

JONILSON MICHEL FONTES GALVÃO

**CLASSIFICAÇÃO DE REGIÕES EM IMAGENS DO SATÉLITE
SENTINEL 2 ATRAVÉS DE INTELIGÊNCIA ARTIFICIAL**

Sorocaba
2023

PROGRAMA DE PÓS-GRADUAÇÃO em

*ciências
ambientais*

JONILSON MICHEL FONTES GALVÃO

**CLASSIFICAÇÃO DE REGIÕES EM IMAGENS DO SATÉLITE
SENTINEL 2 ATRAVÉS DE INTELIGÊNCIA ARTIFICIAL**

Dissertação apresentada como requisito para a obtenção do título de Mestre em Ciências Ambientais da Universidade Estadual Paulista “Júlio de Mesquita Filho” na Área de Concentração Diagnóstico, Tratamento e Recuperação Ambiental

Orientador: Prof. Dr. Antônio Cesar
Germanto Martins

Sorocaba

2023

G182c Galvão, Jonilson Michel Fontes
Classificação de Regiões em Imagens do Satélite Sentinel 2
através de Inteligência Artificial / Jonilson Michel Fontes
Galvão. -- Sorocaba, 2023
49 f.

Dissertação (mestrado) - Universidade Estadual Paulista
(Unesp), Instