOGIA	41
3.4	RESULTADOS E DISCUSSÃO	43
3.5	CONSIDERAÇÕES FINAIS	63
	REFERÊNCIAS	65
4	CONSIDERAÇÕES FINAIS GERAIS	69
	REFERÊNCIAS	71
	APÊNDICE A	77

1 INTRODUÇÃO GERAL

O monitoramento de uso e cobertura do solo se tornou uma importante ferramenta para estudos sobre regiões geográficas, com potencial para obter informações sobre variações nas áreas naturais, urbanas, agrárias, entre outras. Esta técnica permite a quantificação e análise temporal de fenômenos complexos e, em conjunto com outros tipos de estudos, contribui para uma compreensão profunda de tais fenômenos. Este tipo de monitoramento é considerado essencial para a preservação do meio ambiente e para a sustentabilidade (PULIGHE, 2022).

As pesquisas na área avançaram principalmente a partir da década de 1970, fortemente influenciadas pelo programa Landsat, que disponibiliza publicamente imagens digitais da superfície terrestre em um programa contínuo, desde seu primeiro lançamento, em 1972, mas também influenciadas pelos avanços da computação, que permitiu a aplicação de métodos mais eficientes (PHIRI; MORGENROTH, 2017).

As primeiras metodologias desenvolvidas se baseavam em métodos manuais ou numéricos, passíveis de cálculo manual (PHIRI; MORGENROTH, 2017). Foi com os adventos da computação e dos sistemas de informação geográfica, que novas metodologias puderam ser aplicadas, e esse tipo de pesquisa se tornou mais acessível, popularizando-se entre pesquisadores de diferentes áreas de estudo. Na atualidade, métodos baseados em aprendizado de máquina têm ganhado considerável atenção nos estudos da área, com a utilização de métodos gradativamente mais avançados (TALUKDAR *et al.*, 2020), o que é uma tendência que se observa na computação em sua totalidade (PUGLIESE; REGONDI; MARINI, 2021). Gupta, Sethi e Bathija (2021) apontaram que a convergência do *deep learning*, agentes inteligentes e interfaces homem-máquina convergem de tal forma que podem construir o futuro das metodologias para uso e cobertura do solo. Contudo, é importante a realização de pesquisas sistematizadas para a identificação de tal fenômeno, principalmente em áreas aplicadas.

Uma das ferramentas relevantes para a popularização dos estudos de uso e cobertura do solo é o QGIS¹, uma suíte de criação, visualização, análise e

¹O QGIS é um software SIG de código aberto e multiplataforma que permite a criação, visualização, análise e publicação de dados e mapas geoespaciais. Pode ser obtido a partir do endereço eletrônico: <https://qgis.org>.

publicação de dados geoespaciais licenciada na forma de *software* livre, sob licença GPL (QGIS DEVELOPMENT TEAM, 2023). Sua distribuição sob licenciamento aberto e sem custos permite que qualquer indivíduo possa adquirir uma cópia do programa e desenvolver projetos baseados em dados geoespaciais. Este tipo de licenciamento permite ainda que os usuários participem do desenvolvimento do projeto, seja diretamente no repositório oficial, seja na forma de *plugins* para estender as funcionalidades do *software*. Por meio dos *plugins* desenvolvidos pela comunidade, o QGIS permite a aplicação de diferentes metodologias de classificação de uso e cobertura do solo.

Outra ferramenta de relevância na atualidade, em relação à popularização destes tipos de estudo, é a Earth Engine², uma plataforma GIS desenvolvida pela Google. Essa ferramenta, apesar de não ser distribuída sob licenciamento aberto, é oferecida gratuitamente a pesquisadores por meio de cadastro na plataforma. Seu principal diferencial em relação às demais é utilizar de computação em nuvem para a realização do processamento dos dados (GORELICK *et al.*, 2017), que tendem a ser intensivos para computadores convencionais. Desta forma, é possível que pesquisadores sem acesso a recursos avançados de *hardware* possam realizar estudos em áreas de grandes dimensões, e com imagens de alta resolução.

Com a popularização dos sistemas de informação geográfica e dos estudos de uso e cobertura do solo, aliados do progresso da computação, obtêm-se uma vasta gama de aplicações para essas técnicas. Tal expansão se deu por meio de novas metodologias que vêm sendo aplicadas nestes estudos, permitindo menores taxas de erro, distinção de elementos com pouca variação espectral, entre outras vantagens, contribuindo para a superação dos principais desafios enfrentados pelos estudos da área (CHAVES; PICOLI; SANCHES, 2020).

Nessa perspectiva, o presente trabalho visa realizar um levantamento das metodologias mais utilizadas nos trabalhos científicos sobre classificação de uso e cobertura do solo ao longo do período de 2012 a 2022 e realizar uma comparação das principais encontradas, por meio de aplicação dos métodos sobre a área de estudo da região Norte Fluminense. Com isso, visa-se identificar os pontos fortes e

²O Earth Engine é uma plataforma SIG proprietária baseada em computação em nuvem, desenvolvida pela Google e disponibilizada gratuitamente para pesquisadores. Pode ser utilizada a partir do endereço eletrônico: <https://earthengine.google.com/>

fracos das técnicas experimentadas, além de seus campos mais adequados de aplicação.

1.1 OBJETIVOS

O objetivo do trabalho é quantificar o uso de métodos de classificação de uso e cobertura do solo na literatura acadêmica e realizar um estudo comparativo com quatro dos principais identificados.

1.1.1 Objetivos específicos

O projeto apresenta, como objetivos específicos:

- Realizar uma revisão bibliográfica dos artigos publicados sobre o tema no período de 2012 a 2022;
- Identificar as metodologias mais utilizadas e selecionar quatro para realizar o estudo comparativo;
- Selecionar uma área geográfica de estudo e obter uma imagem de sensoriamento remoto da região;
- Obter uma classificação já validada por especialistas para uso como referência (“ground truth”);
- Aplicar os quatro métodos selecionados e comparar os resultados obtidos,

1.2 JUSTIFICATIVA

Os estudos de uso e cobertura do solo são realizados com múltiplas finalidades, em pesquisas de diferentes áreas, e permitem o monitoramento de informações sobre a superfície terrestre, como a evolução histórica dos tipos de cobertura, tendências para o futuro, identificação de fenômenos naturais e antrópicos, entre outras. Essas informações, por sua vez, possuem aplicabilidade em áreas como proteção ambiental, planejamento urbano, agricultura, entre outras (NEDD *et al.*, 2021).

Com a expansão desses estudos, muitos pesquisadores menos experientes se encontram aplicando as diferentes técnicas em seus trabalhos, mas sem a orientação apropriada sobre quais as melhores para as particularidades das suas pesquisas e para as características esperadas nos seus resultados. Com isso, muitas pesquisas podem aplicar métodos pouco eficientes para o tipo de estudo, ou métodos em desuso na literatura.

Nesse sentido, a presente pesquisa visa obter resultados que permitam sua aplicação como guia para a realização de trabalhos futuros sobre uso e cobertura do solo, apresentando as principais técnicas empregadas na atualidade e as principais características de algumas dentre as mais populares.

1.3 DELIMITAÇÃO DO PROBLEMA

O problema abordado pela presente pesquisa consiste da quantificação dos métodos de classificação empregados nos estudos científicos de uso e cobertura do solo mundialmente, para permitir a identificação sistemática dos mais utilizados e realização de comparações entre os mesmos. A partir dos dados de utilização de metodologias obtidos, é possível identificar tendências, como crescimento e redução do uso de determinadas técnicas, além da identificação das principais características das mais utilizadas, como suas vantagens e desvantagens para os diferentes tipos de estudo.

1.4 ESTRUTURA DA PESQUISA

Este trabalho está estruturado em cinco capítulos: O primeiro se trata de uma introdução geral sobre a pesquisa e as principais informações necessárias ao leitor para a efetiva compreensão do trabalho. Os dois capítulos subsequentes constituem dois artigos independentes, mas de temática relacionada, subdivididos com as principais seções de um artigo científico (resumo, introdução, revisão da literatura, metodologia, resultados e discussão, considerações finais e referências). O quarto capítulo apresenta as considerações finais gerais, referentes à totalidade do trabalho, e o quinto, é composto pelas referências bibliográficas para toda a dissertação.

2 ARTIGO A – O ESTADO DA ARTE DOS CLASSIFICADORES DE USO E COBERTURA DO SOLO

Resumo

O presente trabalho visa realizar um levantamento das principais metodologias aplicadas nos estudos de classificação de uso e cobertura do solo, identificar as mais utilizadas e possíveis tendências. Para isso, foi realizada uma pesquisa bibliométrica nas bases de dados de artigos científicos Scopus, na qual se realizou a contagem da utilização dos métodos de classificação presentes no período entre 2012 e 2022. Foi identificada uma tendência geral de crescimento nos estudos da área, com tendência de crescimento do emprego de métodos baseados em aprendizado de máquina e estabilização do uso das demais metodologias, sendo os métodos estatísticos também bastante aplicados, e os demais, com utilização pouco expressiva.

Palavras-chave: uso e cobertura do solo; indicadores; bibliometria.

2.1 INTRODUÇÃO

O mapeamento de uso e cobertura do solo se tornou uma importante ferramenta para estudos de geoprocessamento, com potencial para obter informações sobre variações nas áreas naturais, urbanas, agrárias, entre outras (CHAVES; PICOLI; SANCHES, 2020). Este tipo de mapeamento permite a quantificação e análise temporal de fenômenos complexos e, em conjunto com outros tipos de estudos, pode levar a uma compreensão mais profunda de tais fenômenos.

Wulder *et al.* (2018) descreveram o uso e cobertura do solo como uma série de classes hierárquicas, cada uma representando os grupos bióticos e abióticos predominantes, constituindo-se um descritor crítico da superfície terrestre. Simplificadamente, esse tipo de mapeamento pode ser entendido como a rotulação dos elementos que constituem a superfície do planeta em categorias. Desse modo, a classificação dos elementos é a principal tarefa para a elaboração de um mapeamento de uso e cobertura do solo, cujas características, qualidade e fidelidade do resultado dependem, em parte, do método empregado para tal.

Diversos métodos foram criados para realizar esse tipo de classificação, denominados classificadores. Historicamente, os classificadores dependiam de análises manuais, baseadas na experiência de especialistas e/ou execução de métodos numéricos, mas foi a partir dos adventos da computação, que novas técnicas foram criadas, permitindo a realização de análises em áreas maiores, com melhores resoluções e ainda de forma mais ágil (PHIRI; MORGENROTH, 2017).

Diante da quantidade de classificadores de uso e cobertura do solo utilizados na atualidade, torna-se relevante compreender o quanto estes são empregados na literatura científica, quais os mais utilizados na atualidade, e identificar possíveis tendências, seja para compreender o estado da arte, seja auxiliar pesquisadores a elaborarem suas metodologias, ou para tomadas de decisões estratégicas por profissionais do ramo.

Com esta justificativa, o presente trabalho objetiva-se a identificar os classificadores de uso e cobertura do solo empregados nas pesquisas científicas sobre uso e cobertura do solo catalogadas na base de dados Scopus no período entre 2012 e 2022. Como objetivos específicos, o trabalho visa obter os classificadores utilizados na atualidade, quantificar sua proporção de uso e observar se há tendências de crescimento ou redução do uso destes.

2.2 REVISÃO DA LITERATURA

As metodologias empregadas para realizar a classificação de uso e cobertura do solo variaram ao longo dos anos e com os avanços tecnológicos, tanto na área de processamento de imagens, quanto na área de captura destas. Phiri e Morgenroth (2017) apontaram que, antes da década de 1970, as análises eram realizadas visual e manualmente por especialistas, tendo as primeiras análises digitais se iniciado a partir dessa década, por meio do uso de técnicas de filtragem e de classificação numérica. Segundo os mesmos autores, o reconhecimento de padrões numéricos foi um importante desenvolvimento e serviu de base para os métodos modernos, que permitiram a utilização de abordagens de *software*, implementando diferentes tipos de classificadores.

Os métodos de classificação podem ser estruturados baseados na necessidade de fornecimento de conhecimento do usuário como entrada de dados, dividindo-se em supervisionados e não supervisionados. Podem ainda ser

classificados com base no tipo de unidade de análise, sendo esta os *pixels* ou objetos. Pandey *et al.* (2021) apresentaram alguns dos principais classificadores supervisionados (máxima verossimilhança, mapeamento de ângulo espectral, máquina de vetores de suporte, Random Forest, árvore de decisão, e distância mínima) e alguns dos principais classificadores não supervisionados (“k-means” e ISODATA).

As técnicas supervisionadas e não supervisionadas se diferenciam na necessidade do pesquisador de elaborar uma divisão de classes com base no conhecimento da região e de características de tipos de cobertura. Isto não é necessário para uma classificação não supervisionada, pois esta se baseia na formação de “clusters” naturais e posterior análise e nomeação pelo pesquisador. Já para a classificação supervisionada, é necessário que, além do conhecimento e estabelecimento de classes, que o pesquisador também obtenha um conjunto de dados para treinamento (BORRA; THANKI; DEY, 2019).

Para Jozdani, Johnson e Chen (2019), a escolha de um classificador é um aspecto crítico para o mapeamento de uso e cobertura do solo. Os autores destacaram duas categorias de classificadores muito utilizadas na atualidade: a baseada em aprendizado de máquina, incluindo métodos como “random forest”, “bagging trees” e “boosted trees”, e a baseada em *deep learning*, como o uso de “deep neural networks”.

Embora haja uma diversidade de métodos na literatura científica, estes podem ser agrupados em 5 principais categorias, de acordo com suas naturezas predominantes de funcionamento: clusterização, métodos estatísticos, aprendizado de máquina, segmentação de imagens, e manuais. Existem outros métodos que não se enquadram em nenhuma das categorias mencionadas, mas suas aplicações são mais restritas. É possível ainda utilizar mais de uma abordagem para um mesmo trabalho, seja por meio de um método combinado, seja por meio do uso de diferentes métodos em diferentes etapas do mesmo trabalho.

A clusterização é um tipo de classificação não supervisionada na qual o algoritmo agrupa os *pixels* da imagem em um número pré-definido de “clusters” naturais, com base em suas proximidades das feições e/ou similaridade estatística (BORRA; THANKI; DEY, 2019). Apesar de se tratar de uma metodologia não supervisionada, algumas pesquisas desenvolveram técnicas de clusterização que permitem algumas entradas de dados, denominando a técnica como classificação

semisupervisionada (MAI; NGO, 2018). É importante destacar que a palavra clusterização é, por vezes, empregada na literatura como sinônimo de classificação, podendo gerar confusão em algumas situações.

Os métodos estatísticos são utilizados para a classificação de uso e cobertura do solo desde seus primeiros anos de desenvolvimento (FRAZIER; SHOVIĆ, 1980), e permanecem relevantes na atualidade. Estes métodos utilizam da criação de modelos estatísticos a partir dos dados de treinamento para encontrar a categoria que se aproxima, com maior probabilidade, de cada *pixel* da imagem.

O aprendizado de máquina, ou *machine learning*, é um ramo da computação que estuda algoritmos e técnicas para automação de soluções complexas, difíceis de serem obtidas por meio de programação convencional, cujo design pode ser difícil ou impraticável de se realizar (REBALA; RAVI; CHURIWALA, 2019). De acordo com Alzubi, Nayyar e Kumar (2018), as técnicas de aprendizado de máquina possuem eficácia comprovada na solução de problemas da ciência de dados, dentre eles, os problemas de classificação e clusterização. Desta forma, o aprendizado de máquina se demonstra uma ferramenta viável como classificador, aplicada em diferentes pesquisas e considerada por Wang *et al.* (2022), como central nos estudos de mudanças de uso e cobertura do solo que a empregam, pois tratam-se de problemas de processamento de imagem única e reconhecimento de padrões.

O modelo genérico de aprendizado de máquina consiste em 6 componentes, sendo eles a preparação dos dados, a seleção de feições, a escolha do algoritmo, seleção do modelo e dos seus parâmetros, o treinamento e, por fim, a avaliação do desempenho (ALZUBI; NAYYAR; KUMAR, 2018). No ramo de uso e cobertura do solo, este processo é, por muitas vezes, simplificado por meio de ferramentas, focando na criação de um bom modelo de treinamento e, quando necessário, no ajuste de parâmetros.

A segmentação de imagem é uma abordagem de processamento de imagem que a divide em múltiplos segmentos, ou regiões distintas, conforme a diferenciação dos *pixels*. Não se trata de uma técnica de classificação em si, mas, em alguns contextos, é utilizada também para esta finalidade (HORNING, 2004). A técnica se diferencia das técnicas de classificação de imagem no sentido que as técnicas de classificação categorizam cada *pixel* em uma classe, enquanto a segmentação categoriza regiões inteiras. Contudo, é importante destacar que não há consenso na literatura sobre estas definições, e que nem sempre é feita uma distinção dos

termos, de forma que alguns trabalhos os utilizam como sinônimos. Alguns autores, como Lei e Nandi (2022), diferenciam os dois tipos de procedimentos, mas abordam a segmentação como uma forma de se classificar uma imagem.

Os métodos manuais são os mais antigos e os primeiros a serem desenvolvidos, utilizados mesmo antes do uso de satélites para captura de imagens (PHIRI; MORGENROTH, 2017). Para o uso de tais técnicas, é necessária a aplicação de conhecimentos de especialistas na topografia da região estudada, e demanda longo tempo para interpretação e produção de resultados. Apesar de se tratar de uma técnica antiga, também pode ser auxiliada por meio de SIGs, e permanecem utilizadas na atualidade, normalmente para auxiliar na precisão de imagens classificadas por meio de outras técnicas.

A classificação de uso e cobertura do solo tem evoluído ao longo dos anos, devido aos avanços tecnológicos na área de processamento de imagens e captura de dados. As análises passaram de manual e visual para digitais, por meio do uso de técnicas de filtragem e classificação numérica. Alguns trabalhos na literatura abordam a evolução histórica dos diferentes tipos de metodologias (PHIRI; MORGENROTH, 2017), ou apontam alguma tendência de aumento no uso de determinada técnica (TALUKDAR *et al.*, 2020; MAXWELL *et al.*, 2018; MORA *et al.*, 2014). Nesse sentido, a literatura científica carece de quantificação das principais metodologias empregadas, para permitir uma melhor compreensão na evolução metodológica nos estudos da área, uma lacuna explorada no presente trabalho.

2.3 METODOLOGIA

Foram realizadas buscas na base de dados de artigos científicos Scopus, visando obter os artigos de periódicos que aplicaram métodos de classificação de uso e cobertura do solo no período entre 2012 a 2022. A estratégia final utilizada foi utilizando o termo de busca “TITLE-ABS-KEY (land AND (cover OR use) AND classification) AND PUBYEAR > 2011 AND PUBYEAR < 2023 AND (LIMIT-TO (DOCTYPE, "ar"))”, que resultou em um total de 5406 documentos, os quais foram exportados em formato CSV, separados ano a ano, devido a limitações da plataforma.

Em seguida, os arquivos CSV foram convertidos para tabelas de um banco de dados do tipo SQLite, por meio da ferramenta SQLite DB Browser, e os dados foram

analisados por meio de consultas, ou “queries”, nos campos de “Author Keywords”, “Index Keywords e “Abstract”.

Foram utilizados como termos de busca, os nomes dos métodos encontrados na literatura, assim como suas abreviaturas, e registradas as quantidades de artigos presentes no banco de dados para cada consulta realizada. A relação de palavras-chave utilizadas, assim como os métodos relacionados a estas, se encontram na tabela 1 - Relação de palavras-chave pesquisadas e seus métodos associados.

Tabela 1 - Relação de palavras-chave pesquisadas e seus métodos associados

Categoria	Método	Palavras-chave
Clusterização		clustering
	ISODATA	isodata
	K-Means	kmeans, k-means, k means
	Semantic Clustering	semantic clustering
Aprendizado de máquina		machine learning
	Deep Learning	deep learning
	Random Forest	random forest, decision forest
	Decision Tree	decision tree
	Redes Neurais Artificiais	neural network
	Support Vector Machine	support vector, svm
	K-Nearest Neighbour	nearest neighbo, k-nearest, k nearest, knn
	Classification and Regression Tree	cart, classification and regression tree
Autoencoder	autoencoder, auto encoder	
Estatísticos	Máxima verossimilhança	maximum likelihood, mle
	Markov chain	markov
	mahalanobis	mahalanobis
	Principal Component Analysis	principal component, pca
Índices	Não se aplica	normalized difference, Ndvi, ndbu, ndbi, indice, index, rvi, mkt, ndwi, baei, dbi, dbsi
Outros	Overlay analisys	overlay analysis
	Autômatos celulares	cellular automata
	Segmentação de Imagem	image segmentation, thresholds
	Analytic Hierarchy Process	analytic hierarchy
Manual		visual interpretation
		manual interpretation

Fonte: Elaborado pelo autor, 2022.

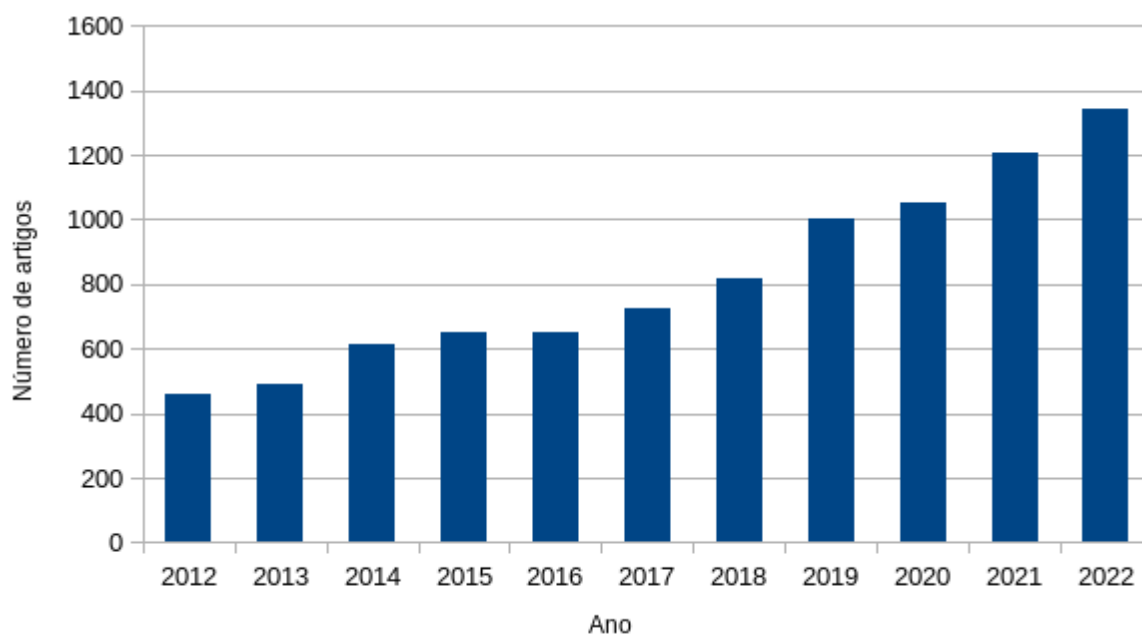
Ao finalizar a busca, foram obtidos os artigos não listados em nenhuma das buscas anteriores, e selecionados para análise manual, para encontrar tipos de classificadores não incluídos nos termos de busca. Este procedimento foi repetido até que a busca manual não retornasse novos classificadores.

Os resultados foram então tabulados e analisados quantitativamente, a fim de encontrar a quantidade de pesquisas realizadas na área, assim como os classificadores mais empregados na atualidade e suas possíveis tendências. Foram ainda elaborados gráficos para auxiliar na visualização e discussão dos resultados. Por fim, foram identificados e avaliados os 5 artigos mais citados, de acordo com informações da base de dados pesquisada, para obter uma maior compreensão da abordagem do tema nos artigos e ampliar a discussão.

2.4 RESULTADOS E DISCUSSÃO

Ao se analisar os resultados, a primeira informação relevante é a quantidade total de artigos que empregam técnicas de classificação de uso e cobertura do solo, o que é apresentado na Figura 1. Pode-se observar um expressivo aumento no período de 2012 a 2022, partindo de quantitativos abaixo dos 500 artigos, até 2013, para valores acima de 1000 artigos, a partir de 2020. Este crescimento corresponde com o esperado, diante da popularização de técnicas e ferramentas, de forma que cada vez mais pesquisadores, de diferentes áreas, realizam estes tipos de estudo.

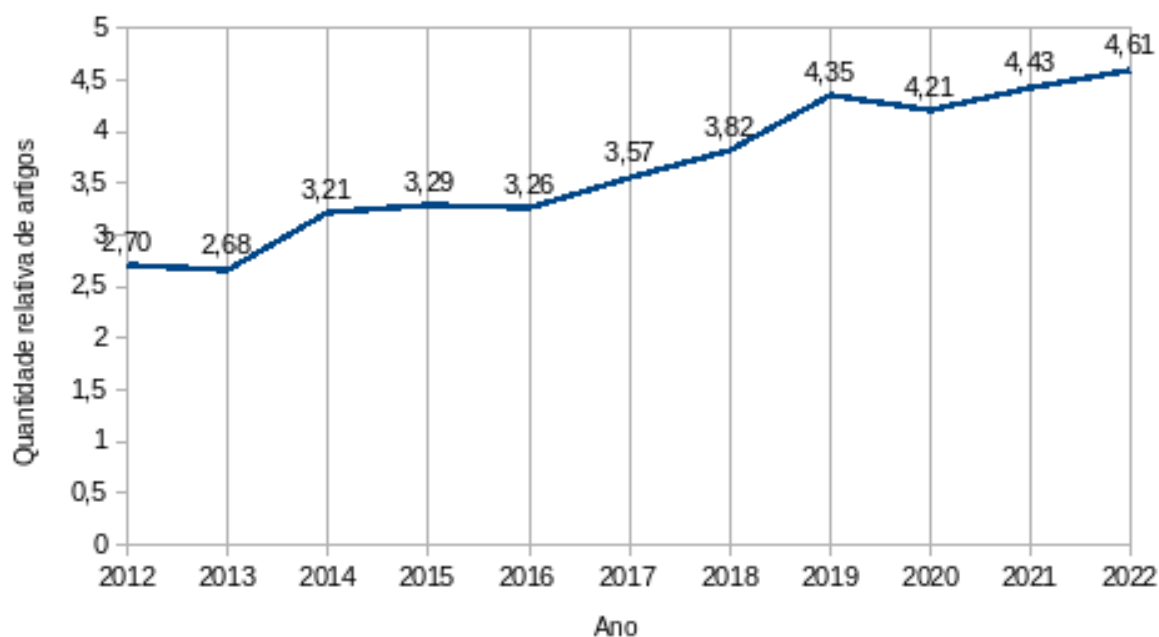
Figura 1 - Número de artigos publicados no período de 2012 a 2022 sobre o tema “classificação de uso e cobertura do solo” na base Scopus



Fonte: Elaboração própria com base na Scopus (2022).

Ao se analisar a quantidade de artigos publicados para o tema para cada 10.000 artigos publicados para todos os temas, apresentada na Figura 2, observa-se um acréscimo desde 2012, quando havia uma quantidade relativa de 460 artigos, e então subiu para 1341, em 2022. Este crescimento é um indicador de aumento no interesse pela área de pesquisa, já que houve um crescimento relativo nas publicações para o tema específico, em comparação com todos os temas da base de dados.

Figura 2 - Quantidade de artigos publicados para o tema para cada 10.000 artigos publicados para todos os temas

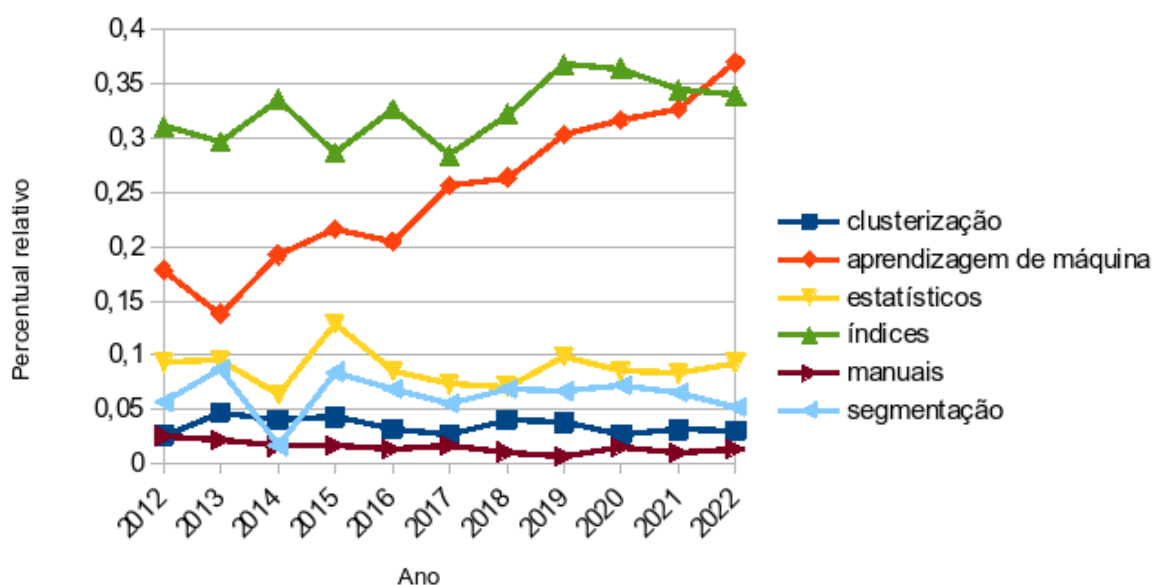


Fonte: Elaboração própria com base na Scopus (2022).

Diferentes fatores podem ter contribuído para a elevação no interesse de pesquisa. Dentre eles, há uma maior disponibilidade de fontes de dados e ferramentas na atualidade, que facilitam e democratizam a realização de pesquisas na área. Há ainda, um maior emprego de técnicas computacionais e entrada de pesquisadores de outras áreas de estudo no ramo, movidos por interesse nos tipos de abordagem empregados. Nesse sentido, a classificação de uso e cobertura do solo se apresenta como um campo de estudo em alta, com potencial para crescimento e desenvolvimento de melhorias.

Na Figura 3, são apresentadas as principais categorias de métodos de classificação de uso e cobertura do solo, com suas respectivas quantidades de artigos encontrados com suas palavras-chave presentes, para cada ano pesquisado. É importante destacar que a categoria “Índices” está relacionada ao emprego de índices calculados a partir dos valores radiométricos dos dados de origem, que, apesar de não se tratarem de um método de classificação, são, muitas vezes, empregados para tal. Nesse sentido, considerou-se relevante a inclusão desta categoria.

Figura 3 - Percentual de utilização das diferentes técnicas no período de 2012 a 2022



Fonte: Elaboração própria com base na Scopus (2022).

Observa-se uma predominância das técnicas de aprendizado de máquina, com seu percentual de uso mais do que dobrando para o período estudado, partindo de uma presença em menos de 20% do total dos estudos até 2014, para uma presença em mais de 30% dos mesmos, a partir de 2019. Já para as demais categorias, não se observa crescimento ou redução, mas uma estabilidade na sua proporção de utilização, exceto por um crescimento em torno de 5% no uso de índices de diferenças normalizadas, tendo o uso das demais técnicas apresentadas um padrão de estabilidade.

Ao comparar as Figuras 1 e 3, percebe-se que a curva de crescimento no uso de métodos de aprendizado de máquina acompanha a curva de crescimento nos estudos de uso e cobertura do solo. Ou seja, o crescimento na área em sua totalidade se deu, principalmente, pela popularização e aplicação das técnicas de aprendizado de máquina, de forma que não houve redução no uso das demais metodologias, mas um expressivo aumento desta. Este resultado reforça a possibilidade de que grande parte do aumento no interesse pelas pesquisas de uso e cobertura do solo tem se dado por pesquisadores vindos de outras áreas de estudo. O crescimento observado no crescimento desse tipo de método pode ter sido o principal motivador da elevação do número de publicações na área.

Já para a identificação das variações para os métodos individuais, assim como todos os dados obtidos, é possível observá-los a partir da tabela 2.

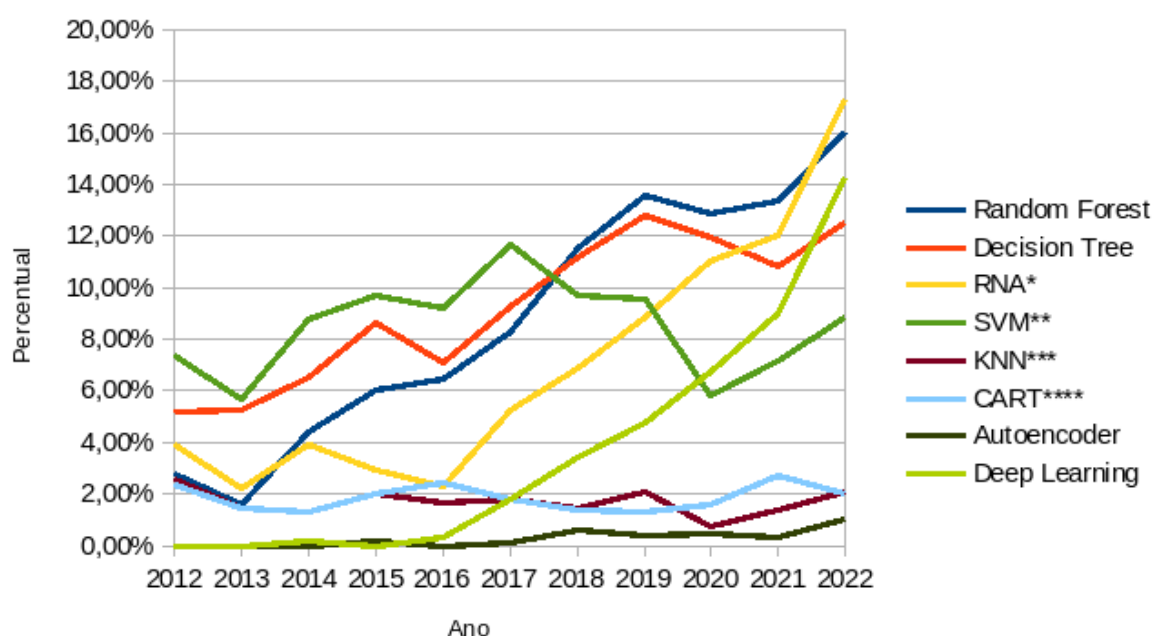
Tabela 2 - Resultados das consultas por palavra-chave

Categoria	Subcategoria	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Clusterização		2,61%	4,67%	4,07%	4,31%	3,22%	2,63%	4,04%	3,83%	2,66%	3,13%	3,68%
	ISODATA	0,00%	1,22%	0,65%	0,92%	0,46%	0,42%	0,25%	0,50%	0,10%	0,08%	0,09%
	K-Means	0,43%	0,81%	1,14%	0,46%	1,23%	0,97%	1,10%	0,50%	0,67%	0,99%	0,64%
	Semantic Clustering	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
aprendizado de máquina		17,83%	13,82%	19,22%	21,57%	20,40%	25,62%	26,35%	30,31%	31,65%	32,70%	45,26%
	Random Forest	2,83%	1,63%	4,40%	6,01%	6,44%	8,31%	11,52%	13,60%	12,83%	13,34%	16,01%
	Deep Learning	0,00%	0,00%	0,16%	0,00%	0,31%	1,80%	3,43%	4,73%	6,75%	8,98%	14,26%
	Decision Tree	5,22%	5,28%	6,51%	8,63%	7,06%	9,28%	11,15%	12,79%	11,98%	10,79%	12,51%
	Redes Neurais	3,91%	2,24%	3,91%	2,93%	2,30%	5,26%	6,86%	8,86%	11,03%	12,03%	17,30%
	SVM	7,39%	5,69%	8,79%	9,71%	9,20%	11,63%	9,68%	9,57%	5,80%	7,17%	8,83%
	K-Nearest Neighbor	2,61%	1,42%	1,30%	2,00%	1,69%	1,80%	1,47%	2,11%	0,76%	1,40%	2,12%
	CART	2,39%	1,42%	1,30%	2,00%	2,45%	1,80%	1,35%	1,31%	1,62%	2,72%	2,02%
	Autoencoder	0,00%	0,00%	0,00%	0,15%	0,00%	0,14%	0,61%	0,40%	0,48%	0,33%	1,01%
Estatísticos		9,35%	9,55%	6,35%	12,94%	8,59%	7,34%	7,11%	9,87%	8,56%	8,32%	11,41%
	Maximum Likelihood	4,13%	6,50%	2,93%	7,86%	6,60%	4,85%	4,53%	5,64%	4,94%	4,78%	8,46%
	Cadeia de Markov	1,74%	1,63%	2,12%	3,24%	1,38%	1,11%	1,35%	2,42%	2,19%	2,80%	1,75%
	Distância de Mahalanobis	0,43%	0,41%	0,65%	0,15%	0,00%	0,42%	0,49%	0,30%	0,19%	0,25%	0,37%
	PCA	0,43%	0,81%	0,98%	1,54%	0,46%	1,11%	0,98%	1,11%	1,14%	0,91%	1,84%
Índices		31,09%	29,67%	33,55%	28,66%	32,67%	28,39%	32,23%	36,76%	36,41%	34,43%	41,58%
Manuais		2,61%	2,24%	1,63%	1,69%	1,38%	1,66%	1,10%	0,70%	1,52%	0,99%	1,66%
Segmentação de imagem		5,65%	8,74%	7,82%	8,47%	6,90%	5,54%	6,86%	6,65%	7,22%	6,51%	6,35%
Outros		0,22%	0,41%	0,33%	0,77%	0,46%	0,55%	0,74%	0,91%	0,76%	1,65%	1,84%

Fonte: Elaboração própria com base na Scopus (2022).

Na Tabela 2, se observa que na categoria de aprendizado de máquina, nem todos os métodos cresceram em uso, tendo alguns sofrido reduções, como o método SVM, o mais empregado até o ano de 2017, mas que sofreu uma considerável redução nos últimos anos. Enquanto isso, os métodos baseados em redes neurais artificiais sofreram um aumento expressivo no mesmo período, chegando a ser o método mais utilizado para o último ano analisado. Para melhor visualização destes dados, foi elaborada a Figura 4, que apresenta a variação das técnicas de aprendizado de máquina identificadas ao longo do período estudado.

Figura 4 - Percentual de utilização de diferentes técnicas de aprendizado de máquina



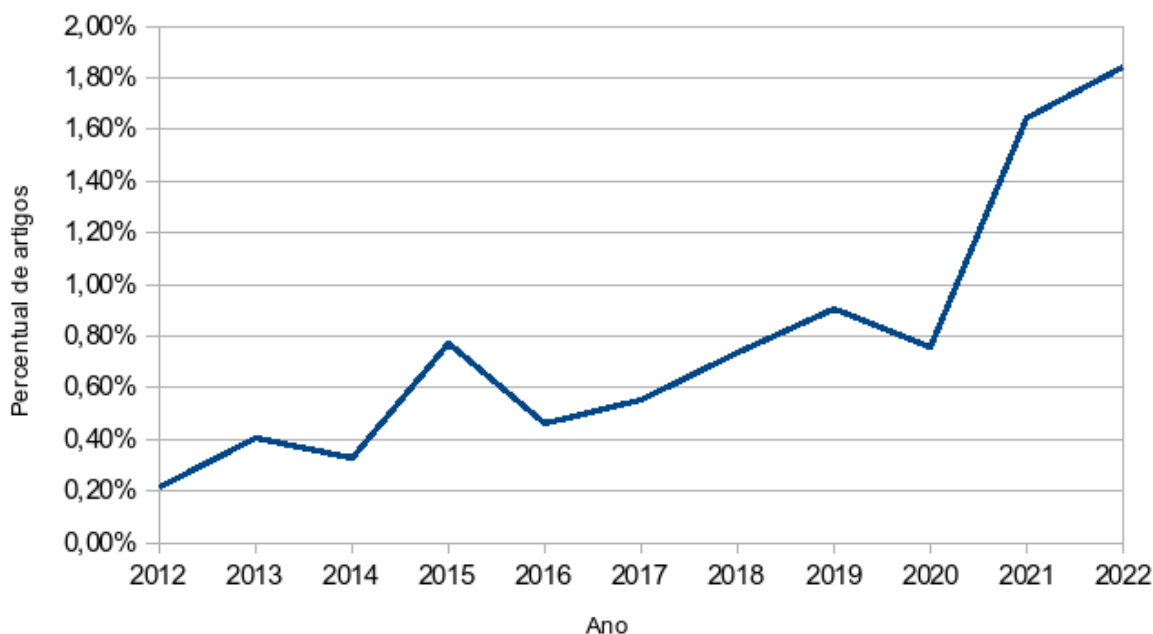
* Redes Neurais Artificiais **Support Vector Machine ***K-Nearest Neighbor ****Classification And Regression Tree

Fonte: Elaboração própria com base na Scopus (2022).

Uma possível explicação para a elevação do uso de redes neurais artificiais a partir de 2016, apesar de não haver tendência anterior, é a utilização de técnicas de *deep learning* em estudos mais recentes. Trata-se de um conjunto de técnicas baseadas em RNA, que consistem da utilização de múltiplas camadas de aprendizado, em oposição aos outros métodos, considerados “rasos”.

Ainda na tabela 2, não se observam tendências de crescimento ou redução no uso de técnicas específicas dentre métodos estatísticos, ou de clusterização, evidenciando o estado de estabilidade no uso destas. Entretanto, é possível observar um crescimento na categoria “outros”, apresentado na Figura 5.

Figura 5 - Percentual de artigos que aplicaram tipos de técnicas não categorizadas



Fonte: Elaboração própria com base na Scopus (2022).

A categoria “outros” constitui-se de artigos que possuem palavras-chave relacionadas à classificação de uso e cobertura do solo, mas não referentes aos métodos pesquisados. Os artigos desta categoria foram analisados manualmente, visando encontrar métodos não contemplados na busca e realizar a construção de novas categorias, restando aqueles que não possuíram quantitativo relevante para uma categoria individual. Desse modo, um crescimento nesta categoria pode estar relacionado ao crescimento no interesse da comunidade científica em pesquisas de uso e cobertura do solo, levando a maiores tentativas de elaboração de novos classificadores, assim como ao maior emprego de métodos mais raramente utilizados. Dessa forma, um crescimento nesta categoria pode representar uma boa saúde nesta área de estudo, com perspectivas futuras de disponibilidade de novos métodos para classificação, que, por sua vez, podem levar a melhorias nos resultados dos estudos ou menores barreiras de entrada a novos pesquisadores interessados.

Dentre os artigos encontrados, os cinco com maiores quantidades de citações foram os trabalhos realizados por Curtis *et al.* (2018), com 852 citações, Gong *et al.* (2019), com 607 citações, Thanh Noi e Kappas (2018), com 516 citações, Belgiu e Csillik (2018), com 493 citações, e Yang *et al.* (2018), com 493. Observa-se que os trabalhos mais citados são recentes, com menos de 5 anos desde sua publicação, sugerindo um dinamismo na área de estudo. Ao se analisar os trabalhos mais citados, é possível obter uma visão geral das dinâmicas do tema pesquisado, já que estes podem representar as contribuições que obtiveram maior interesse e utilidade para o desenvolvimento do conhecimento científico na área.

O artigo de Curtis *et al.* (2018), de título "Classifying drivers of global forest loss", apresenta o desenvolvimento de um modelo de classificação para agentes causadores de perda florestal, com o intuito da sua aplicação na detecção de causas desmatamento em escalas temporais. Foi utilizada a técnica de árvore de decisão, um ramo do aprendizado de máquina, em imagens fornecidas pelo *software* Google Earth. Como resultados, foi realizada a classificação da área de todo o planeta, em células quadradas de 10 km de lado, no período entre 2000 e 2015, que permitiu a identificação da principal causa de perda florestal global para mudanças permanentes do uso do solo em atividades de produção de commodities. O trabalho ainda fornece os dados subdivididos por regiões geográficas ao longo do tempo, facilitando sua utilização em novos trabalhos científicos.

Já o artigo de Gong *et al.* (2019), aborda a problemática de se trabalhar com imagens de diferentes resoluções e de origens diferentes. Os autores utilizaram um modelo previamente treinado, utilizando o método Random Forest e com imagens com resolução espacial de 30 m, provenientes do satélite LandSat 8, e aplicaram o modelo para classificação de imagens de maior resolução, com 10 m, e provenientes do satélite Sentinel 2. Como resultados, identificaram que o método Random Forest apresenta baixas perdas nesse tipo de situação, permitindo o uso de modelos treinados em conjuntos de imagens de fontes distintas, desde que com características semelhantes.

Thanh Noi e Kappas (2018), por sua vez, realizaram uma comparação dos métodos de classificação Random Forest, k-Nearest Neighbor, e SVM utilizando imagens do satélite Sentinel 2 e seis classes. Foi identificado que todos os métodos produziram resultados de precisão entre 90 a 95%, tendo o método SVM apresentado a maior precisão dentre os 3. Também na forma de estudo

comparativo, Belgiu e Csillik (2018) realizaram uma avaliação dos métodos “time-weighted dynamic time warping” (TWDTW) e Random Forest, aplicados sobre *pixels* e objetos para classificação de plantações em uma mesma área de estudo. O trabalho identificou que, apesar de resultados próximos, o TWDTW apresentou melhores resultados de forma geral, mas que o Random Forest permitiu uma melhor classificação quando houve maiores variações espectrais entre classes.

Yang *et al.* (2018) realizaram um trabalho que apresentou a elaboração de um novo banco de dados nacional de uso e cobertura do solo nos EUA. Os autores detalham os processos automatizados envolvidos na metodologia da construção dos dados, e avaliam sua precisão, identificada entre 71% a 97%.

Os trabalhos mais citados abordaram temas distintos, demonstrando a diversidade da área de estudo, que pode ter sido um dos principais fatores que levaram ao crescimento encontrado no número de pesquisas. Em linhas gerais, observa-se que os 5 trabalhos mais citados tenderam a realizar a comparações entre métodos, ou elaborar novas metodologias, ou realizar grandes mapeamentos e disponibilizar os dados obtidos para os demais pesquisadores.

2.5 CONSIDERAÇÕES FINAIS

O presente trabalho analisou os artigos científicos sobre classificação de uso e cobertura do solo, publicados na última década, compreendendo os anos de 2012 a 2022, e identificou um crescimento na realização de pesquisas na área, assim como realizou um mapeamento do uso de diferentes metodologias neste período.

O crescimento no número de publicações para o tema superou crescentemente o número de publicações para todos os temas na base de dados. Uma possível explicação se dá pela facilitação na realização de estudos no ramo, motivados por meio de fontes de dados, técnicas e equipamentos modernos, reduzindo as barreiras de entrada para novos pesquisadores, assim como atraindo pesquisadores de outras áreas de estudo.

Foi identificado um expressivo aumento no uso de técnicas baseadas em aprendizado de máquina a partir do ano de 2013. Estas se tornaram as mais populares na atualidade, presente em mais de 35% dos estudos. Já as demais técnicas, não sofreram redução ou aumento consideráveis para o mesmo período. Isto sugere que o aumento na quantidade de estudos na área de classificação de

uso e cobertura do solo se deu a partir do crescimento no uso das técnicas de aprendizado de máquina, sem uma redução, ou desinteresse, no uso das demais metodologias.

A área de estudo de uso e cobertura do solo vem se tornando mais popular nos últimos anos, e novas metodologias vêm sendo empregadas, a partir de novas técnicas e aprimoramentos na computação. Contudo, a escassez de informações quantitativas a respeito dificultava a realizações de discussões, que se apoiam, muitas vezes, na percepção pessoal de especialistas. Nesse sentido, o presente trabalho correspondeu com seu objetivo de obter dados de pesquisas científicas e identificar o perfil de uso das diferentes técnicas para classificação de uso e cobertura do solo. As informações obtidas permitem ampliar a compreensão do estado da arte, assim como auxiliar pesquisas futuras.

REFERÊNCIAS

ALZUBI, Jafar; NAYYAR, Anand; KUMAR, Akshi. Machine Learning from Theory to Algorithms: An Overview. **Journal of Physics: Conference Series**, [S. l.], v. 1142, n. 1, p. 012012, 2018. Disponível em: <http://doi.org/10.1088/1742-6596/1142/1/012012>. Acesso em: 10 nov. 2022.

BELGIU, Mariana; CSILLIK, Ovidiu. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. **Remote Sensing of Environment**, [s. l.], v. 204, p. 509–523, 2018. Disponível em: <http://doi.org/10.1016/j.rse.2017.10.005>. Acesso em: 12 nov. 2022.

BORRA, Surekha; THANKI, Rohit; DEY, Nilanjan. **Satellite Image Analysis: Clustering and Classification**. Singapura: Springer, 2019.

CHAVES, Michel E. D.; PICOLI, Michelle C. A.; SANCHES, Ieda D. Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review. **Remote Sensing**, [s. l.], v. 12, n. 18, p. 3062, 2020. Disponível em: <http://doi.org/10.3390/rs12183062>. Acesso em: 12 nov. 2022.

CURTIS, Philip G.; SLAY, Christy M.; HARRIS, Nancy L.; TYUKAVINA, Alexandra; HANSEN, Matthew C. Classifying drivers of global forest loss. **Science**, [s. l.], v. 361, n. 6407, p. 1108–1111, 2018. Disponível em: <http://doi.org/10.1126/science.aau3445>. Acesso em: 10 nov. 2022.

FRAZIER, Bruce E.; SHOVIK, HF. Statistical methods for determining land-use change with aerial photographs. **Photogrammetric Engineering and Remote Sensing**, [s. l.], v. 46, n. 8, p. 1067–1077, 1980.

GONG, Peng *et al.* Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. **Science Bulletin**, [s. l.], v. 64, n. 6, p. 370–373, 2019. Disponível em: <http://doi.org/10.1016/j.scib.2019.03.002>. Acesso em: 14 nov. 2022.

HORNING, Ned. **Land cover classification methods, Version 1.0**. American Museum of Natural History, Center for Biodiversity and Conservation, 2004. Disponível em: https://www.amnh.org/content/download/74344/1391366/file/LandCoverClassification_Final.pdf. Acesso em: 10 nov. 2022.

JOZDANI, Shahab Eddin; JOHNSON, Brian Alan; CHEN, Dongmei. Comparing Deep Neural Networks, Ensemble Classifiers, and Support Vector Machine Algorithms for Object-Based Urban Land Use/Land Cover Classification. **Remote Sensing**, [s. l.], v. 11, n. 14, p. 1713, 2019. Disponível em: <http://doi.org/10.3390/rs11141713>. Acesso em: 10 nov. 2022.

LEI, Tao; NANDI, Asoke K. **Image Segmentation: Principles, Techniques, and Applications**. [s.l.]: Wiley, 2022.

MAI, Sinh Dinh; NGO, Long Thanh. Multiple kernel approach to semi-supervised fuzzy clustering algorithm for land-cover classification. **Engineering Applications of Artificial Intelligence**, [s. l.], v. 68, p. 205–213, 2018. Disponível em: <http://doi.org/10.1016/j.engappai.2017.11.007>. Acesso em: 14 nov. 2022.

MAXWELL, Aaron E.; WARNER, Timothy A.; FANG, Fang. Implementation of machine-learning classification in remote sensing: an applied review. **International Journal of Remote Sensing**, [s. l.], v. 39, n. 9, p. 2784–2817, 2018. Disponível em: <http://doi.org/10.1080/01431161.2018.1433343>. Acesso em: 18 nov. 2022.

MORA, Brice; TSENDBAZAR, Nandin-Erdene; HEROLD, Martin; ARINO, Olivier. Global Land Cover Mapping: Current Status and Future Trends. *Em*: MANAKOS, Ioannis; BRAUN, Matthias (org.). **Land Use and Land Cover Mapping in Europe: Practices & Trends**. Remote Sensing and Digital Image Processing Dordrecht: Springer Netherlands, 2014. p. 11–30. Disponível em: http://doi.org/10.1007/978-94-007-7969-3_2. Disponível em: https://doi.org/10.1007/978-94-007-7969-3_2. Acesso em: 17 jan. 2023.

PANDEY, Prem Chandra; KOUTSIAS, Nikos; PETROPOULOS, George P.; SRIVASTAVA, Prashant K.; BEN DOR, Eyal. Land use/land cover in view of earth observation: data sources, input dimensions, and classifiers—a review of the state of the art. **Geocarto International**, [s. l.], v. 36, n. 9, p. 957–988, 2021. Disponível em: <http://doi.org/10.1080/10106049.2019.1629647>. Acesso em: 18 nov. 2022.

PHIRI, Darius; MORGENROTH, Justin. Developments in Landsat Land Cover Classification Methods: A Review. **Remote Sensing**, [s. l.], v. 9, n. 9, p. 967, 2017. Disponível em: <http://doi.org/10.3390/rs9090967>. Acesso em: 18 out. 2022.

REBALA, Gopinath; RAVI, Ajay; CHURIWALA, Sanjay. Machine Learning Definition and Basics. *Em*: REBALA, Gopinath; RAVI, Ajay; CHURIWALA, Sanjay (org.). **An**

Introduction to Machine Learning. Cham: Springer International Publishing, 2019. p. 1–17. Disponível em: http://doi.org/10.1007/978-3-030-15729-6_1. Disponível em: https://doi.org/10.1007/978-3-030-15729-6_1. Acesso em: 10 dez. 2022.

TALUKDAR, Swapan; SINGHA, Pankaj; MAHATO, Susanta; SHAHFAHAD; PAL, Swades; LIOU, Yuei-An; RAHMAN, Atiqur. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. **Remote Sensing**, [s. l.], v. 12, 2020. Disponível em: <http://doi.org/10.3390/rs12071135>. Acesso em: 18 nov. 2022.

THANH NOI, Phan; KAPPAS, Martin. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. **Sensors**, [s. l.], v. 18, n. 1, p. 18, 2018. Disponível em: <http://doi.org/10.3390/s18010018>. Acesso em: 20 nov. 2022.

WANG, Junye; BRETZ, Michael; DEWAN, M. Ali Akber; DELAVAR, Mojtaba Aghajani. Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. **Science of The Total Environment**, [s. l.], v. 822, p. 153559, 2022. Disponível em: <http://doi.org/10.1016/j.scitotenv.2022.153559>. Acesso em: 16 nov. 2022.

WULDER, Michael A.; COOPS, Nicholas C.; ROY, David P.; WHITE, Joanne C.; HERMOSILLA, Txomin. Land cover 2.0. **International Journal of Remote Sensing**, [s. l.], v. 39, n. 12, p. 4254–4284, 2018. Disponível em: <http://doi.org/10.1080/01431161.2018.1452075>. Acesso em: 22 nov. 2022.

YANG, Limin; JIN, Suming; DANIELSON, Patrick; HOMER, Collin; GASS, Leila; BENDER, Stacie M.; CASE, Adam; COSTELLO, Catherine; DEWITZ, Jon; FRY, Joyce. A new generation of the United States National Land Cover Database: requirements, research priorities, design, and implementation strategies. **Isprs Journal Of Photogrammetry And Remote Sensing**, [s.l.], v. 146, p. 108-123, dez. 2018. Disponível em: <http://dx.doi.org/10.1016/j.isprs.2018.09.006>. Acesso em: 10 nov. 2022.

3 ARTIGO B – ESTUDO COMPARATIVO DE CLASSIFICADORES DE USO E COBERTURA DO SOLO

Resumo

O presente trabalho visa comparar quatro dos principais métodos de classificação de uso e cobertura do solo empregados na literatura científica (máxima verossimilhança, SVM, Random Forest, e redes neurais artificiais) e realizar um comparativo entre eles, para identificar suas principais características, semelhanças e divergências. A região geográfica selecionada para estudo foi a região Norte Fluminense e os dados de referência utilizados foram o mapeamento de uso e cobertura do solo do projeto MapBiomass. A classificação foi realizada utilizando o mesmo conjunto de dados para o treinamento em todos os métodos e a comparação foi realizada por meio do uso de matrizes de erros e análise manual. O comparativo permitiu identificar que a eficiência dos métodos depende da diferenciação espectral entre as classes a serem identificadas, mas com particularidades entre cada um. Em geral, as classes com maiores distinções espectrais foram classificadas com maior acurácia, enquanto as com menores distinções geraram maiores erros. O método MLE apresentou os melhores resultados para classes bastante distintas espectralmente, enquanto o Random Forest apresentou alguma eficácia para distinguir classes com pouca diferenciação espectral, mas dificuldade para classificação de cobertura não vegetal, cuja classe apresentou alta dispersão em relação à média de valores dos seus *pixels*. Os métodos SVM e RNA, não apresentaram eficácia para classes espectralmente similares, mas o SVM apresentou a maior acurácia na classificação de cobertura não vegetal.

Palavras-chave: uso e cobertura do solo; classificação; SVM; Random Forest.

3.1 INTRODUÇÃO

O uso e cobertura do solo estão relacionados à forma como se apresenta a superfície terrestre, sendo a cobertura do solo relacionada às formações naturais, como vegetação e corpos de água, e o uso do solo, relacionado às formações antrópicas, como áreas urbanas e atividade agrícola. Para Meyer e Turner (1994),

ambos fenômenos estão interligados, uma vez que os objetivos humanos de uso do solo produzem efeitos sobre a cobertura do mesmo, de forma que esta relação se torna um ponto de interseção entre as ciências naturais e sociais.

O estudo do uso e cobertura do solo é considerado um campo dinâmico e de rápidas mudanças. A quantidade de pesquisas realizadas na área cresceu rapidamente, a uma taxa de 36,5% ao ano, entre 1991 a 2015 (ZHANG *et al.*, 2019). Junto desse crescimento na quantidade de pesquisas realizadas, ocorreram avanços tecnológicos que trouxeram novas fontes de dados e permitiram a execução de novas metodologias, partindo de um cenário de difícil acesso a bases de dados geoespaciais, que, por sua vez, possuíam baixa resolução; baixo nível de integração entre pesquisas; limitações de recursos computacionais e ferramentas de processamento; e apresentação de resultados de forma estática, para um cenário de alta disponibilidade de fontes de dados, com altas resoluções; alta integração entre pesquisas; alta disponibilidade de recursos computacionais e ferramentas; além de resultados dinâmicos e interativos, focados no formato *web* (WULDER *et al.*, 2018).

Essas transformações evidenciam o interesse dos pesquisadores em estudos de uso e cobertura do solo. Contudo, diante de um cenário de alta disponibilidade de recursos, ferramentas e metodologias, novos pesquisadores da área podem encontrar dificuldades para encontrar a metodologia mais adequada para o seu projeto e, sem guiamento adequado, podem utilizar de técnicas ineficientes para o objeto de estudo.

Nesse sentido, o presente trabalho realizou uma comparação entre quatro das metodologias mais utilizadas para classificação de uso e cobertura do solo na atualidade. Três delas são baseadas em métodos de aprendizado de máquina (Random Forest, SVM e Redes Neurais Artificiais), que têm crescido fortemente no uso, e uma consiste em um método estatístico, que vem sendo usado há mais tempo e permanece como um método relevante (máxima verossimilhança).

O trabalho busca, por meio da comparação entre os quatro métodos supracitados, identificar suas principais características, pontos fortes e fracos, permitindo estimar seus melhores campos de aplicação, auxiliando na seleção de metodologia para estudos futuros, no sentido que pode prover de embasamento para a escolha.

3.2 REVISÃO DA LITERATURA

3.2.1 Classificação de uso e cobertura do solo

O uso e cobertura do solo é definido por Meyer e Turner (1992) como a união de dois campos de estudo que exploram as alterações na superfície terrestre e sua cobertura biótica. Para Duhamel (2012), este tipo de classificação permite responder a perguntas comuns que surgem ao se observar uma porção da superfície terrestre, como “O que é isso?” e “Para que serve isso?”. A primeira pergunta se refere diretamente ao conceito de cobertura do solo, ou seja, descrever o que está apresentado naquela porção de área. A segunda pergunta se refere ao conceito de uso do solo, que apresenta a função daquela porção de superfície para as atividades humanas.

O processo tradicional de classificação de uso e cobertura do solo consiste em 5 etapas, iniciando com a aquisição de dados, que são preprocessados a fim de gerar imagens apropriadas para a execução dos algoritmos de classificação, cujo resultado é avaliado quanto à sua exatidão, e, por fim, é gerado o mapa classificado (WULDER *et al.*, 2018).

Os dados de sensoriamento remoto são normalmente obtidos por meio de satélites que carregam equipamento de sensoriamento radiométrico e obtém fotografias multibanda de porções da superfície do planeta em intervalos de tempo, enquanto orbitam a Terra. As imagens são, a seguir, recebidas por meio de ondas eletromagnéticas, processadas e armazenadas em coleções de dados (YOUNG *et al.*, 2017).

Devido ao alto custo dessas missões espaciais, elas são consideradas uma ferramenta estratégica de poderio geopolítico, por vezes, realizadas por meio de esforços conjuntos entre instituições ou países, além de missões realizadas para projetos comerciais (BELWARD; SKØIEN, 2015). Portanto, as imagens obtidas nem sempre são oferecidas gratuitamente para o público geral, o que pode se tornar um obstáculo para pesquisas que as requerem, especialmente quando realizadas por pesquisadores de países menos desenvolvidos economicamente. Alguns dos satélites mais populares em uso, cujas imagens podem ser obtidas sem custo, são as famílias Sentinel, Landsat e MODIS (RADOČAJ *et al.*, 2020).

As imagens são obtidas pelas instituições ou grupos responsáveis pelos satélites e, por vezes, disponibilizadas em seus próprios portais, sob diferentes formas de acesso. Contudo, as entidades responsáveis pela obtenção das imagens costumam relicenciá-las para que outros órgãos ou empresas possam fornecê-las de maneira mais simplificada para o usuário final. Algumas fontes gratuitas de imagens de sensoriamento remoto incluem o portal EarthExplorer da USGS (2022), o Worldview da EOSDIS (2023), o portal brasileiro do INPE (2023), que provê imagens de várias fontes, incluindo o Satélite Sino-Brasileiro de Recursos Terrestres, e o Earth Engine da Google (2023a).

As imagens obtidas pelos satélites são armazenadas em formatos próprios, que variam para cada projeto. Entretanto, tais formatos são de mais difícil manipulação para os usuários. Por isso, os portais utilizam, em sua maioria, o formato Geotiff, uma especificação de domínio público criada na década de 1990 que utiliza de um formato maduro e independente de plataforma, o TIFF, adicionando-lhe metadados que permitem sua utilização em aplicações geográficas. Atualmente é um formato amplamente utilizado pelas comunidades de ciências da terra e geoespaciais e sua padronização é definida pelo Open Geospatial Consortium (OGC GEOTIFF STANDARD | OGC, 2023).

Os recortes espaciais são normalmente oferecidos na forma de cenas ou de ladrilhos (“tiles”). As cenas são as regiões abrangidas pelas fotografias dos satélites, sem controle ou repetibilidade, apenas demarcadas pelos metadados de localização e horário. Já o sistema de ladrilhos consiste em porções de área retangulares justapostas para abranger o território terrestre. São normalmente construídos a partir do processamento das cenas visando permitir a seleção determinística de regiões espaciais (MCNEILL; HALE, 2017).

A captura de imagens por sensoriamento remoto apresenta desafios quanto à sua precisão, sofrendo erros estatísticos no processo de medição, interferências, variações por efeitos de posicionamento solar e ângulo da captura da imagem, alterações geométricas devido à velocidade do satélite e curvatura da Terra, entre outros. As imagens obtidas diretamente pelos satélites são chamadas de imagens “raw”, ou cruas. A quantidade de distorções presentes nestas imagens, normalmente as impedem de serem usadas diretamente em SIGs, sendo necessária a realização de uma etapa de pré-processamento para gerar imagens prontas para a análise e com distorções em limites aceitáveis (DAVE; JOSHI; SRIVASTAVA, 2015).

Young *et al.* (2017) destacaram os principais tipos de correção aplicadas em imagens de sensoriamento remoto, com um enfoque nas obtidas a partir da família de satélites Landsat, sendo elas as correções dos tipos:

- Geométricas: buscam garantir o exato posicionamento da imagem e incluem georreferenciamento e ortorretificação;
- Radiométricas (absoluta ou relativa): amenizam os efeitos sensoriais, solares, atmosféricos e topográficos;
- Conversão para radiância: Também chamada de calibração de sensor, converte os números digitais e valores relativos às grandezas físicas por meio de fatores de conversão;
- Solares: Relativa às influências de fatores solares (em grande parte, angulares) nos *pixels*, converte os valores de radiância no sensor em valores equivalentes ao topo da atmosfera;
- Atmosféricas: Reduzem as perturbações causadas pelos efeitos de absorção e espalhamento da energia refletida por interações com partículas atmosféricas, como gases, vapores e aerossóis;
- Topográficas: um processo radiométrico que considera os efeitos de iluminação causados pelo relevo terrestre, tornando a imagem “plana”;
- Preprocessamento para índices espectrais: corrigem efeitos aditivos e diferenciação entre bandas espectrais.

Shrawankar e Khandare (2016) abordaram ainda outros tipos de correções que podem ser realizadas, como melhoramento da imagem, melhoramento de contraste, e transformações matemáticas, a depender do escopo do projeto e das suas especificidades. Portanto, para que se defina o nível de preprocessamento necessário para um conjunto de imagens aéreas, é necessário terem sido levantados os requisitos necessários, uma vez que, tanto a falta de etapas de processamento quanto a aplicação de métodos desnecessários, podem comprometer os resultados obtidos.

O processo de classificação consiste em atribuir, a cada *pixel* da imagem, uma categoria, seja na forma de metadados ou sobrescrevendo seu valor original. Para isso, faz-se uso de técnicas de classificação, ou classificadores. Horning (2004)

destacou que existem muitos métodos de classificação, cada um com diferentes características e aplicações, de forma que não há uma abordagem que possa ser considerada melhor do que a outra, mas que a escolha deve ser feita baseada na disponibilidade, *software* utilizado e familiaridade do pesquisador com o método.

É ainda importante, para a escolha da metodologia, considerar as características da imagem a ser classificada. Nesse sentido, Naboureh *et al.* (2021, tradução nossa) descreveram que “reconhecer similaridades e variâncias espectrais em um espaço espectral multidimensional, seguido de conectá-las a tipos de cobertura do solo, é identificado como um processo para a classificação de sensoriamento remoto”. Ou seja, diferentes classificadores podem apresentar diferentes desempenhos, conforme as características da imagem, se apresentando como mais um critério para a escolha do método.

A avaliação de acurácia visa verificar a qualidade dos dados utilizados na classificação e avaliar a precisão dos resultados obtidos. Este processo pode, assim como a classificação, ser realizado de diferentes formas. Janssen e Vanderwel (1994) descreveram alguns métodos, que incluem amostragem e conferência em campo, matriz de erros, processos baseados em distribuição binomial e testagem de hipóteses. O resultado final da classificação é um mapa temático que representa as diferentes classes de uso e cobertura do solo, presentes na área de interesse, como áreas urbanas, florestas, culturas agrícolas, corpos d'água, entre outros. Este mapa temático pode ser utilizado para diversas aplicações, como o planejamento territorial, monitoramento ambiental, estudos de mudanças climáticas, entre outros.

Devido à sua importância para diferentes áreas e estudos, algumas instituições realizam mapeamentos e divulgam dados de uso e cobertura do solo para regiões extensas. Alguns exemplos incluem o mapeamento realizado pelo IBGE (2020), que publica dados de uso e cobertura do solo para todo o território brasileiro, a USGS (2018), que publica dados para os EUA, a EUROPEAN SPACE AGENCY (2023), que divulga dados de uso e cobertura do solo para todo o globo terrestre, entre outras. No Brasil, um projeto que se destaca é o MapBiomass (2019), uma rede colaborativa, formada por ONGs, universidades e startups de tecnologia que elabora e disponibiliza para acesso livre mapas de uso e cobertura do solo de todo o território brasileiro, com dados a partir de 1985. Este tipo de iniciativa permite que pesquisadores utilizem as imagens para a construção de novos estudos, sem

precisarem realizar o próprio mapeamento, e, dessa forma focando nas características a serem observadas das suas áreas de estudo.

3.2.2 Métodos de classificação supervisionada para uso e cobertura do solo

A classificação supervisionada é uma técnica que visa prever a classe de um dado com base em dados previamente rotulados. É necessário que o analista forneça as classes a serem utilizadas para a classificação, assim como o conjunto de dados de treinamento, permitindo que o algoritmo identifique, para cada *pixel* da imagem, qual das classes mais se aproxima das suas características espectrais (DHINGRA; KUMAR, 2019). Alguns dos métodos de classificação supervisionada incluem o método da máxima verossimilhança (MLE), o SVM, o Random Forest, e o uso de redes neurais artificiais (RNA).

O método da máxima verossimilhança opera considerando que cada classe determinada, para cada uma das bandas de dados da imagem, pode ser descrita como uma distribuição normal, de forma que cada *pixel* pode ser classificado conforme a classe com que este possui maior probabilidade de pertencer (BAYARSAIKHAN *et al.*, 2009).

Otukei e Blaschke (2010) consideram o MLE como um método paramétrico bem conhecido, e destacam sua característica de considerar a variância e covariância nas distribuições de classes e para dados distribuídos de forma estatisticamente normal. Esta característica permite um melhor desempenho do método, mas pode não trazer bons resultados quando os dados não constituírem uma distribuição normal.

O método Random Forest utiliza de uma combinação de várias estruturas em forma de árvore, que “votam” em qual classe pertence o *pixel*, o qual recebe a classificação decidida pela maioria das estruturas (BREIMAN, 2001). Cutler, Cutler e Stevens. (2011) consideram o método atrativo devido uma série de características. Do ponto de vista computacional, os autores destacam sua velocidade de treinamento e execução, dependência apenas de 1 ou 2 parâmetros, a capacidade de lidar com regressão e multiclasse, presença de uma forma de predição de erro, possibilidade de uso para problemas com muitas dimensões e facilidade de implementação paralela. Já do ponto de vista estatístico, os autores destacam a capacidade de medir importância de variáveis, peso diferencial de classe, imputação

de valores faltantes e visualização. Dessa forma, o método Random Forest se apresenta como um método vantajoso, já que possui robustez estatística, é fácil de utilizar, e não é tão demandante computacionalmente, além de fornecer recursos para usuários mais avançados.

O método SVM foi trabalhado por Cortes e Vapnik (1995), a partir de ideias já existentes, desenvolvendo uma técnica para problemas de classificação baseados em vetores de entrada mapeados não-linearmente a um espaço de feições de muitas dimensões. Noble (2006, tradução nossa) descreveu as máquinas de vetor de suporte como “um algoritmo (receita) para maximização de uma função matemática particular com respeito a um conjunto de dados preestabelecido”.

Huang, Davis e Townshend (2002) avaliaram o desempenho do método SVM para classificação de uso e cobertura do solo e identificaram uma boa acurácia nos resultados, especialmente quando maiores quantidades de variáveis foram empregadas. Entretanto, os autores observaram que o método é intensivo computacionalmente, já que necessitou, para o seu estudo, de tempo superior a um dia para a execução do algoritmo. É importante destacar que desde a realização do estudo, as ferramentas computacionais se tornaram mais eficientes, de forma que o tempo de processamento em equipamentos modernos pode ser substancialmente reduzido.

Por fim, as redes neurais artificiais são uma técnica de aprendizado de máquina inspirada em redes biológicas de neurônios, que utiliza de inúmeros processadores simples com muitas interconexões (GUPTA, 2013). O uso de redes neurais artificiais para classificação de uso e cobertura do solo foi proposto por Yoshida e Omatu (1994) como um método mais poderoso do que o método Bayesiano tradicional, e com a vantagem de não necessitar de elevados conhecimentos geográficos sobre a área de estudo.

3.2.3 Ferramentas populares para classificação de uso e cobertura do solo

Os estudos de uso e cobertura do solo são normalmente apoiados por sistemas de informação geográfica (SIG), que permitem processar grandes volumes de informação e aplicação de algoritmos complexos. Entre os SIGs mais populares, o QGIS e o Earth Engine se demonstram ferramentas influentes na popularização deste tipo de estudo.

O QGIS é um SIG de código aberto que oferece uma ampla gama de ferramentas para visualização, análise e manipulação de dados geográficos (QGIS DEVELOPMENT TEAM, 2023). O *software* consegue lidar com dados vetoriais e do tipo *raster*, e pode ser integrado a bancos de dados geográficos. Uma das suas principais características se dá por ser distribuído sob uma licença livre (QGIS, 2023a), permitindo que qualquer indivíduo possa executá-lo, copiá-lo, modificá-lo e redistribuí-lo conforme os termos definidos. Esta característica, apoiado pelo formato de distribuição sem custo para o usuário, levaram à formação de uma comunidade ativa de desenvolvedores e usuários que criam conteúdo, fornecem dados e constroem extensões para adicionar novas funcionalidades ao QGIS. Sua página oficial de *plugins* consta, em dezembro de 2022, com 1739 extensões (QGIS, 2023b) que podem ser baixadas gratuitamente, para adicionar diferentes funcionalidades, inclusive para classificação de imagens.

É importante destacar que o seu formato de desenvolvimento e distribuição contribui para a remoção de barreiras de acesso à tecnologia, de forma que até mesmo pesquisadores com poucos recursos, muitas vezes de países em desenvolvimento ou subdesenvolvidos, possam realizar pesquisas científicas, sem prejuízo técnico em relação a pesquisadores de países desenvolvidos, exceto pelas possíveis limitações de acesso a *hardware*.

Já o Earth Engine, trata-se uma plataforma de processamento de dados geográficos baseada em nuvem desenvolvida pela Google. O *software* permite acesso a uma gama de conjuntos de dados de sensoriamento remoto preprocessados, incluindo imagens dos programas Landsat, Sentinel e MODIS, além de oferecer ferramentas para análise dos dados, incluindo classificação de uso e cobertura do solo (GORELICK *et al.*, 2017).

Diferentemente do QGIS, o Earth Engine é um *software* distribuído via acesso *web*, sob licença proprietária, com algumas restrições de acesso, como a necessidade de cadastro. A ferramenta possui custo para o usuário, mas a Google a oferece gratuitamente (sem alguns recursos da ferramenta) a indivíduos, organizações sem fins lucrativos, pesquisadores e outros usuários de impacto para seus projetos de pesquisa, para não comerciais (GOOGLE, 2023b).

Desse modo, apesar de a ferramenta não ser livre, o processamento dos dados em nuvem permite que os usuários não necessitem de recursos avançados de *hardware*, funcionando até mesmo em computadores mais antigos. Isto, por sua

vez, também se torna uma forma de inclusão para pesquisadores com poucos recursos tecnológicos, mas com acesso à *internet*. Entretanto, cabe ressaltar que, como a ferramenta não é livre, é possível que a empresa modifique suas políticas no futuro. Ou seja, o acesso gratuito não é uma garantia do usuário.

Adicionalmente, o Earth Engine possui uma interface de programação de aplicativos (API) que permite aos desenvolvedores integrar as capacidades do Earth Engine em seus próprios aplicativos e ferramentas (GORELICK *et al.*, 2017). Dentre as possibilidades que a API fornece, estão o uso do serviço de computação em nuvem oferecido a partir de clientes de *software* livre e uma facilitada integração com diferentes pesquisas. O próprio QGIS possui uma extensão, desenvolvida por Donchyts *et al.* (2023) para integração com o Earth Engine, permitindo o uso de ambas ferramentas simultaneamente, extraindo as melhores vantagens de cada uma.

3.3 METODOLOGIA

A área de estudo escolhida é composta pela mesorregião classificada pelo IBGE como Norte Fluminense, localizada ao norte do estado do Rio de Janeiro e composta pelos municípios de Campos dos Goytacazes, Carapebus, Cardoso Moreira, Conceição de Macabu, Macaé, São Fidélis, São Francisco de Itabapoana, São João da Barra e Quissamã. Foi obtido o recorte da região distribuído pelo IBGE (2021) em formato *shapefile*^C. Como referência, ou “ground truth”, foram utilizados os dados do estudo de uso e cobertura do solo do projeto MapBiomass (2019) Os dados são compostos por mapas com resolução de 30 m por *pixel* classificados em 5 macroclasses e suas subdivisões.

Já as imagens de sensoriamento remoto, estas foram obtidas por meio do *software* Earth Engine (2023a), que disponibiliza imagens de diferentes satélites e permite a manipulação das mesmas por meio de linguagem JavaScript. Foram obtidas imagens do satélite Sentinel-2 ao longo do ano de 2021 que possuíssem cobertura de nuvens inferior a 20% e então aplicado um filtro mediano, obtendo-se assim um resultado que representa a mediana dos *pixels* capturados durante o ano

^CO *shapefile* é um padrão de arquivo utilizado na área de geoprocessamento que consiste em vários arquivos individuais que, em conjunto, permitem armazenar informações sobre a geometria dos objetos geográficos, como pontos, linhas ou polígonos, bem como permitem armazenar informações de atributos, como nome, área ou população.

de 2021 na mesorregião Norte Fluminense. Esta técnica permite amenizar os efeitos sazonais e obter uma imagem representativa anual. Foram selecionadas as bandas de imagem B2, B3, B4, B8, B11 e B12, correspondentes a tons de azul, verde, vermelho, tons próximos de infravermelho (NIR), e infravermelhos de onda curta (SWIR 1 e SWIR 2). O algoritmo utilizado encontra-se no Apêndice A.

Todas as imagens obtidas foram recortadas utilizando os limites geográficos da mesorregião do Norte Fluminense, a partir do *shapefile* obtido. Para a realização do treinamento dos algoritmos utilizados, foram selecionadas amostras aleatórias de regiões contíguas selecionadas manualmente, pertencentes a cada classe objetivada.

Antes da etapa a classificação, foram analisadas as características espectrais das classes da imagem, a fim de auxiliar na compreensão do funcionamento dos métodos empregados. A análise foi realizada a partir da plotagem espectral das classes envolvidas. Foi ainda utilizado o índice de similaridade Bray-Curtis (BRAY; CURTIS, 1957) para quantificar a diferenciação entre cada uma.

Para a classificação pelo método Random Forest, foi utilizado o *software* Earth Engine, por meio da função “*smileRandomForest*” oferecida pela plataforma, que admite ajustar o número de árvores do algoritmo como parâmetro. Foram experimentados os valores de 10, 100 e 500. Para a classificação pelo método SVM, foi também utilizado o *software* Earth Engine, que oferece a função de classificador denominada “*libSVM*”. Foi utilizado o *kernel* linear, cuja função oferecida não oferece ajuste de parâmetros. Já para a realização da classificação pelo método da máxima verossimilhança, foi utilizado o *software* QGIS e o *plugin* Semi-Automatic Classification, desenvolvido por Congedo (2021). Por fim, foi utilizado o QGIS e o *plugin* Neural Network MLPClassifier, desenvolvido por Crabbé (2020), para a realização da classificação pelo método de redes neurais artificiais. Foram utilizados 10 neurônios na camada escondida, função de ativação logística e 200 iterações máximas.

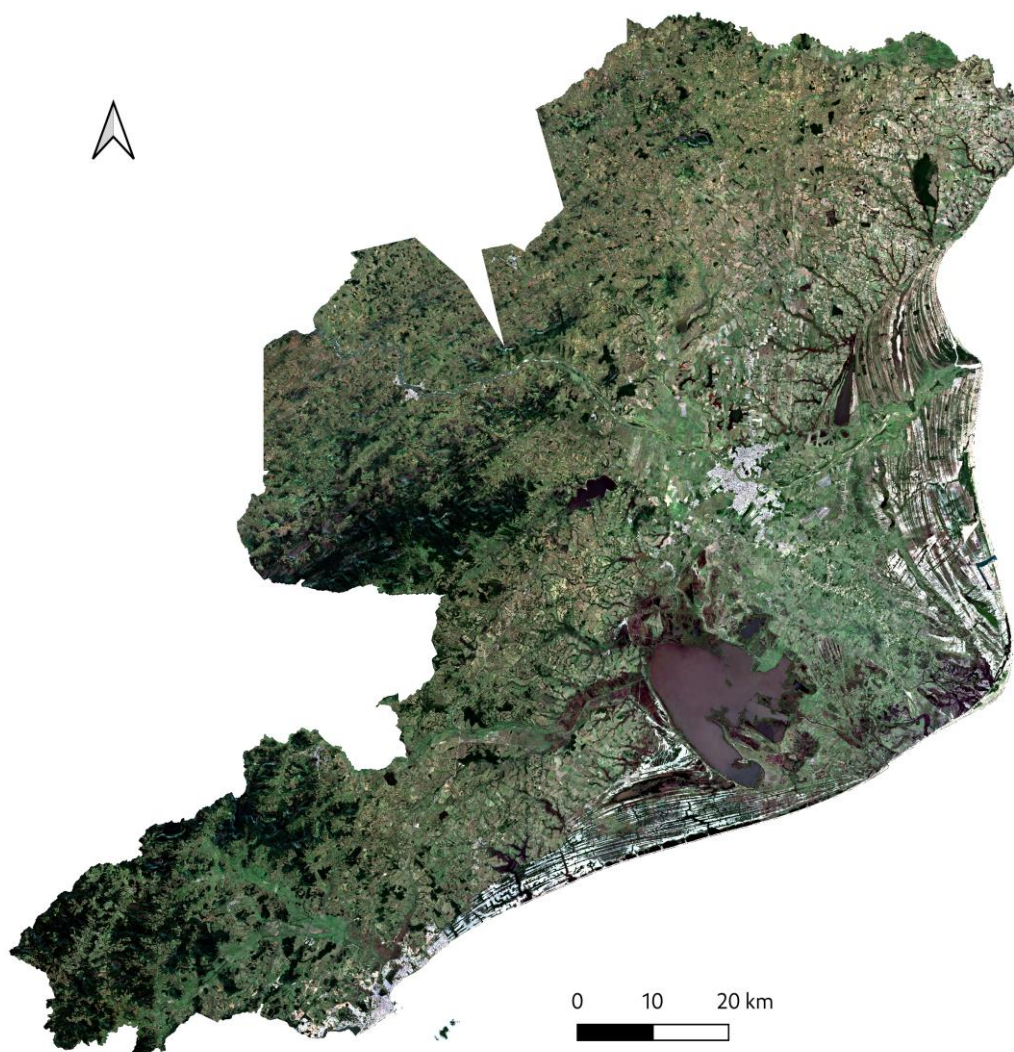
Para a análise da acurácia dos resultados, foram utilizadas as técnicas de matriz de confusão, e os valores associados de taxa de verdadeiros positivos, sensibilidade, erro de omissão e erro de comissão (BARSI *et al.*, 2018), assim como o coeficiente *kappa*. Este conjunto de técnicas é considerado um padrão na avaliação de acurácia de classificação de imagens (RWANGA; NDAMBUKI, 2017).

Adicionalmente, foi utilizada de análise manual para ampliar o estudo. Os resultados foram comparados e discutidos.

3.4 RESULTADOS E DISCUSSÃO

A imagem base para o estudo foi elaborada por meio da aplicação de um filtro mediano em todas as imagens obtidas pelo Sentinel-2 na região durante o ano de 2021. Esta imagem é apresentada na Figura 6 e oferece uma representação mediana da região Norte Fluminense para todo o ano de 2021.

Figura 6 - Imagem da região de estudo obtida a partir do Sentinel-2 (exibindo bandas

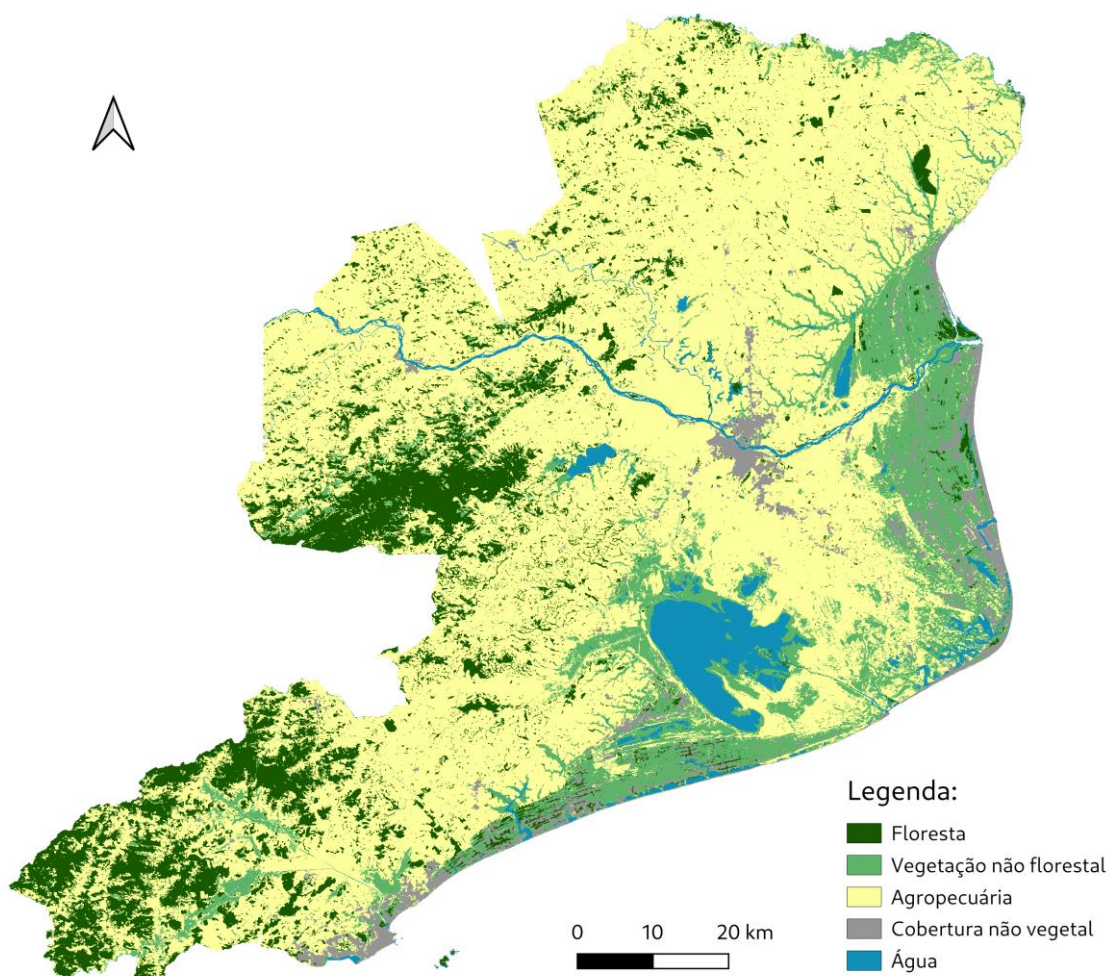


vermelho, verde e azul)

Fonte: Elaboração própria com base no Sentinel-2 (2022).

Os dados obtidos a partir do projeto MapBiomas foram utilizados neste trabalho como imagens de referência, também conhecidas como *ground truth*. Com estes dados, permitiu-se selecionar manualmente conjuntos de amostras da imagem de satélite original representando cada classe para treinamento dos algoritmos. A imagem de referência é apresentada na Figura 7.

Figura 7 - Imagem da região de estudo obtida a partir do projeto Mapbiomas e utilizada como referência

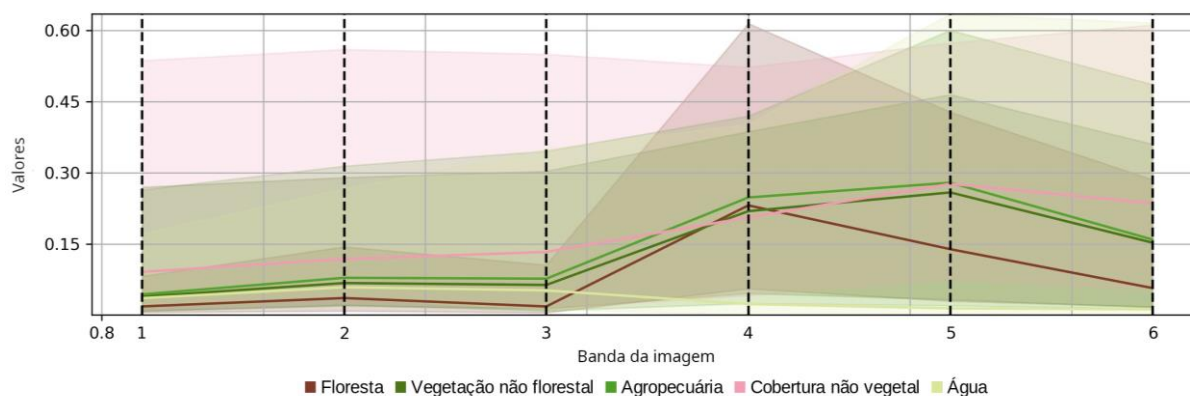


Fonte: Elaboração própria com base no MapBiomas (2022).

A partir dos dados utilizados para treinamento dos algoritmos, é possível identificar as características espectrais das diferentes classes estudadas, permitindo melhor compreender as características e distinções entre cada uma delas. Na Figura 8, é apresentada uma plotagem espectral em que são exibidas as concentrações de

valores para cada banda de imagem, dentro cada classe. Nesta figura, cada linha representa a média de valores dos *pixels* associados a cada classe, e cada faixa da mesma cor de cada linha, representa a dispersão.

Figura 8 - Plotagem espectral para os dados de treinamento



Fonte: Elaboração própria com base no Sentinel-2 (2022).

Identifica-se que a classe de água possui diferenciação elevada em relação às outras classes nas bandas numeradas como 4 (NIR), 5 (SWIR 1) e 6 (SWIR 2), enquanto a classe de floresta se diferencia principalmente nas bandas 5 (SWIR 1) e 6 (SWIR 2), e a classe de cobertura não vegetal se diferencia nas bandas 1 (azul), 2 (verde) e 6 (SWIR 2). Já as classes de vegetação não florestal e área de uso agropecuário, possuem baixa diferença espectral entre si, em todas as bandas de imagem. Isto torna mais desafiador a correta diferenciação entre estas duas classes e permite verificar o desempenho das metodologias estudadas para as duas condições, a de alta diferenciação espectral entre classes, e a de baixa diferenciação. Outro fator observável na Figura 8, é a dispersão elevada dos valores, o que pode dificultar o funcionamento de alguns modelos, principalmente para a classe de cobertura não vegetal, que possui dispersão elevada nas bandas em que há sua maior diferenciação média em relação às outras classes.

Esta baixa diferenciação entre as classes de agropecuária e vegetação não florestal pode ser confirmada ao realizar o cálculo do índice de similaridade Bray-Curtis entre as assinaturas das classes, apresentada na Tabela 3. Com base nestes dados, observa-se que as duas classes possuem uma similaridade acima de 90%, gerando um fator dificultador para os algoritmos utilizados. A partir destes dados, espera-se que os algoritmos tenham maior precisão na classificação de água e

floresta e maiores dificuldades na classificação de agropecuária e vegetação não florestal. Para a classe de cobertura não vegetal, cuja diferenciação média é moderada, mas a dispersão dos valores dos *pixels* é elevada, espera-se que o desempenho possa variar entre baixo a moderado.

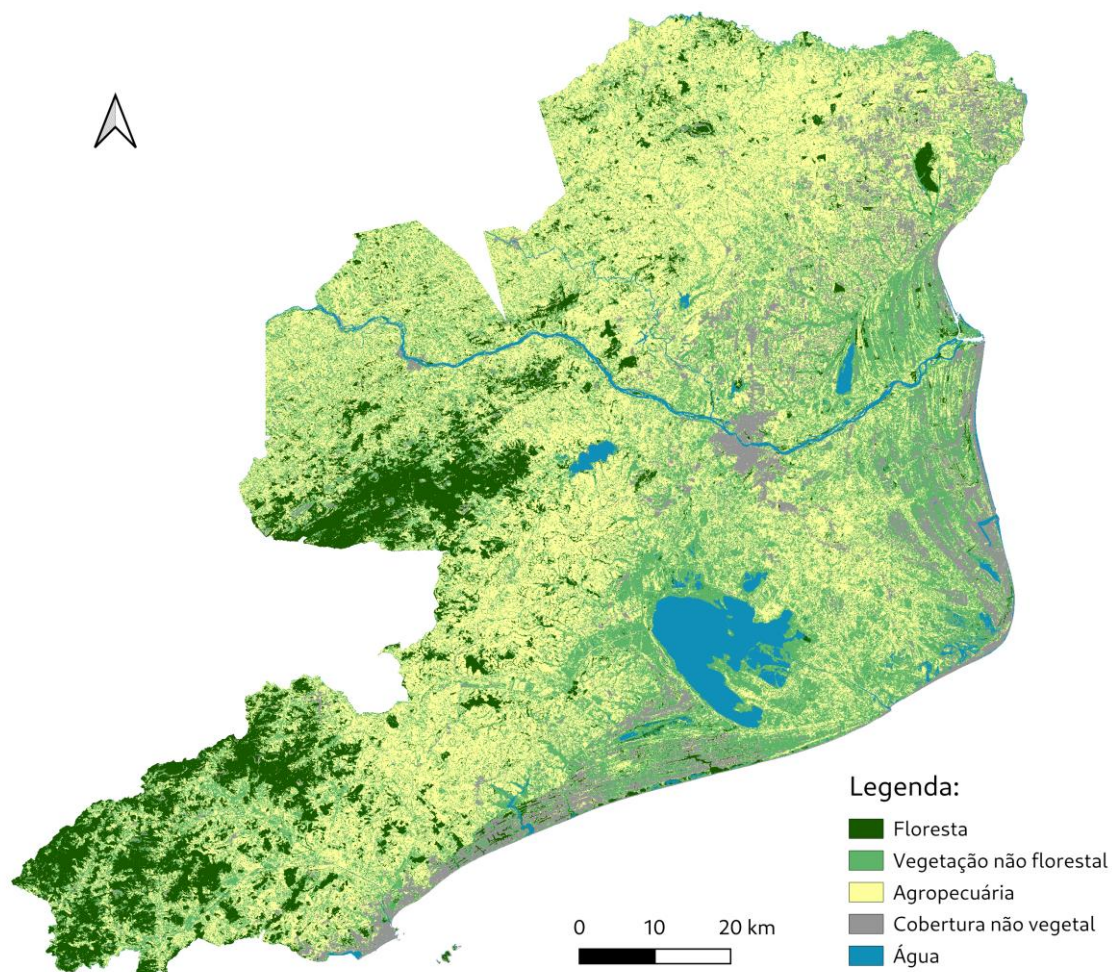
Tabela 3 - Coeficiente de similaridade Bray-Curtis para cada par de classes

Par de classes	Similaridade
Floresta / Vegetação não florestal	75,12%
Floresta / Agropecuária	72,28%
Floresta / Cobertura não vegetal	60,96%
Floresta / Água	35,70%
Vegetação não florestal / Agropecuária	94,99%
Vegetação não florestal / Cobertura não vegetal	84,66%
Vegetação não florestal / Água	39,77%
Agropecuária / Cobertura não vegetal	86,46%
Agropecuária / Água	36,67%
Cobertura não vegetal / Água	31,58%

Fonte: Elaboração própria com base no Sentinel-2 (2022).

Após a análise da região de estudo, foram aplicadas as metodologias de classificação. O primeiro classificador utilizado foi o da máxima verossimilhança, que não possui parâmetros de ajuste. O método se demonstrou de fácil utilização e de baixas exigências de recursos computacionais. A imagem classificada se encontra na Figura 9.

Figura 9 - Imagem classificada pelo método da máxima verossimilhança (MLE)

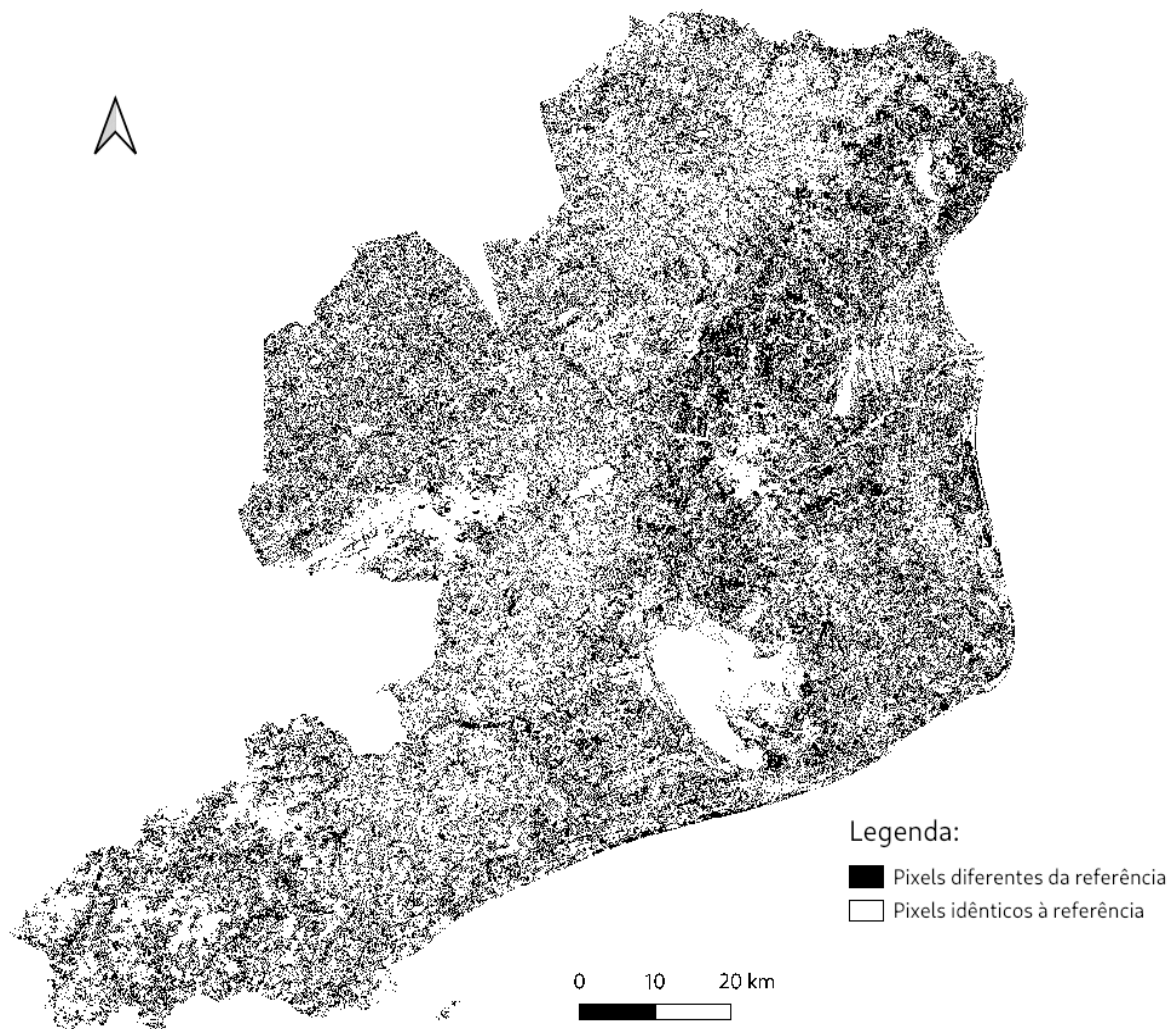


Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomas (2022).

É possível identificar visualmente que houve uma baixa diferenciação entre as classes de vegetação não florestal e de uso agropecuário, com uma mistura de *pixels* das duas categorias nas regiões centrais da imagem. Já para as demais classes, observa-se um bom desempenho, especialmente nas regiões florestais.

Para destacar os erros, foi realizada a operação de diferença entre a imagem gerada e a imagem de referência. Na Figura 10, são exibidos os *pixels* classificados em classes distintas das da imagem de referência

Figura 10 - Pixels classificados erroneamente pelo método da máxima verossimilhança (MLE)



Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomias (2022).

Observa-se uma distribuição homogeneizada de *pixels*, com uma região um pouco mais densa ao centro da imagem, ao redor de uma região de densidade urbana. A contagem total dos *pixels* classificados incorretamente soma 33,41% do total dos *pixels* da área de estudo. O desempenho individual das classes pode ser observado na matriz de confusão apresentada na Tabela 4.

Tabela 4 - Matriz de confusão para o método da máxima verossimilhança (MLE)

Imagem classificada	Imagem de referência						UA	CE
	Floresta	Vegetação não florestal	Agropecuária	Cobertura não vegetal	Água	Total		
Floresta	1300492	19252	90280	492	15679	1426195	0,91	0,09
Vegetação não florestal	132899	1081642	2502118	57694	49761	3824114	0,28	0,72
Agropecuária	193015	290464	4688740	21062	2910	5196191	0,9	0,1
Cobertura não vegetal	8301	147101	312123	372022	9479	849026	0,44	0,56
Água	190	5395	4771	7230	372803	390389	0,95	0,05
Total	1634897	1543854	7598032	458500	450632	11685915		
PA	0,8	0,7	0,62	0,81	0,83			
OE	0,2	0,3	0,38	0,19	0,17			

UA: User's Accuracy (Acurácia do usuário) CE: Commission Error (Erro de comissão) PA: Producer's Accuracy (Acurácia do produtor) OE: Omission Error (Erro de omissão)

Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomas (2022).

O resultado apresentou alta taxa de verdadeiros positivos na identificação de floresta (80%), cobertura não vegetal (81%) e água (83%), mas destas, apenas as classes de água e floresta apresentaram também elevada precisão (95 e 91%, respectivamente). Desse modo, estas duas classes apresentaram o melhor desempenho na representação da superfície da região. A classe de agropecuária apresentou taxa de verdadeiros positivos moderada (62%) e precisão alta (90%), indicando que uma parcela expressiva de *pixels* dessa categoria não foi identificada, mas que os *pixels* apresentados na imagem resultante estão, em sua quase totalidade, correspondentes com a referência. As classes de cobertura não vegetal e vegetação não florestal apresentaram altas taxas de verdadeiros positivos, mas baixa precisão, o que significa que *pixels* de outras classes foram incluídos, gerando uma imagem imprecisa. Para uma avaliação consistente do desempenho estatístico para cada classe, são apresentados, na Tabela 5, os coeficientes *kappa* calculados.

Tabela 5 - Coeficiente *kappa* de cada classe para o método da máxima verossimilhança (MLE)

Classe	Coeficiente de concordância kappa
Floresta	0,90
Vegetação não florestal	0,17
Agropecuária	0,72
Cobertura não vegetal	0,42
Água	0,95

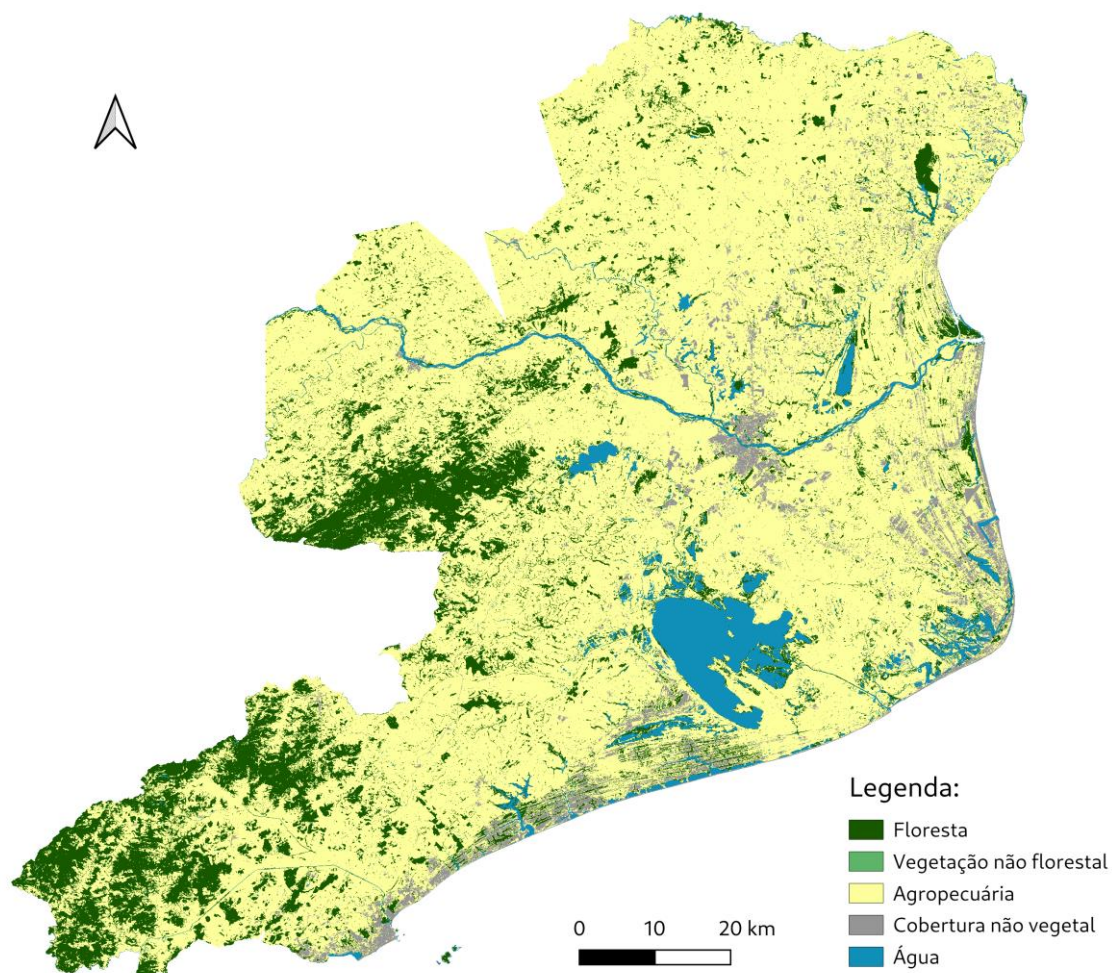
Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomas (2022).

Os coeficientes apresentados permitem auxiliar na sumarização dos resultados observados. As classes de floresta e água apresentaram os melhores desempenhos e melhor representam as regiões na imagem de referência, com coeficientes *kappa* de 0,90 e 0,95, respectivamente, enquanto as classes de vegetação não florestal e cobertura não vegetal, os piores desempenhos, com valores de 0,17 e 0,42. Já a classe de agropecuária, que apresentou alta precisão, mas não classificou uma quantidade elevada dos *pixels* esperados, apresentou o valor de 0,72. O coeficiente *kappa* próximo de 0 para a classe de vegetação não florestal confirma a percepção visual de que a classificação se deu de forma aleatória para esta classe.

O resultado se aproximou do esperado, uma vez que as classes de água e floresta, que possuem as maiores diferenciações espectrais, apresentaram melhores coeficientes. Porém, observa-se que o comportamento do método na situação de 2 classes com alta similaridade foi o de “embaralhar” os *pixels* resultantes de ambas. Isto acarretou uma imagem resultante desequilibrada visualmente, quando comparada com a referência, já que a confusão entre as classes similares foi distribuída pelo mapa.

Foi também estudado o método Support Vector Machine, ou SVM. Executado a partir da plataforma de computação em nuvem da Earth Engine, que entregou a imagem classificada apresentada na Figura 11.

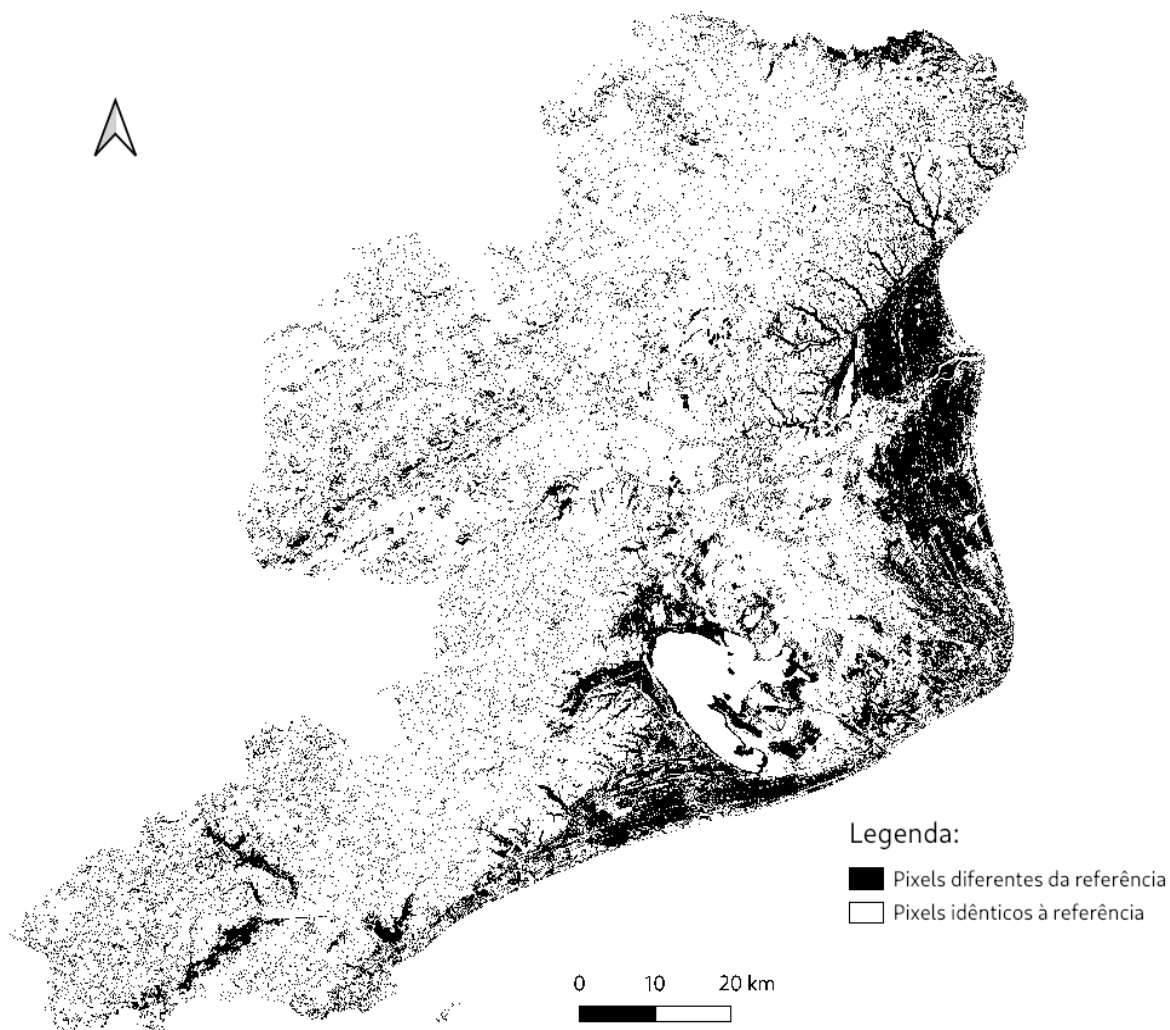
Figura 11 - Imagem classificada pelo método máquina de vetores de suporte (SVM)



Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomass (2022).

Nesta classificação, se observa um resultado com melhor distribuição geográfica dos *pixels* classificados, apresentando uma melhor coerência visual. Em contrapartida, já não se identificam *pixels* classificados como vegetação não florestal, demonstrando que provavelmente, o método lida com a baixa diferenciação espectral priorizando uma das classes, neste caso, a de agropecuária. A distribuição geoespacial dos erros de classificação é apresentada na Figura 12.

Figura 12 - Pixels classificados erroneamente pelo método máquina de vetores de suporte (SVM)



Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomias (2022).

Diferentemente do método MLE, o SVM demonstrou uma concentração espacial dos *pixels* classificados erroneamente, se alongando pela região costeira, enquanto as regiões centrais da imagem apresentaram poucos erros. Esta característica permite que, a imagem resultante tenha uma maior coerência visual, mesmo na presença de erros de classificação, que podem ser mais facilmente isolados ou desconsiderados em uma análise. Este padrão espacial nos erros de classificação se deu devido à falha na identificação de *pixels* na classe de vegetação não florestal, que possui um padrão de concentração na costa. O resultado da

classificação para cada classe pode ser observado na matriz de confusão, na Tabela 6.

Tabela 6 - Matriz de confusão para o método máquina de vetores de suporte (SVM)

Imagem classificado	Imagem de referência					Total	UA	CE
	Floresta	Vegetação não florestal	Agropecuária	Cobertura não vegetal	Água			
Floresta	1420217	223879	216469	1847	10354	1872766	0,76	0,24
Vegetação não florestal	33	1573	1312	377	393	3688	0,43	0,57
Agropecuária	210259	1176884	7244162	128525	5830	8765660	0,83	0,17
Cobertura não vegetal	374	40000	124505	326227	4628	495734	0,66	0,34
Água	4014	101518	11584	1524	429427	548067	0,78	0,22
Total	1634897	1543854	7598032	458500	450632	11685915		
PA	0,87	0,00	0,95	0,71	0,95			
OE	0,13	1,00	0,05	0,29	0,05			

UA: User's Accuracy (Acurácia do usuário) CE: Commission Error (Erro de comissão) PA: Producer's Accuracy (Acurácia do produtor) OE: Omission Error (Erro de omissão)

Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomias (2022).

O resultado apresentou alta taxa de verdadeiros positivos na identificação de floresta (87%), agropecuária (95%) e água (95%), assim como também apresentou moderada precisão para as mesmas (76%, 83% e 78%, respectivamente), gerando um resultado que bastante se aproxima da referência para essas classes, exceto pela de agropecuária, na qual ocorreu uma concentração dos falsos positivos envolvendo *pixels* da classe vegetação não florestal. Na Tabela 6, pode-se notar que a classe de vegetação não florestal apresentou uma taxa de verdadeiros positivos aproximadamente nula, ou seja, o modelo não conseguiu identificar *pixels* dessa categoria. Este foi, portanto, o comportamento observado na classificação de 2 classes com alta similaridade entre si. É importante ressaltar que, apesar do aparente bom resultado numérico na classificação da classe agropecuária, isto apenas ocorreu porque sua área total na imagem é substancialmente maior do que a área da sua classe espectralmente similar. Já a classe de cobertura não vegetal, esta apresentou um resultado moderado, tanto em acertos, quanto na precisão,

superando o encontrado pelo método MLE. Os coeficientes *kappa* calculados para cada classe são apresentados na Tabela 7.

Tabela 7 - Coeficiente *kappa* de cada classe para o método da máxima verossimilhança (SVM)

Classe	Coeficiente de concordância kappa
Floresta	0,72
Vegetação não florestal	0,34
Agropecuária	0,51
Cobertura não vegetal	0,64
Água	0,77

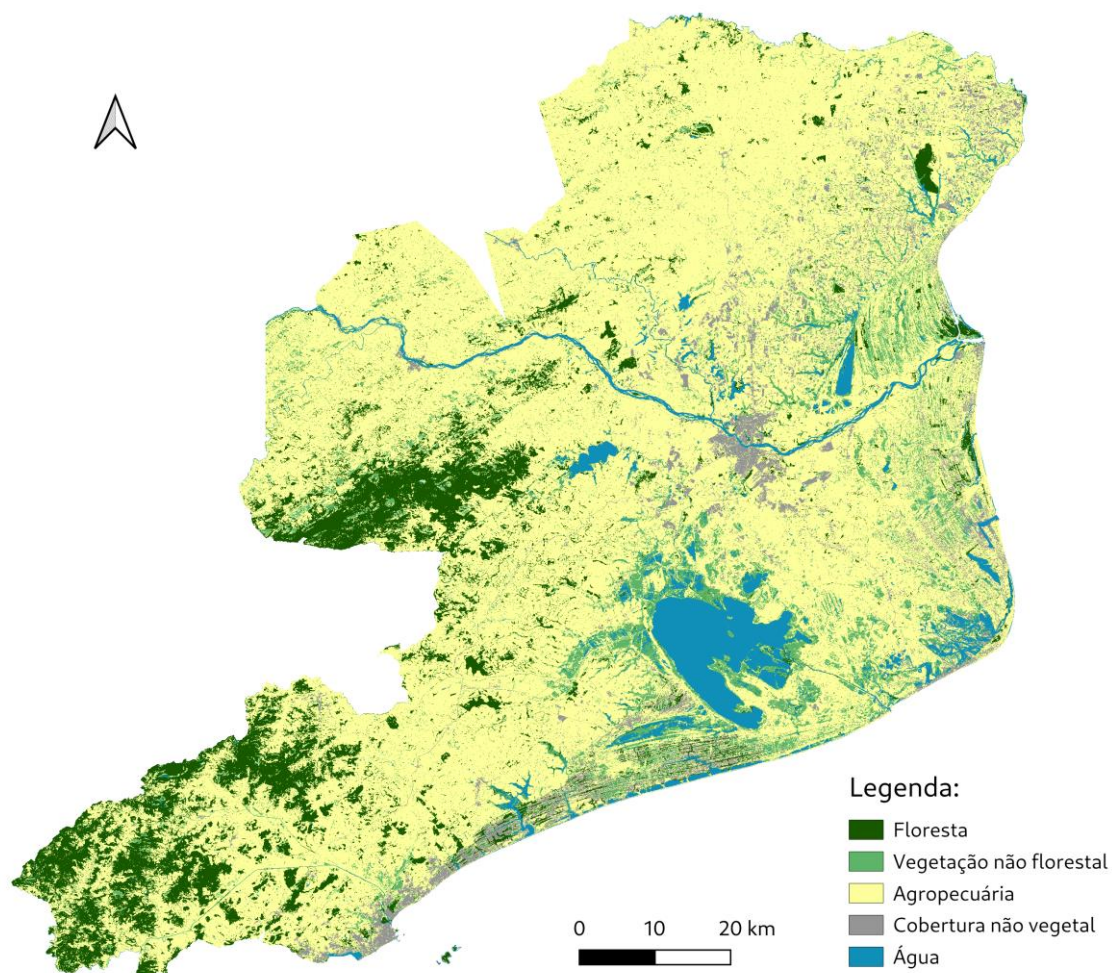
Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomias (2022).

Os coeficientes *kappa* evidenciam que as classes de floresta e água apresentaram os melhores desempenhos e o modelo não foi eficiente na classificação da categoria vegetação não florestal. A classe de agropecuária apresentou um valor correspondente a um desempenho marginalmente satisfatório. Contudo, é importante destacar que esta classe possui grande extensão territorial em comparação com a classe de vegetação não florestal, a qual houve maior confusão na classificação de *pixels*, o que pode ter mascarado parte do mau desempenho.

O resultado para as classes com maiores diferenças espectrais se aproximou do esperado, com melhores resultados para as classes de água e floresta, e resultado moderado para a classificação de cobertura não vegetal. O comportamento observado na situação de 2 classes com baixa diferenciação espectral foi de não realizar a classificação da classe de menor extensão territorial e identificar a maioria dos *pixels* de ambas classes na categoria da de maior extensão territorial.

Já o método Random Forest, foi também executado a partir de computação em nuvem disponibilizada pela plataforma Earth Engine foi testado com 3 diferentes valores de quantidade de árvores, e o resultado de maior precisão (500 árvores) é apresentado na Figura 13.

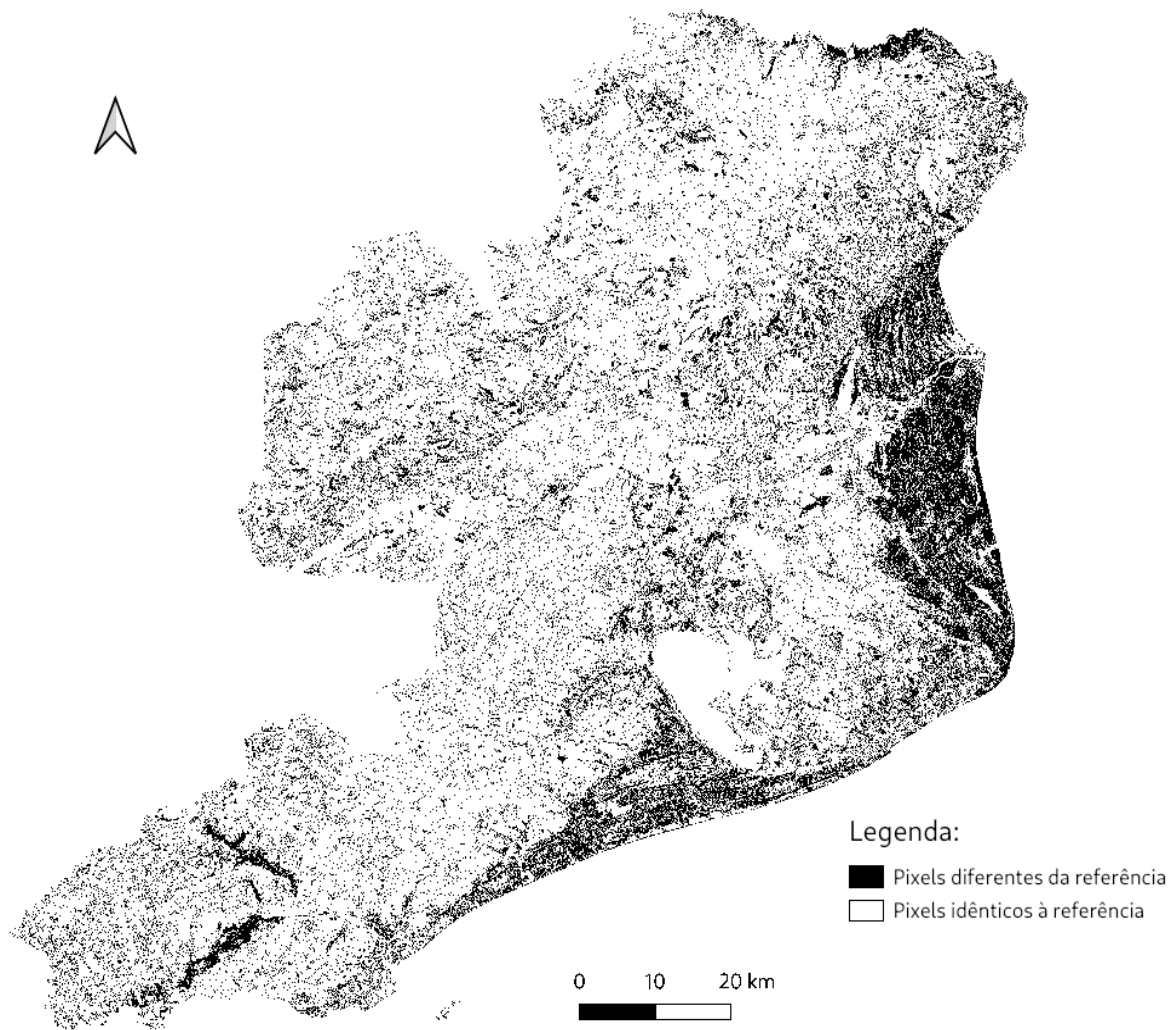
Figura 13 - Imagem classificada pelo método Random Forest



Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomass (2022).

Observa-se uma imagem que se assemelha à obtida pelo método SVM, mas com a presença de *pixels* classificados na classe de vegetação não florestal, demonstrando a capacidade do método de diferenciar classes mesmo com baixa diferença espectral entre si. A distribuição geoespacial dos erros é apresentada na Figura 14.

Figura 14 - Pixels classificados erroneamente pelo método Random Forest



Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomas (2022).

O perfil geoespacial dos erros de classificação utilizando o Random Forest apresentou, assim como no SVM, uma concentração nas regiões costeiras, mas com uma densidade mais baixa, assim como apresentou uma maior presença de erros nas regiões centrais, de forma distribuída, assemelhando-se, nesse aspecto, ao resultado obtido com o MLE. O desempenho individual das classes pode ser observado na matriz de confusão, apresentada na Tabela 8.

Tabela 8 - Matriz de confusão para o método Random Forest

Imagem classificado	Imagem de referência						UA	CE
	Floresta	Vegetação não florestal	Agropecuária	Cobertura não vegetal	Água	Total		
Floresta	1190352	47816	69635	595	1283	1309681	0,91	0,09
Vegetação não florestal	34568	599276	354605	13155	11691	1013295	0,59	0,41
Agropecuária	403634	725124	6945246	226552	5332	8305888	0,84	0,16
Cobertura não vegetal	1242	75112	217949	215716	3043	513062	0,42	0,58
Água	5101	96526	10597	2482	429283	543989	0,79	0,21
Total	1634897	1543854	7598032	458500	450632	11685915		
PA	0,73	0,39	0,91	0,47	0,95			
OE	0,27	0,61	0,09	0,53	0,05			

UA: User's Accuracy (Acurácia do usuário) CE: Commission Error (Erro de comissão) PA: Producer's Accuracy (Acurácia do produtor) OE: Omission Error (Erro de omissão)

Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomias (2022).

O resultado também apresentou alta taxa de verdadeiros positivos na identificação de água (95%) e agropecuária (91%), com precisão também elevada (79 e 84%). Já a classe de floresta, não apresentou taxa de verdadeiros positivos tão elevada como nos métodos anteriores, com um total de 73%, mas obteve uma precisão de 91%. Na situação das 2 classes de baixa diferenciação espectral, o método obteve o resultado mais equilibrado dentre os analisados, identificando 39% dos verdadeiros positivos para a classe de vegetação não florestal, com uma precisão de 59%, enquanto obteve uma taxa de verdadeiros positivos de 91% para agropecuária, com precisão de 84%. Já a classe de cobertura não vegetal, apresentou um resultado inferior em comparação com os demais, com taxas de verdadeiros positivos e precisão de, respectivamente, 47% e 42%. Isto indica que, apesar de o método conseguir diferenciar classes com alta similaridade, pode ser perder desempenho quando há elevada dispersão em relação à média de valores espectrais para os *pixels* de uma categoria. Os coeficientes *kappa* calculados são apresentados na Tabela 9.

Tabela 9 - Coeficiente *kappa* de cada classe para o método Random Forest

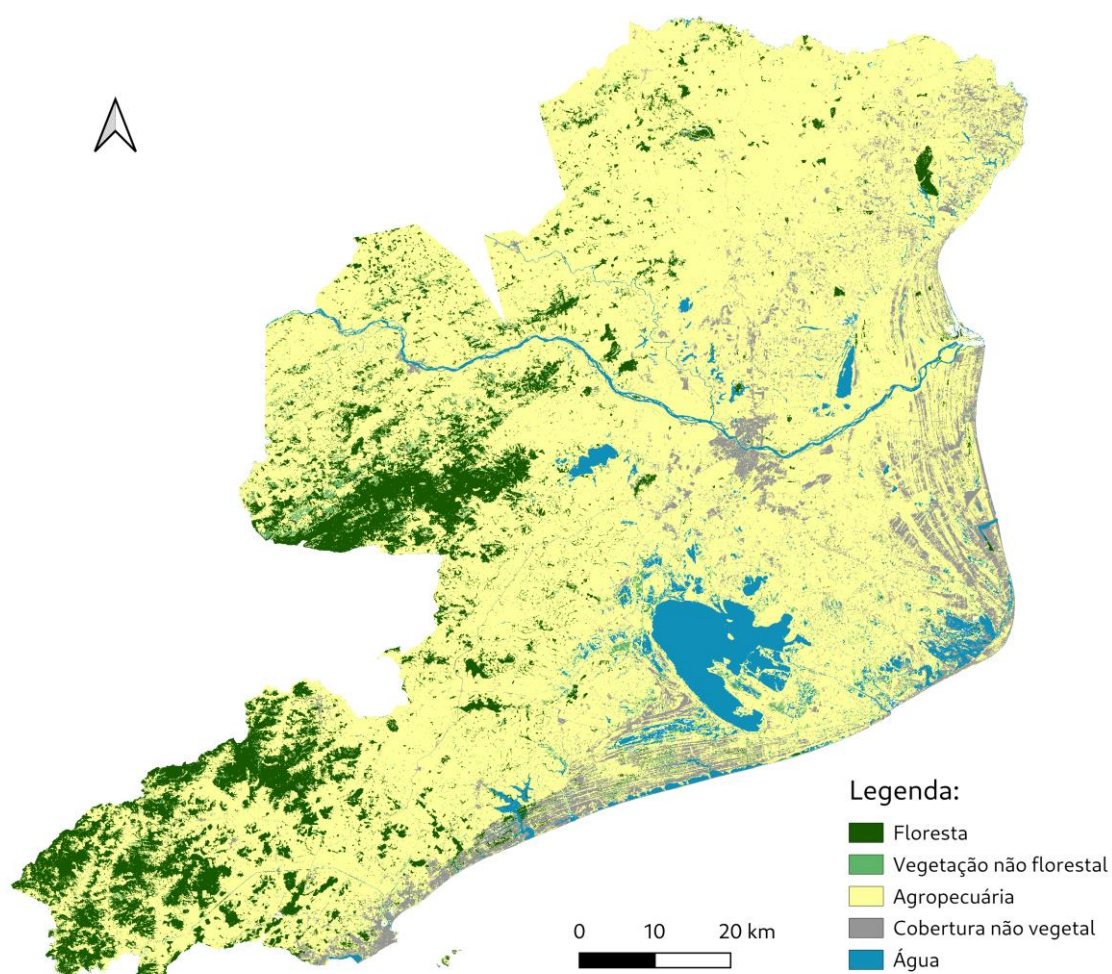
Classe	Coeficiente de concordância kappa
Floresta	0,89
Vegetação não florestal	0,53
Agropecuária	0,53
Cobertura não vegetal	0,40
Água	0,78

Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomias (2022).

Os coeficientes *kappa* destacaram os resultados mais elevados na classificação de floresta e água, e resultados menos elevados na classificação de vegetação não florestal, agropecuária e cobertura não vegetal. A principal característica do método foi apresentar o melhor resultado na diferenciação das 2 classes com alta similaridade espectral, apresentando o coeficiente *kappa* mais elevado, dentre os classificadores estudados, para vegetação não florestal. Entretanto, houve um desempenho reduzido na diferenciação das regiões de cobertura não vegetal, em que grande parcela de *pixels* desta classe foram classificados como agropecuária, conforme Tabela 8. Isto pode ter contribuído para um coeficiente reduzido para a classe de agropecuária, que apresentou o menor valor dentre os métodos testados.

O método RNA, executado localmente, se demonstrou intensivo computacionalmente, necessitando de pouco mais de uma hora de execução para apresentar os resultados da Figura 15.

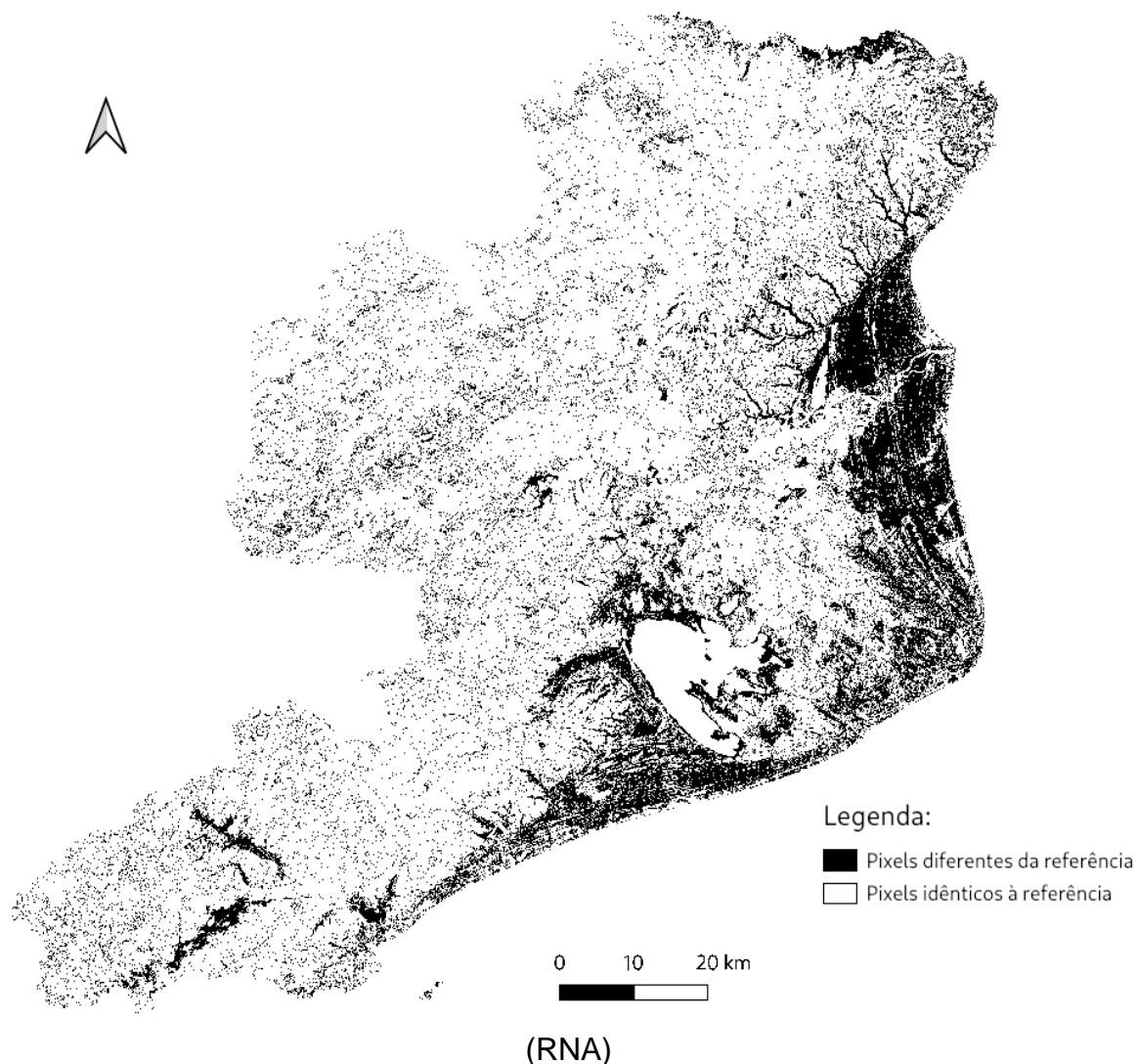
Figura 15 - Imagem classificada pelo método redes neurais artificiais (RNA)



Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomias (2022).

A imagem resultante se assemelha à obtida por meio do método SVM, no sentido que não se identificam *pixels* classificados como vegetação não florestal, além de também se observar uma menor quantidade de *pixels* classificados como cobertura não vegetal. Por outro lado, se identifica uma boa qualidade da identificação de regiões de água, com a identificação correta de áreas alagadas no sudeste da região, assim como de canais de água estreitos não identificados pelos demais métodos. A distribuição geoespacial dos erros é apresentada na Figura 16.

Figura 16 - Pixels classificados erroneamente pelo método redes neurais artificiais



Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomias (2022).

Os erros da classificação se concentraram densamente nas proximidades da costa, principalmente nas áreas de vegetação próximas a corpos de água, além de uma densidade ligeiramente mais elevada nas regiões de floresta, em comparação com os demais métodos. É importante observar que, apesar da menor eficiência na identificação de cobertura não vegetal, as regiões de aglomerados urbanos foram, em geral, classificadas corretamente, com baixa densidade de erros. O desempenho individual das classes pode ser observado na matriz de confusão apresentada na Tabela 10.

Tabela 10 - Matriz de confusão para o método redes neurais artificiais (RNA)

Imagem classificado	Imagem de referência					Total	UA	CE
	Floresta	Vegetação não florestal	Agropecuária	Cobertura não vegetal	Água			
Floresta	1246959	17507	78900	582	2345	1346293	0,93	0,07
Vegetação não florestal	7740	83616	43895	1403	62	136716	0,61	0,39
Agropecuária	372933	1145794	7227484	90185	25429	8861825	0,82	0,18
Cobertura não vegetal	5651	147120	232634	364299	3381	753085	0,48	0,52
Água	1614	149817	15119	2031	419415	587996	0,71	0,29
Total	1634897	1543854	7598032	458500	450632	11685915		
PA	0,76	0,05	0,95	0,79	0,93			
OE	0,24	0,95	0,05	0,21	0,07			

UA: User's Accuracy (Acurácia do usuário) CE: Commission Error (Erro de comissão) PA: Producer's Accuracy (Acurácia do produtor) OE: Omission Error (Erro de omissão)

Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomas (2022).

O resultado apresentou resultados mais elevados na identificação de água e floresta. O desempenho do método para a classificação das 2 classes espectralmente similares foi semelhante ao obtido pelo método SVM, pois a taxa de verdadeiros positivos para vegetação não florestal foi aproximadamente nula, enquanto para agropecuária foi de 95%. O método obteve ainda uma taxa de verdadeiros positivos moderada para cobertura não vegetal (79%), mas com precisão reduzida (48%). Os coeficientes *kappa* calculados são apresentados na Tabela 11.

Tabela 11 - Coeficiente *kappa* de cada classe para o método da máxima verossimilhança (RNA)

Classe	Coeficiente de concordância kappa
Floresta	0,91
Vegetação não florestal	0,55
Agropecuária	0,47
Cobertura não vegetal	0,46
Água	0,7

Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomias (2022).

Apesar da quantidade insatisfatória de *pixels* corretamente identificados como vegetação não florestal, o coeficiente *kappa* desta classe se apresentou elevado em comparação com os demais métodos. Já para a classe de agropecuária, o coeficiente foi de apenas 0,47, semelhante ao valor para a classe de cobertura não vegetal, de 0,46. Os valores mais elevados foram para as classes de floresta e água.

Para fins de comparação entre os quatro métodos estudados, as taxas de acerto foram agregadas lado a lado na Tabela 12.

Tabela 12 - Coeficiente de concordância kappa e acurácias do usuário e produtor para os quatro métodos

Classe	Método											
	MLE			SVM			Random Forest			RNA		
	PA	UA	Kappa	PA	UA	Kappa	PA	UA	Kappa	PA	UA	Kappa
Floresta	0,80	0,91	0,90	0,87	0,76	0,72	0,73	0,91	0,89	0,76	0,93	0,91
Vegetação não florestal	0,70	0,28	0,17	0,00	0,43	0,34	0,39	0,59	0,53	0,05	0,61	0,55
Agropecuária	0,62	0,90	0,72	0,95	0,83	0,51	0,91	0,84	0,53	0,95	0,82	0,47
Cobertura não vegetal	0,81	0,44	0,42	0,71	0,66	0,64	0,47	0,42	0,40	0,79	0,48	0,46
Água	0,83	0,95	0,95	0,95	0,78	0,77	0,95	0,79	0,78	0,93	0,71	0,70

MLE: Maximum Likelihood Estimation (estimação de máxima verossimilhança) SVM: Support Vector Machine (máquina de vetor de suporte) RNA: Redes Neurais Artificiais UA: User's Accuracy (Acurácia do usuário) PA: Producer's Accuracy (Acurácia do produtor)

Fonte: Elaboração própria com base no Sentinel-2 (2022) e MapBiomias (2022).

Ao comparar os resultados dos métodos utilizados, observa-se que, no geral, as classes com maiores diferenças espectrais foram classificadas com maiores acertos e precisão, apresentando resultados satisfatórios por todos os métodos estudados, enquanto as classes com menores diferenças, apresentaram resultados inferiores. O método MLE demonstrou-se mais dependente de ampla diferenciação espectral entre as classes estimadas, enquanto o Random Forest apresentou alguma eficácia na diferenciação de classes pouco distintas entre si, mas não teve bom resultado na classificação de cobertura não vegetal, cujas características espectrais apresentaram maiores dispersões da média. O método SVM, apesar de ineficaz na diferenciação de classes espectralmente semelhantes, apresentou o melhor resultado na identificação de cobertura não vegetal, cujas diferenças espectrais eram moderadas em relação às demais classes. O método RNA também se demonstrou dependente de maiores diferenças espectrais.

3.5 CONSIDERAÇÕES FINAIS

As diferentes técnicas de classificação de uso e cobertura do solo muito se diferenciam entre si na forma como atuam, tanto na condição de classificação de áreas com alta diferenciação espectral, quanto nas áreas de baixa diferenciação.

O método da máxima verossimilhança, um método estatístico, foi capaz de produzir o resultado com as maiores taxas de verdadeiros positivos, precisão e coeficiente *kappa* para as classes de floresta e água, mas apresentou dificuldade na identificação das classes com menor diferenciação espectral. No caso das 2 classes com alta similaridade, produziu um resultado com um espalhamento de *pixels* das 2 categorias, tornando a imagem pouco representativa da região geográfica.

Tratando-se de um método simples e sem a necessidade de ajuste de parâmetros, torna-se uma escolha adequada para trabalhos em regiões de classes bem distintas entre si. Outra importante característica do método é sua baixa demanda de *hardware*, podendo ser utilizado até mesmo em computadores de baixo custo ou mais antigos, sem a inviabilização ou dificuldade dos trabalhos.

O método SVM também demonstrou baixa eficiência na diferenciação das classes de agropecuária e de vegetação não florestal, que possuem baixas variações espectrais entre si, mas a forma como se portou diante do problema foi o oposto do método da máxima verossimilhança, classificando a grande maioria dos

pixels como área de agropecuária, gerando poucos *pixels* de vegetação não florestal. Apesar da ineficiência na diferenciação entre ambas classes, a classe de agropecuária apresentou um valor *kappa* elevado, demonstrando uma limitação deste coeficiente. Contudo, a imagem gerada possui mais homogeneidade nos *pixels* classificados, de forma que mais se assemelha à imagem real. Um fator de destaque para o método SVM, foi sua capacidade de classificar as áreas de cobertura não vegetal, cujas diferenças espectrais eram moderadas em relação às demais, apresentando o melhor resultado para esta classe, dentre os métodos estudados.

Um resultado semelhante foi obtido pelo método RNA, que classificou excessivamente os *pixels* como agropecuária, com poucos *pixels* classificados como vegetação não florestal. Entretanto, o método foi bastante eficaz na classificação de floresta e água.

Por fim, o método Random Forest produziu o resultado visualmente mais equilibrado e representativo da imagem de referência. Não foi o mais eficiente estatisticamente para nenhuma das classes objetivadas, mas também não produziu grandes percentuais de erros, comparativamente. O principal destaque deste método foi sua capacidade de lidar com a situação de duas classes com baixa diferenciação espectral entre si, conseguindo ser eficiente na classificação de ambas, tanto estatisticamente, quando espacialmente. Contudo, foi o método menos eficiente na classificação de cobertura não florestal, cuja diferenciação espectral era moderada em relação às demais e com alta dispersão em relação à média do valor dos *pixels* em cada banda da imagem.

Diante do exposto, o presente trabalho reforça a necessidade de se conhecer, tanto as características da região de estudo, quanto as características do classificador a ser empregado. Considerando o bom desempenho obtido pelo método Random Forest ao lidar com classes de alta similaridade entre si, se comparado com os demais estudados, este pode apresentar maior potencial para ser indicado como um método para uso geral, atendendo a diferentes tipos de pesquisa e fornecendo taxas de acerto e erro aceitáveis para a maioria dos casos. Contudo, a aplicabilidade dos métodos abordados ainda carece de mais estudos para uma melhor compreensão dos melhores campos de aplicação para cada.

Destaca-se ainda que nenhum dos métodos estudados se demonstrou livre de grandes percentuais de erros, de forma que o papel dos especialistas permanece

importante na revisão dos resultados, sendo a função das ferramentas computacionais, facilitar o trabalho e permitir a classificação de áreas mais extensas, ou séries temporais, o que seria inviável para análise manual.

REFERÊNCIAS

BARSI, Á. *et al.* Accuracy dimensions in remote sensing. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, [S. l.], v. 42, p. 61–67, 2018. Disponível em: <https://isprs-archives.copernicus.org/articles/XLII-3/61/2018/isprs-archives-XLII-3-61-2018.html>. Acesso em: 22 jan. 2023.

BAYARSAIKHAN, Uudus; BOLDGIV, Bazartseren; KIM, Kyung-Ryul; PARK, Kyung-Ae; LEE, Donkoo. Change detection and classification of land cover at Hustai National Park in Mongolia. **International Journal of Applied Earth Observation and Geoinformation**, [S. l.], v. 11, n. 4, p. 273–280, 2009. Disponível em: <http://doi.org/10.1016/j.jag.2009.03.004>. Acesso em: 22 jan. 2023.

BELWARD, Alan S.; SKØIEN, Jon O. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. **ISPRS Journal of Photogrammetry and Remote Sensing**, Global Land Cover Mapping and Monitoring. [S. l.], v. 103, p. 115–128, 2015. Disponível em: <http://doi.org/10.1016/j.isprsjprs.2014.03.009>. Acesso em: 20 jan. 2023.

BREIMAN, Leo. Random Forests. **Machine Learning**, [S. l.], v. 45, n. 1, p. 5–32, 2001. Disponível em: <http://doi.org/10.1023/A:1010933404324>. Acesso em: 22 jan. 2023.

CONGEDO, Luca. **Semi-Automatic Classification Plugin**. Zenodo, 2021. Disponível em: <https://zenodo.org/record/5259995>. Acesso em: 22 jan. 2023.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine Learning**, [S. l.], v. 20, n. 3, p. 273–297, 1995. Disponível em: <http://doi.org/10.1007/BF00994018>. Acesso em: 22 jan. 2023.

CRABBÉ, Ann. **Neural Network MLPClassifier — QGIS Python Plugins Repository**. 2020. Disponível em: <https://plugins.qgis.org/plugins/Inns/>. Acesso em: 22 fev. 2023. Acesso em: 22 jan. 2023.

CUTLER, Adele; CUTLER, David; STEVENS, John. Random Forests. *In*: ZHANG, C.; MA, Y. (ed.). **Ensemble Machine Learning - ML**. [s.l.]: Springer, 2011. v. 45p. 157–176. Disponível em: http://doi.org/10.1007/978-1-4419-9326-7_5. Acesso em: 22 jan. 2023.

DAVE, Chintan P.; JOSHI, Rahul; SRIVASTAVA, S. S. A Survey on Geometric Correction of Satellite Imagery. **International Journal of Computer Applications**,

[S. l.], v. 116, n. 12, p. 24–27, 2015. Disponível em: <http://doi.org/10.5120/20389-2655>. Acesso em: 22 jan. 2023.

DHINGRA, Sakshi; KUMAR, Dharminder. A review of remotely sensed satellite image classification. **International Journal of Electrical and Computer Engineering**, [S. l.], v. 9, n. 3, p. 1720, 2019.

DONCHYTS, Gennadii; LLANO, Xavier C.; BAART, Fedor; BRAATEN, Justin; HEREÑÚ, Darío. **Google Earth Engine plugin for QGIS**. 2023. Disponível em: <https://github.com/gee-community/qgis-earthengine-plugin>. Acesso em: 19 fev. 2023.

DUHAMEL, Christophe. Land use, Land cover, including their classification. **Encyclopedia of life support system**, [S. l.], v. 1, p. 1–9, 2012.

EOSDIS. **EOSDIS Worldview**. 2023. Disponível em: <https://worldview.earthdata.nasa.gov/>. Acesso em: 19 fev. 2023.

EUROPEAN SPACE AGENCY. **WorldCover | WORLDCOVER**. 2023. Disponível em: <https://esa-worldcover.org/en>. Acesso em: 10 jan. 2022.

GOOGLE. **Preços do Earth Engine Google Cloud**. 2023a. Disponível em: <https://cloud.google.com/earth-engine/pricing?hl=pt-br>. Acesso em: 17 fev. 2023.

GOOGLE. **Google Earth Engine**. 2023b. Disponível em: <https://earthengine.google.com>. Acesso em: 19 fev. 2023.

GORELICK, Noel; HANCHER, Matt; DIXON, Mike; ILYUSHCHENKO, Simon; THAU, David; MOORE, Rebecca. Google Earth Engine: planetary-scale geospatial analysis for everyone. **Remote Sensing Of Environment**, [S. l.], v. 202, p. 18-27, dez. 2017. Elsevier BV. Disponível em: <http://dx.doi.org/10.1016/j.rse.2017.06.031>. Acesso em: 22 mar. 2023.

GUPTA, Neha. Artificial neural network. **Network and Complex Systems**, [S. l.], v. 3, n. 1, p. 24–28, 2013.

HORNING, Ned. **Land cover classification methods, Version 1.0**. American Museum of Natural History, Center for Biodiversity and Conservation, 2004. Disponível em: https://www.amnh.org/content/download/74344/1391366/file/LandCoverClassification_Final.pdf. Acesso em: 10 nov. 2022.

HUANG, C.; DAVIS, L. S.; TOWNSHEND, J. R. G. An assessment of support vector machines for land cover classification. **International Journal of Remote Sensing**, [S. l.], v. 23, n. 4, p. 725–749, 2002. Disponível em: <http://doi.org/10.1080/01431160110040323>. Acesso em: 22 fev. 2023.

IBGE. **Monitoramento da cobertura e uso da terra do Brasil: 2016 - 2018**. Rio de Janeiro: IBGE, 2020.

IBGE. **Malha Municipal IBGE**. 2021. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=acesso-ao-produto>. Acesso em: 19 fev. 2023.

INPE. **Catálogo de Imagens**. 2023. Disponível em: <http://www.dgi.inpe.br/CDSR/>. Acesso em: 19 fev. 2023.

JANSSEN, L. L. F.; VANDERWEL, F. J. M. Accuracy assessment of satellite derived land-cover data: A review. **Photogrammetric Engineering and Remote Sensing**, United States, v. 60, n. 4, 1994. Disponível em: <https://www.osti.gov/biblio/6448244>. Acesso em: 16 fev. 2023.

MAPBIOMAS. **Mapbiomas Brasil - O projeto**. 2019. Disponível em: <https://mapbiomas.org/o-projeto>. Acesso em: 18 fev. 2023.

MCNEILL, Graham; HALE, Scott A. Generating Tile Maps. **Computer Graphics Forum**, [S. l.], v. 36, n. 3, p. 435–445, 2017. Disponível em: <http://doi.org/10.1111/cgf.13200>. Acesso em: 22 abr. 2023.

MEYER, William B.; TURNER, B. L. Human population growth and global land-use/cover change. **Annual Review of Ecology, Evolution, and Systematics**, [S. l.], v. 23, n. 1, p. 39–61, 1992. Disponível em: <http://doi.org/10.1146/annurev.es.23.110192.000351>. Acesso em: 22 abr. 2023.

MEYER, William B.; TURNER, B. L.; UNIVERSITY CORPORATION FOR ATMOSPHERIC RESEARCH (org.). **Changes in land use and land cover: a global perspective**. Cambridge, England: Cambridge University Press, 1994.

NABOUREH, Amin; BIAN, Jinhui; LEI, Guangbin; LI, Ainong. A review of land use/land cover change mapping in the China-Central Asia-West Asia economic corridor countries. **Big Earth Data**, [S. l.], v. 5, p. 237–257, 2021. Disponível em: <http://doi.org/10.1080/20964471.2020.1842305>. Acesso em: 22 abr. 2023.

NOBLE, William S. What is a support vector machine? **Nature Biotechnology**, [S. l.], v. 24, n. 12, p. 1565–1567, 2006. Disponível em: <http://doi.org/10.1038/nbt1206-1565>. Acesso em: 22 abr. 2023.

OGC **GeoTIFF Standard OGC**. 2023. Disponível em: <https://www.ogc.org/standards/geotiff>. Acesso em: 7 mar. 2022.

OTUKEI, J. R.; BLASCHKE, T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. **International Journal of Applied Earth Observation and Geoinformation**, [S. l.], v. 12, p. S27–S31, 2010. Disponível em: <http://doi.org/10.1016/j.jag.2009.11.002>. Acesso em: 22 fev. 2023.

QGIS. — **QGIS Python Plugins Repository**. 2023a. Disponível em: <https://plugins.qgis.org/plugins/>. Acesso em: 17 fev. 2023.

QGIS. **Descubra o QGIS**. 2023b. Disponível em: https://qgis.org/pt_BR/site/about/index.html. Acesso em: 17 fev. 2023.

QGIS DEVELOPMENT TEAM. **QGIS Geographic Information System**. [S. l.]: QGIS Association, 2023. Disponível em: <https://www.qgis.org>. Acesso em: 16 fev. 2023.

RADOČAJ, Dorijan; OBHOĐAŠ, Jasmina; JURISIĆ, Mladen; GAŠPAROVIĆ, Mateo. Global Open Data Remote Sensing Satellite Missions for Land Monitoring and Conservation: A Review. **Land**, [S. l.], v. 9, n. 11, p. 402, 2020. Disponível em: <http://doi.org/10.3390/land9110402>. Acesso em: 22 abr. 2023.

RWANGA, Sophia S.; NDAMBUKI, J. M. Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. **International Journal of Geosciences**, [S. l.], v. 8, n. 4, p. 611–622, 2017. DOI: 10.4236/ijg.2017.84033.

SHRAWANKAR, Urmila; B. KHANDARE, Shrinivas. Preprocessing Algorithm for High Resolution Satellite Imagery. *In*: INTERNATIONAL CONFERENCE ON RESEARCH IN INTELLIGENT COMPUTING IN ENGINEERING, 2016, **Anais** [...]. Disponível em: https://www.academia.edu/download/46728238/RICE-2016_ShrinivasKhandare.pdf. Acesso em: 26 abr. 2023.

USGS. **National Land Cover Database - U.S. Geological Survey**. 2018. Disponível em: <https://www.usgs.gov/centers/eros/science/national-land-cover-database>. Acesso em: 16 fev. 2023.

USGS. **EarthExplorer**. 2022. Disponível em: <https://earthexplorer.usgs.gov/>. Acesso em: 19 fev. 2023.

WULDER, Michael A.; COOPS, Nicholas C.; ROY, David P.; WHITE, Joanne C.; HERMOSILLA, Txomin. Land cover 2.0. **International Journal of Remote Sensing**, [S. l.], v. 39, n. 12, p. 4254–4284, 2018. Disponível em: <http://doi.org/10.1080/01431161.2018.1452075>. Acesso em: 22 abr. 2023.

YOSHIDA, T.; OMATU, S. Neural network approach to land cover mapping. **IEEE Transactions on Geoscience and Remote Sensing**, [S. l.], v. 32, n. 5, p. 1103–1109, 1994. Disponível em: <http://doi.org/10.1109/36.312899>. Acesso em: 22 mar. 2023.

YOUNG, Nicholas E.; ANDERSON, Ryan S.; CHIGNELL, Stephen M.; VORSTER, Anthony G.; LAWRENCE, Rick; EVANGELISTA, Paul H. A survival guide to Landsat preprocessing. **Ecology**, [S. l.], v. 98, n. 4, p. 920–932, 2017. Disponível em: <http://doi.org/10.1002/ecy.1730>. Acesso em: 22 abr. 2023.

ZHANG, J.; WANG, B.; CHEN, X.; WU, X.; ZHANG, D. Research trends on land use changes during 1991–2015: A Bibliometric Analysis. **Lowland Technology International**, [S. l.], v. 21, n. 1, June, p. 61–70, 2019. Acesso em: 22 fev. 2023.

4 CONSIDERAÇÕES FINAIS GERAIS

Neste trabalho, foram realizados dois estudos: um na base de dados científica Scopus, com o objetivo de identificar o grau de utilização de diferentes metodologias para uso e cobertura do solo, e o outro, um estudo comparativo de quatro dos métodos mais relevantes da atualidade. A análise de artigos científicos sobre classificação de uso e cobertura do solo foi realizada com trabalhos publicados na última década, compreendendo os anos de 2012 a 2022. Houve um crescimento no número de pesquisas na área, sendo mapeado o uso de diferentes metodologias nesse período.

O número de publicações para o tema superou o crescimento no número de publicações para todos os temas na base de dados, o que pode ser indicador de maior interesse no tema e facilitação na realização de estudos por meio de técnicas e equipamentos modernos, reduzindo barreiras de entrada. As técnicas baseadas em aprendizado de máquina se tornaram as mais populares na atualidade, presentes em mais de 35% dos estudos. O aumento na quantidade de estudos na área de classificação de uso e cobertura do solo como um todo, se deu a partir do crescimento no uso dessas técnicas, sem que houvesse uma redução ou desinteresse no uso das demais metodologias.

Outra observação relevante foi a identificação de um aumento aproximado de 10 vezes na categoria classificada como “outros”, que consistiu de artigos utilizando técnicas sem representatividade suficiente para a criação de sua própria categoria e discussão. Isso pode estar relacionado a uma exploração de novas metodologias na área.

Já na avaliação de métodos, observou-se que as diferentes técnicas de classificação de uso e cobertura do solo diferenciam-se entre si na forma como atuam, principalmente em áreas com baixa diferenciação espectral. O método da máxima verossimilhança, um método estatístico, foi capaz de produzir o resultado com as maiores taxas de verdadeiros positivos, precisão e coeficiente *kappa* para classes com elevadas diferenciações espectrais em relação às demais classificadas, mas apresentou dificuldade na identificação de classes com baixa diferenciação espectral, produzindo um resultado com um espalhamento de *pixels* entre categorias, tornando a imagem menos representativa da região geográfica nessas condições. Entretanto, considerando o bom desempenho em situações de alta

diferenciação, torna-se um método viável e de baixos requisitos computacionais para diferentes tipos de estudo.

Os métodos SVM e de Redes Neurais Artificiais também demonstraram baixa eficiência na diferenciação das classes de agropecuária e de vegetação não florestal, que possuem poucas variações espectrais entre si. Contudo, estes classificaram maioria dos *pixels* como área de agropecuária. Para as classes com menores similaridades, também apresentaram resultados satisfatórios, com destaque para a capacidade do método SVM de gerar o melhor resultado para a cobertura não vegetal, cuja diferenciação espectral era moderada em comparação com as demais e apresentou elevada dispersão de valores em relação à média.

Em contrapartida, o método Random Forest produziu o resultado visualmente mais equilibrado e representativo da imagem de referência. Embora não tenha sido estatisticamente o mais eficiente para nenhuma das classes em particular, esse método também não apresentou grandes erros em comparação aos outros métodos. O principal destaque desse método foi a sua capacidade de lidar com situações em que duas classes apresentam pouca diferenciação espectral entre si, conseguindo classificar ambas com relativa eficiência tanto espacialmente quanto estatisticamente. Entretanto, não apresentou bom desempenho para a classe de cobertura não vegetal.

Por fim, é importante destacar que nenhum dos métodos avaliados foi livre de percentuais de erros e que, a depender das condições espectrais da área de estudo, do conhecimento do pesquisador sobre esta, e dos recursos disponíveis, o método mais apropriado pode variar. Destaca-se ainda que o papel dos especialistas na revisão dos resultados é importante, para que se minimizem os erros e se obtenha um resultado em maior conformidade visual com o esperado.

REFERÊNCIAS

ALZUBI, Jafar; NAYYAR, Anand; KUMAR, Akshi. Machine Learning from Theory to Algorithms: An Overview. **Journal of Physics: Conference Series**, [S. l.], v. 1142, n. 1, p. 012012, 2018. Disponível em: <http://doi.org/10.1088/1742-6596/1142/1/012012>. Acesso em: 10 nov. 2022.

BARSI, Á. *et al.* Accuracy dimensions in remote sensing. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, [S. l.], v. 42, p. 61–67, 2018. Disponível em: <https://isprs-archives.copernicus.org/articles/XLII-3/61/2018/isprs-archives-XLII-3-61-2018.html>. Acesso em: 22 jan. 2023.

BAYARSAIKHAN, Uudus; BOLDGIV, Bazartseren; KIM, Kyung-Ryul; PARK, Kyung-Ae; LEE, Donkoo. Change detection and classification of land cover at Hustai National Park in Mongolia. **International Journal of Applied Earth Observation and Geoinformation**, [S. l.], v. 11, n. 4, p. 273–280, 2009. Disponível em: <http://doi.org/10.1016/j.jag.2009.03.004>. Acesso em: 22 jan. 2023.

BELGIU, Mariana; CSILLIK, Ovidiu. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. **Remote Sensing of Environment**, [S. l.], v. 204, p. 509–523, 2018. Disponível em: <http://doi.org/10.1016/j.rse.2017.10.005>. Acesso em: 12 nov. 2022.

BELWARD, Alan S.; SKØIEN, Jon O. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. **ISPRS Journal of Photogrammetry and Remote Sensing**, Global Land Cover Mapping and Monitoring. [S. l.], v. 103, Global Land Cover Mapping and Monitoring, p. 115–128, 2015. Disponível em: <http://doi.org/10.1016/j.isprsjprs.2014.03.009>. Acesso em: 20 jan. 2023.

BORRA, Surekha; THANKI, Rohit; DEY, Nilanjan. **Satellite Image Analysis: Clustering and Classification**. Singapura : Springer, 2019.

BREIMAN, Leo. Random Forests. **Machine Learning**, [S. l.], v. 45, n. 1, p. 5–32, 2001. Disponível em: <http://doi.org/10.1023/A:1010933404324>. Acesso em: 22 jan. 2023.

CHAVES, Michel E. D.; PICOLI, Michelle C. A.; SANCHES, Ieda D. Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review. **Remote Sensing**, [S. l.], v. 12, n. 18, p. 3062, 2020. Disponível em: <http://doi.org/10.3390/rs12183062>. Acesso em: 12 nov. 2022.

CONGEDO, Luca. **Semi-Automatic Classification Plugin**. Zenodo, 2021. Disponível em: <http://doi.org/10.5281/zenodo.5259995>. Disponível em: <https://zenodo.org/record/5259995>. Acesso em: 22 jan. 2023.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine Learning**, [S. l.], v. 20, n. 3, p. 273–297, 1995. Disponível em: <http://doi.org/10.1007/BF00994018>. Acesso em: 22 jan. 2023.

CRABBÉ, Ann. **Neural Network MLPClassifier — QGIS Python Plugins Repository**. 2020. Disponível em: <https://plugins.qgis.org/plugins/lnns/>. Acesso em: 22 fev. 2023. Acesso em: 22 jan. 2023.

CURTIS, Philip G.; SLAY, Christy M.; HARRIS, Nancy L.; TYUKAVINA, Alexandra; HANSEN, Matthew C. Classifying drivers of global forest loss. **Science**, [S. l.], v. 361, n. 6407, p. 1108–1111, 2018. Disponível em: <http://doi.org/10.1126/science.aau3445>. Acesso em: 10 nov. 2022.

CUTLER, Adele; CUTLER, David; STEVENS, John. Random Forests. In: ZHANG, C.; MA, Y. (ed.). **Ensemble Machine Learning - ML**. [s.l.]: Springer, 2011. v. 45p. 157–176. Disponível em: http://doi.org/10.1007/978-1-4419-9326-7_5. Acesso em: 22 jan. 2023.

DAVE, Chintan P.; JOSHI, Rahul; SRIVASTAVA, S. S. A Survey on Geometric Correction of Satellite Imagery. **International Journal of Computer Applications**, [S. l.], v. 116, n. 12, p. 24–27, 2015. Disponível em: <http://doi.org/10.5120/20389-2655>. Acesso em: 22 jan. 2023.

DHINGRA, Sakshi; KUMAR, Dharminder. A review of remotely sensed satellite image classification. **International Journal of Electrical and Computer Engineering**, [S. l.], v. 9, n. 3, p. 1720, 2019.

DONCHYTS, Gennadii; LLANO, Xavier C.; BAART, Fedor; BRAATEN, Justin; HEREÑÚ, Darío. **Google Earth Engine plugin for QGIS**. 2023. Disponível em: <https://github.com/gee-community/qgis-earthengine-plugin>. Acesso em: 19 fev. 2023.

DUHAMEL, Christophe. Land use, Land cover, including their classification. **Encyclopedia of life support system**, [S. l.], v. 1, p. 1–9, 2012.

EOSDIS. **EOSDIS Worldview**. 2023. Disponível em: <https://worldview.earthdata.nasa.gov/>. Acesso em: 19 fev. 2023.

EUROPEAN SPACE AGENCY. **WorldCover | WORLDCOVER**. 2023. Disponível em: <https://esa-worldcover.org/en>. Acesso em: 10 jan. 2022.

FRAZIER, Bruce E.; SHOVIK, HF. Statistical methods for determining land-use change with aerial photographs. **Photogrammetric Engineering and Remote Sensing**, [S. l.], v. 46, n. 8, p. 1067–1077, 1980.

GONG, Peng *et al.* Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. **Science Bulletin**, [S. l.], v. 64, n. 6, p. 370–373, 2019. Disponível em: <http://doi.org/10.1016/j.scib.2019.03.002>. Acesso em: 14 nov. 2022.

GOOGLE. **Preços do Earth Engine Google Cloud**. 2023a. Disponível em: <https://cloud.google.com/earth-engine/pricing?hl=pt-br>. Acesso em: 17 fev. 2023.

GOOGLE. **Google Earth Engine**. 2023b. Disponível em: <https://earthengine.google.com>. Acesso em: 19 fev. 2023.

GORELICK, Noel; HANCHER, Matt; DIXON, Mike; ILYUSHCHENKO, Simon; THAU, David; MOORE, Rebecca. Google Earth Engine: planetary-scale geospatial analysis for everyone. **Remote Sensing Of Environment**, [S.L.], v. 202, p. 18-27, dez. 2017. Elsevier BV. Disponível em: <http://dx.doi.org/10.1016/j.rse.2017.06.031>. Acesso em: 22 mar. 2023.

GUPTA, Neha. Artificial neural network. **Network and Complex Systems**, [S. l.], v. 3, n. 1, p. 24–28, 2013.

HORNING, Ned. **Land cover classification methods, Version 1.0**. American Museum of Natural History, Center for Biodiversity and Conservation, 2004. Disponível em: https://www.amnh.org/content/download/74344/1391366/file/LandCoverClassification_Final.pdf. Acesso em: 10 nov. 2022.

HUANG, C.; DAVIS, L. S.; TOWNSHEND, J. R. G. An assessment of support vector machines for land cover classification. **International Journal of Remote Sensing**, [S. l.], v. 23, n. 4, p. 725–749, 2002. Disponível em: <http://doi.org/10.1080/01431160110040323>. Acesso em: 22 fev. 2023.

IBGE. **Monitoramento da cobertura e uso da terra do Brasil: 2016 - 2018**. Rio de Janeiro: IBGE, 2020.

IBGE. **Malha Municipal IBGE**. 2021. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=acesso-ao-produto>. Acesso em: 19 fev. 2023.

INPE. **Catálogo de Imagens**. 2023. Disponível em: <http://www.dgi.inpe.br/CDSR/>. Acesso em: 19 fev. 2023.

JANSSEN, L. L. F.; VANDERWEL, F. J. M. Accuracy assessment of satellite derived land-cover data: A review. **Photogrammetric Engineering and Remote Sensing**, United States, v. 60, n. 4, 1994. Disponível em: <https://www.osti.gov/biblio/6448244>. Acesso em: 16 fev. 2023.

JOZDANI, Shahab Eddin; JOHNSON, Brian Alan; CHEN, Dongmei. Comparing Deep Neural Networks, Ensemble Classifiers, and Support Vector Machine Algorithms for Object-Based Urban Land Use/Land Cover Classification. **Remote Sensing**, [S. l.], v. 11, n. 14, p. 1713, 2019. Disponível em: <http://doi.org/10.3390/rs11141713>. Acesso em: 10 nov. 2022.

LEI, Tao; NANDI, Asoke K. **Image Segmentation: Principles, Techniques, and Applications**. [s.l.]: Wiley, 2022.

MAI, Sinh Dinh; NGO, Long Thanh. Multiple kernel approach to semi-supervised fuzzy clustering algorithm for land-cover classification. **Engineering Applications of**

Artificial Intelligence, [S. l.], v. 68, p. 205–213, 2018. Disponível em: <http://doi.org/10.1016/j.engappai.2017.11.007>. Acesso em: 14 nov. 2022.
MAPBIOMAS. Mapbiomas Brasil - O projeto. 2019. Disponível em: <https://mapbiomas.org/o-projeto>. Acesso em: 18 fev. 2023.

MAXWELL, Aaron E.; WARNER, Timothy A.; FANG, Fang. Implementation of machine-learning classification in remote sensing: an applied review. **International Journal of Remote Sensing**, [S. l.], v. 39, n. 9, p. 2784–2817, 2018. Disponível em: <http://doi.org/10.1080/01431161.2018.1433343>. Acesso em: 18 nov. 2022.

MORA, Brice; TSENDBAZAR, Nandin-Erdene; HEROLD, Martin; ARINO, Olivier. Global Land Cover Mapping: Current Status and Future Trends. *Em*: MANAKOS, Ioannis; BRAUN, Matthias (org.). **Land Use and Land Cover Mapping in Europe: Practices & Trends**. Remote Sensing and Digital Image Processing Dordrecht: Springer Netherlands, 2014. p. 11–30. Disponível em: https://doi.org/10.1007/978-94-007-7969-3_2. Acesso em: 17 jan. 2023.

MCNEILL, Graham; HALE, Scott A. Generating Tile Maps. **Computer Graphics Forum**, [S. l.], v. 36, n. 3, p. 435–445, 2017. Disponível em: <http://doi.org/10.1111/cgf.13200>. Acesso em: 22 abr. 2023.

MEYER, William B.; TURNER, B. L. Human population growth and global land-use/cover change. **Annual Review of Ecology, Evolution, and Systematics**, [S. l.], v. 23, n. 1, p. 39–61, 1992. Disponível em: <http://doi.org/10.1146/annurev.es.23.110192.000351>. Acesso em: 22 abr. 2023.

MEYER, William B.; TURNER, B. L.; UNIVERSITY CORPORATION FOR ATMOSPHERIC RESEARCH (org.). **Changes in land use and land cover: a global perspective**. Cambridge, England: Cambridge University Press, 1994.

NABOUREH, Amin; BIAN, Jinhui; LEI, Guangbin; LI, Ainong. A review of land use/land cover change mapping in the China-Central Asia-West Asia economic corridor countries. **Big Earth Data**, [S. l.], v. 5, p. 237–257, 2021. Disponível em: <http://doi.org/10.1080/20964471.2020.1842305>. Acesso em: 22 abr. 2023.

NOBLE, William S. What is a support vector machine? **Nature Biotechnology**, [S. l.], v. 24, n. 12, p. 1565–1567, 2006. Disponível em: <http://doi.org/10.1038/nbt1206-1565>. Acesso em: 22 abr. 2023.

OGC GeoTIFF Standard OGC. 2023. Disponível em: <https://www.ogc.org/standards/geotiff>. Acesso em: 7 mar. 2022.

OTUKEI, J. R.; BLASCHKE, T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. **International Journal of Applied Earth Observation and Geoinformation**, [S. l.], v. 12, p. S27–S31, 2010. Disponível em: <http://doi.org/10.1016/j.jag.2009.11.002>. Acesso em: 22 fev. 2023.

PANDEY, Prem Chandra; KOUTSIAS, Nikos; PETROPOULOS, George P.; SRIVASTAVA, Prashant K.; BEN DOR, Eyal. Land use/land cover in view of earth

observation: data sources, input dimensions, and classifiers—a review of the state of the art. **Geocarto International**, [S. l.], v. 36, n. 9, p. 957–988, 2021. Disponível em: <http://doi.org/10.1080/10106049.2019.1629647>. Acesso em: 18 nov. 2022.

PHIRI, Darius; MORGENROTH, Justin. Developments in Landsat Land Cover Classification Methods: A Review. **Remote Sensing**, [S. l.], v. 9, n. 9, p. 967, 2017. Disponível em: <http://doi.org/10.3390/rs9090967>. Acesso em: 18 out. 2022.

QGIS. — **QGIS Python Plugins Repository**. 2023a. Disponível em: <https://plugins.qgis.org/plugins/>. Acesso em: 17 fev. 2023.

QGIS. **Descubra o QGIS**. 2023b. Disponível em: https://qgis.org/pt_BR/site/about/index.html. Acesso em: 17 fev. 2023.

QGIS DEVELOPMENT TEAM. **QGIS Geographic Information System**. [S. l.]: QGIS Association, 2023. Disponível em: <https://www.qgis.org>. Acesso em: 16 fev. 2023.

REBALA, Gopinath; RAVI, Ajay; CHURIWALA, Sanjay. Machine Learning Definition and Basics. *In*: REBALA, Gopinath; RAVI, Ajay; CHURIWALA, Sanjay (org.). **An Introduction to Machine Learning**. Cham: Springer International Publishing, 2019. p. 1–17. Disponível em: http://doi.org/10.1007/978-3-030-15729-6_1. Disponível em: https://doi.org/10.1007/978-3-030-15729-6_1. Acesso em: 10 dez. 2022.

RADOČAJ, Dorijan; OBHOĐAŠ, Jasmina; JURISIĆ, Mladen; GAŠPAROVIĆ, Mateo. Global Open Data Remote Sensing Satellite Missions for Land Monitoring and Conservation: A Review. **Land**, [S. l.], v. 9, n. 11, p. 402, 2020. Disponível em: <http://doi.org/10.3390/land9110402>. Acesso em: 22 abr. 2023.

RWANGA, Sophia S.; NDAMBUKI, J. M. Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. **International Journal of Geosciences**, [S. l.], v. 8, n. 4, p. 611–622, 2017. DOI: 10.4236/ijg.2017.84033.

SHRAWANKAR, Urmila; B. KHANDARE, Shrinivas. Preprocessing Algorithm for High Resolution Satellite Imagery. *In*: INTERNATIONAL CONFERENCE ON RESEARCH IN INTELLIGENT COMPUTING IN ENGINEERING, 2016, **Anais** [...]. Disponível em: https://www.academia.edu/download/46728238/RICE-2016_ShrinivasKhandare.pdf. Acesso em: 26 abr. 2023.

TALUKDAR, Swapan; SINGHA, Pankaj; MAHATO, Susanta; SHAHFAHAD; PAL, Swades; LIOU, Yuei-An; RAHMAN, Atiqur. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. **Remote Sensing**, [S. l.], v. 12, 2020. Disponível em: <http://doi.org/10.3390/rs12071135>. Acesso em: 18 nov. 2022.

THANH NOI, Phan; KAPPAS, Martin. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. **Sensors**, [S. l.], v. 18, n. 1, p. 18, 2018. Disponível em: <http://doi.org/10.3390/s18010018>. Acesso em: 20 nov. 2022.

USGS. **National Land Cover Database - U.S. Geological Survey**. 2018. Disponível em: <https://www.usgs.gov/centers/eros/science/national-land-cover-database>. Acesso em: 16 fev. 2023.

USGS. **EarthExplorer**. 2022. Disponível em: <https://earthexplorer.usgs.gov/>. Acesso em: 19 fev. 2023.

WANG, Junye; BRETZ, Michael; DEWAN, M. Ali Akber; DELAVAR, Mojtaba Aghajani. Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. **Science of The Total Environment**, [S. l.], v. 822, p. 153559, 2022. Disponível em: <http://doi.org/10.1016/j.scitotenv.2022.153559>. Acesso em: 16 nov. 2022.

WULDER, Michael A.; COOPS, Nicholas C.; ROY, David P.; WHITE, Joanne C.; HERMOSILLA, Txomin. Land cover 2.0. **International Journal of Remote Sensing**, [S. l.], v. 39, n. 12, p. 4254–4284, 2018. Disponível em: <http://doi.org/10.1080/01431161.2018.1452075>. Acesso em: 22 nov. 2022.

YANG, Limin; JIN, Suming; DANIELSON, Patrick; HOMER, Collin; GASS, Leila; BENDER, Stacie M.; CASE, Adam; COSTELLO, Catherine; DEWITZ, Jon; FRY, Joyce. A new generation of the United States National Land Cover Database: requirements, research priorities, design, and implementation strategies. **Isprs Journal Of Photogrammetry And Remote Sensing**, [S.L.], v. 146, p. 108-123, dez. 2018. Disponível em: <http://dx.doi.org/10.1016/j.isprsjprs.2018.09.006>. Acesso em: 10 nov. 2022.

YOSHIDA, T.; OMATU, S. Neural network approach to land cover mapping. **IEEE Transactions on Geoscience and Remote Sensing**, [S. l.], v. 32, n. 5, p. 1103–1109, 1994. Disponível em: <http://doi.org/10.1109/36.312899>. Acesso em: 22 mar. 2023.

YOUNG, Nicholas E.; ANDERSON, Ryan S.; CHIGNELL, Stephen M.; VORSTER, Anthony G.; LAWRENCE, Rick; EVANGELISTA, Paul H. A survival guide to Landsat preprocessing. **Ecology**, [S. l.], v. 98, n. 4, p. 920–932, 2017. Disponível em: <http://doi.org/10.1002/ecy.1730>. Acesso em: 22 abr. 2023.

ZHANG, J.; WANG, B.; CHEN, X.; WU, X.; ZHANG, D. Research trends on land use changes during 1991–2015: A Bibliometric Analysis. **Lowland Technology International**, [S. l.], v. 21, n. 1, June, p. 61–70, 2019. Acesso em: 22 fev. 2023.

APÊNDICE A – Algoritmo para classificação no Earth Engine

```

/**** Start of imports. If edited, may not auto-convert in the playground. ****/
var floresta = /* color: #d63000 */ee.FeatureCollection(
  [ee.Feature(
    ee.Geometry.Polygon(
      [[[-41.94064818202731, -22.16573689840616],
        [-41.94167815028903, -22.181315583164015],
        [-41.90889082729098, -22.17940808187007],
        [-41.915585620992154, -22.167008692479506]]]],
    {
      "landcover": 1,
      "system:index": "0"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[-41.97281794657616, -22.14737115512871],
        [-41.96869807352928, -22.163110852182662],
        [-41.95153193583397, -22.16120310389094],
        [-41.955651808880845, -22.14673517077221]]]],
    {
      "landcover": 1,
      "system:index": "1"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[-42.107879991358715, -22.296564241058423],
        [-42.08848225576301, -22.29624658733744],
        [-42.10925328237434, -22.28147489192407]]]],
    {
      "landcover": 1,
      "system:index": "2"
    }
  )),
  ee.Feature(

```

```

ee.Geometry.Polygon(
  [[[-42.1736861013183, -22.36752280320633],
    [-42.16132648217768, -22.368475275343474],
    [-42.16510303247065, -22.36212533802761],
    [-42.1762610219726, -22.361649081056335]]]),
  {
    "landcover": 1,
    "system:index": "3"
  }
)),
ee.Feature(
  ee.Geometry.Polygon(
    [[[-42.17537310294686, -22.40423810916811],
      [-42.169193293376544, -22.416457780313397],
      [-42.16198351554451, -22.416854504868457],
      [-42.16927912406502, -22.400429160831663]]]),
    {
      "landcover": 1,
      "system:index": "4"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[-42.00286218588269, -22.31317886063867],
      [-42.004235476898316, -22.296344220515298],
      [-42.01728174154675, -22.30142658983627]]]),
    {
      "landcover": 1,
      "system:index": "5"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[-41.74397778412581, -21.81474391391312],
      [-41.70380902191878, -21.8408780012979],
      [-41.69282269379378, -21.83434492710107],
      [-41.71925854584456, -21.80390642903968],

```



```

        [-41.74191784760237, -21.80119692966345]]]),
    {
      "landcover": 1,
      "system:index": "6"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[-41.81169819733382, -21.815222019351438],
        [-41.773331879584795, -21.81625790898941],
        [-41.763118027656084, -21.809245586601822],
        [-41.78440403839827, -21.798965538751688],
        [-41.80397343537093, -21.805898675291672]]]],
    {
      "landcover": 1,
      "system:index": "7"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[-41.83679255778829, -21.885553383927792],
        [-41.82511958415548, -21.87583637593712],
        [-41.8576494150881, -21.871216582678088],
        [-41.86709079082052, -21.878066567433308]]]],
    {
      "landcover": 1,
      "system:index": "8"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[-41.100936469349655, -21.389421410728403],
        [-41.09630161217192, -21.404285578620133],
        [-41.08325534752348, -21.378711880477212],
        [-41.092696723255905, -21.373436749141632]]]],
    {
      "landcover": 1,

```

```

      "system:index": "9"
    })),
  veg_nao_florestal = /* color: #98ff00 */ee.FeatureCollection(
    [ee.Feature(
      ee.Geometry.Polygon(
        [[[-41.36474079719119, -21.94956594471734],
          [-41.361307569652126, -21.93889792517187],
          [-41.36774487128787, -21.93682791823781],
          [-41.38122028937869, -21.943197074090975]]]],
      {
        "landcover": 2,
        "system:index": "0"
      }
    ),
    ee.Feature(
      ee.Geometry.Polygon(
        [[[-41.29410214057498, -22.14436242010548],
          [-41.28800816169314, -22.143885421913264],
          [-41.280970045238064, -22.14229541626794],
          [-41.27015537849002, -22.138956345970325],
          [-41.261228986888455, -22.137525291605805],
          [-41.25281757941775, -22.136412239269127],
          [-41.25728077521853, -22.132118955005947],
          [-41.267237135081814, -22.134345118736086],
          [-41.276678510814236, -22.13680975897057],
          [-41.28114170661502, -22.136173726909732],
          [-41.28680653205447, -22.140466887578967],
          [-41.2951321088367, -22.138161317563057],
          [-41.300882764964626, -22.14229541626794],
          [-41.30697674384646, -22.14340842210501],
          [-41.304401823192165, -22.145554908515606]]]],
      {
        "landcover": 2,
        "system:index": "1"
      }
    )
  ]

```

```

ee.Feature(
  ee.Geometry.Polygon(
    [[[-41.981630668344614, -22.323943850133457],
      [-41.96231876343739, -22.307983985979597],
      [-41.98678050965321, -22.31330414344043]]]],
    {
      "landcover": 2,
      "system:index": "2"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[-41.147748493749354, -21.59368643942177],
      [-41.13384392221615, -21.586663267410202],
      [-41.13779213388607, -21.571897533766144],
      [-41.14268448312924, -21.570780065645742],
      [-41.143542790014, -21.566230285062108],
      [-41.14637520273373, -21.564314545250394]]]],
    {
      "landcover": 2,
      "system:index": "3"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[-41.09937486010592, -21.720623749350775],
      [-41.09765824633639, -21.716317905272223],
      [-41.10426720934908, -21.70858300982852],
      [-41.10924538928072, -21.712251156412155]]]],
    {
      "landcover": 2,
      "system:index": "4"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[-41.13149367825237, -21.235796957193298],

```

```

        [-41.131064524809986, -21.23779700827914],
        [-41.11973487393108, -21.234436906957722],
        [-41.11329757229534, -21.238197015240452],
        [-41.11098014370647, -21.22931660533044]]]),
    {
      "landcover": 2,
      "system:index": "5"
    }
  ]]),
agro = /* color: #0b4a8b */ee.FeatureCollection(
  [ee.Feature(
    ee.Geometry.Polygon(
      [[[[-41.548033878010905, -21.68501391034157],
        [-41.59043423811833, -21.73907858654711],
        [-41.502371951741374, -21.740832557180337]]]],
    {
      "landcover": 3,
      "system:index": "0"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[[-41.50035625694124, -21.39551593263885],
        [-41.39735943076936, -21.482438242008175],
        [-41.31633526084749, -21.41469442153796],
        [-41.420018732527176, -21.392319273205516]]]],
    {
      "landcover": 3,
      "system:index": "1"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[[-41.75475841758577, -22.126648127018274],
        [-41.7949271797928, -22.158448361834193],
        [-41.76574474571077, -22.203274481368123],
        [-41.705663263777176, -22.171802318012123],

```

```

        [-41.71630626914827, -22.147319096657192],
        [-41.69639354942171, -22.080525038801397]]]),
    {
      "landcover": 3,
      "system:index": "2"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[-41.1736604160384, -21.81798839754239],
      [-41.22189726296223, -21.8307370402442],
      [-41.18516172829426, -21.866267882083037],
      [-41.14842619362629, -21.84826456858081]]]),
    {
      "landcover": 3,
      "system:index": "3"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[-41.2458884252544, -21.28882161982463],
      [-41.34373541011768, -21.274745560841477],
      [-41.27301092281299, -21.363977738631455]]]),
    {
      "landcover": 3,
      "system:index": "4"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[-41.86958572610147, -22.277266737451583],
      [-41.82392379983194, -22.30315666284991],
      [-41.82529709084756, -22.29362718672411],
      [-41.85911438210733, -22.273454318937315]]]),
    {
      "landcover": 3,
      "system:index": "5"
    }
  )),

```

```

    })),
  agua = /* color: #ffc82d */ee.FeatureCollection(
    [ee.Feature(
      ee.Geometry.Polygon(
        [[[[-41.37100865541169, -21.963173389947997],
          [-41.40602757631013, -21.954894595050582],
          [-41.43109013734529, -21.957760386325965],
          [-41.43383671937654, -21.993100391989394],
          [-41.40637089906404, -22.026522026452437],
          [-41.36105229554841, -22.08443417461553],
          [-41.341482898575755, -22.07648061464799],
          [-41.37856175599763, -22.03352365578052],
          [-41.35315587220857, -22.015382357306517],
          [-41.34079625306794, -22.011244542831665],
          [-41.3548724859781, -21.973362013754596]]]],
        {
          "landcover": 5,
          "system:index": "0"
        }
      )),
    ee.Feature(
      ee.Geometry.Polygon(
        [[[[-41.53232138485179, -21.7745712671296],
          [-41.52442496151195, -21.7779189101507],
          [-41.50691550106273, -21.769310527382075],
          [-41.48889105648265, -21.75926676131943],
          [-41.49884741634593, -21.75097613923616]]]],
        {
          "landcover": 5,
          "system:index": "1"
        }
      )),
    ee.Feature(
      ee.Geometry.Polygon(
        [[[[-41.592970142142754, -22.25312675600689],
          [-41.59254098870037, -22.248916478452056],
          [-41.592970142142754, -22.25312675600689]]]]],

```

```

        [-41.59897829033611, -22.242719990719532],
        [-41.60172487236736, -22.244785517086264],
        [-41.599149951713066, -22.24685101299784],
        [-41.5968325231242, -22.250425837790775],
        [-41.5982916448283, -22.25352394546847],
        [-41.59700418450115, -22.254874381210342]]]),
    {
      "landcover": 5,
      "system:index": "2"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[[-41.61916383088673, -22.21141524238103],
        [-41.6267598468169, -22.219083081773476],
        [-41.624227841506844, -22.22019547961828],
        [-41.61787637055958, -22.211574165584928]]]],
    {
      "landcover": 5,
      "system:index": "3"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[[-41.000608387083545, -21.846013812833572],
        [-41.01228136071636, -21.851829382115827],
        [-41.01502794274761, -21.85644980266444],
        [-41.00678819665386, -21.851192071002714],
        [-40.998119297117725, -21.84696981308031]]]],
    {
      "landcover": 5,
      "system:index": "4"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[[-41.2025325701458, -21.684842277942717],

```

```

        [-41.1875121996624, -21.681891257919432],
        [-41.18519477107353, -21.656206857325742],
        [-41.190859596512986, -21.651659763355767]]]),
    {
      "landcover": 5,
      "system:index": "5"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[[-41.15575484492607, -21.703743323859843],
      [-41.14528350093193, -21.69568875106459],
      [-41.14339522578545, -21.69098339611415],
      [-41.13987616755791, -21.687872992354354],
      [-41.137730400345994, -21.689627587347562],
      [-41.13541297175713, -21.683725682879057],
      [-41.14176444270439, -21.685958864353566],
      [-41.14751509883232, -21.694651991144227],
      [-41.15738562800713, -21.700633195689132]]]],
    {
      "landcover": 5,
      "system:index": "6"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[[-41.263131318653066, -21.763852987374026],
      [-41.25231665190502, -21.757595406277137],
      [-41.2503854614143, -21.757196824991638],
      [-41.24549311217113, -21.75376898025672],
      [-41.2512866836433, -21.75536333682533],
      [-41.26476210173412, -21.761660872175057]]]],
    {
      "landcover": 5,
      "system:index": "7"
    }
  )),

```



```

ee.Feature(
  ee.Geometry.Polygon(
    [[[-41.30755536810471, -21.75466588162398],
      [-41.29884355322434, -21.75458616382878],
      [-41.29528157965256, -21.75155685481533],
      [-41.30699746862961, -21.7517960131136],
      [-41.31223314062668, -21.75167643401426]]]),
    {
      "landcover": 5,
      "system:index": "8"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[-41.845087966068284, -21.590903062095894],
      [-41.84792037878801, -21.586074575110473],
      [-41.850323638065355, -21.587112114821142],
      [-41.845045050724046, -21.59321748636995],
      [-41.84195514593889, -21.597407298254335],
      [-41.84186931525041, -21.593895845093517]]]),
    {
      "landcover": 5,
      "system:index": "9"
    }
  )),
ee.Feature(
  ee.Geometry.Polygon(
    [[[-41.045435814846975, -22.017814123637027],
      [-41.04208841799639, -22.02362267500085],
      [-41.042174248684866, -22.02768056270823],
      [-41.040457634915334, -22.026327946390797],
      [-41.040715126980764, -22.022588292873145],
      [-41.03934183596514, -22.020201228366066],
      [-41.04474916933916, -22.0163818414945]]]),
    {
      "landcover": 5,

```

```

      "system:index": "10"
    })),
  nao_veg = /* color: #00ffff */ee.FeatureCollection(
    [ee.Feature(
      ee.Geometry.Polygon(
        [[[-41.32650034076855, -21.717364265158295],
          [-41.32598535663769, -21.732354315655847],
          [-41.294571324655266, -21.72462028260387],
          [-41.29302637226269, -21.714972091267025],
          [-41.30736009723827, -21.719437450353222],
          [-41.313539906808586, -21.7168060948083],
          [-41.32160799152538, -21.717842695163064]]]],
        {
          "landcover": 4,
          "system:index": "0"
        }
      )),
    ee.Feature(
      ee.Geometry.Polygon(
        [[[-41.35545170529187, -21.738244876128366],
          [-41.364549758270385, -21.74494177856928],
          [-41.357339980438354, -21.74868871873181],
          [-41.35090267880261, -21.74964536864087],
          [-41.34377873165906, -21.75594315479907],
          [-41.338113906219604, -21.76295808501989],
          [-41.328844191864135, -21.765110324424082],
          [-41.32895521275706, -21.7756434880907],
          [-41.33144430272288, -21.78496880098521],
          [-41.32234624974436, -21.789113189807665],
          [-41.31101659886546, -21.777636127262248],
          [-41.304321805164285, -21.768708887981823],
          [-41.306170603638314, -21.75786208803208],
          [-41.318873545532846, -21.757622939837837],
          [-41.32719912231507, -21.754673412686742],
          [-41.33457934620016, -21.74349297344475],

```

```

        [-41.34470736744039, -21.743094353016918]]]),
    {
      "landcover": 4,
      "system:index": "1"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[-41.03209460737742, -22.030399454175935],
        [-41.04458297255076, -22.037281551979873],
        [-41.05041945936716, -22.039668328567988],
        [-41.04827369215525, -22.041657278311565],
        [-41.033124575639135, -22.034536709148686],
        [-41.02333987715281, -22.029126428324137],
        [-41.02677310469187, -22.025983596750823]]]],
    {
      "landcover": 4,
      "system:index": "2"
    }
  )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[-41.07018981135622, -21.489515896464287],
        [-41.07418930564352, -21.496685912796593],
        [-41.074618459085904, -21.505310558680577],
        [-41.077365041117154, -21.509143570452107],
        [-41.0783091786904, -21.525273054033015],
        [-41.073845982889615, -21.523516464478405],
        [-41.074618459085904, -21.515292149088594],
        [-41.07215222058484, -21.50467690063172],
        [-41.069319807865114, -21.496451519363102],
        [-41.065543257572145, -21.49006255281053],
        [-41.06794651684949, -21.48634883709315]]]],
    {
      "landcover": 4,
      "system:index": "3"
    }
  )
)

```

```

    }),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[-41.27425674034963, -21.83285116321331],
        [-41.27554420067678, -21.837790883807845],
        [-41.272368465203144, -21.838189240936018],
        [-41.26987937523732, -21.835241371905852],
        [-41.26404288842092, -21.834843006564824],
        [-41.257348094719745, -21.83922496427482],
        [-41.25554565026174, -21.836675477944343],
        [-41.26352790429006, -21.830142211734593],
        [-41.27502921654592, -21.828389333437237]]]],
      {
        "landcover": 4,
        "system:index": "4"
      }
    )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[-41.79215697339846, -22.366339890511643],
        [-41.79790762952639, -22.369832286163646],
        [-41.79001120618655, -22.374673871578377],
        [-41.78889540723635, -22.381261332076466],
        [-41.80142668775393, -22.39665731241929],
        [-41.80352381836768, -22.40672594755007],
        [-41.79296664368506, -22.398552597224807],
        [-41.7715948022544, -22.38744241853182],
        [-41.77235640107245, -22.377811565739055]]]],
      {
        "landcover": 4,
        "system:index": "5"
      }
    )),
  ee.Feature(
    ee.Geometry.Polygon(
      [[[-41.72287966939705, -22.311235301409702],

```

```

        [-41.71541239949959, -22.31695234428238],
        [-41.702366134851154, -22.30694736562428],
        [-41.70871760579842, -22.299959336321496]]]),
    {
      "landcover": 4,
      "system:index": "6"
    })),
    norte_fluminense = ee.FeatureCollection("users/vitordsgoncalves/NF");
/***** End of imports. If edited, may not auto-convert in the playground. *****/

var regioes = floresta.merge(veg_ao_florestal)
    .merge(agro)
    .merge(nao_veg)
    .merge(agua);

/*****
* Função para aplicar máscara de nuvens
* Fonte: https://developers.google.com/earth-
engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED
*/
function maskS2clouds(image) {
  var qa = image.select('QA60');

  var cloudBitMask = 1 << 10;
  var cirrusBitMask = 1 << 11;

  var mask = qa.bitwiseAnd(cloudBitMask).eq(0)
    .and(qa.bitwiseAnd(cirrusBitMask).eq(0));

  return image.updateMask(mask).divide(10000);
}
/*****/

//Obtém as imagens do Sentinel-2 ao longo do ano de 2021

```

```
//com percentual de nuvens inferior a 20% e aplica funação de máscara
var dataset = ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED')
    .filterDate('2021-01-01', '2021-12-31')
    .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE',20))
    .map(maskS2clouds);
```

```
//aplica o filtro mediano
var mediana = dataset.median().clip(norte_fluminense)
```

```
//visualis a imagem
var visualization = {
  min: 0.0,
  max: 0.3,
  bands: ['B4', 'B3', 'B2'],
};
```

```
Map.setCenter(-41.323889, -21.753889);
Map.addLayer(mediana, visualization, 'RGB');
```

```
//obtem amostras da região seleciona para treinamento e validação
var sample = mediana.sampleRegions({
  collection: regioes,
  properties: ['landcover'],
  scale: 200
});
```

```
sample = sample.randomColumn();
var trainingSample = sample.filter('random <= 0.8');
var validationSample = sample.filter('random > 0.8');
```

```
// aplica o algoritmo classificador
```

```

//var trainedClassifier = ee.Classifier.smileRandomForest(10).train({
var trainedClassifier = ee.Classifier.libsvm().train({
  features: trainingSample,
  classProperty: "landcover",
  inputProperties: ['B2','B3','B4','B8','B11','B12']
});

print('Results of trained classifier', trainedClassifier.explain());

var trainAccuracy = trainedClassifier.confusionMatrix();
print('Training error matrix', trainAccuracy);
print('Training overall accuracy', trainAccuracy.accuracy());

validationSample = validationSample.classify(trainedClassifier);
var validationAccuracy = validationSample.errorMatrix("landcover",
'classification');
print('Validation error matrix', validationAccuracy);
print('Validation accuracy', validationAccuracy.accuracy());

var trainAccuracy = trainedClassifier.confusionMatrix();
print('Resubstitution error matrix: ', trainAccuracy);
print('Training overall accuracy: ', trainAccuracy.accuracy());

var classified = mediana.classify(trainedClassifier);

Map.addLayer(classified, {min: 1, max: 12}, "imagem classificada");

// exporta a imagem classificada
Export.image.toDrive({
  image: classified,
  scale: 30,
  region: norte_fluminense
});

```

```
// exporta a região selecionada para o treinamento
Export.table.toDrive({
  collection: regioes,
  description:'regioes_treino',
  fileFormat: 'shp'
});
```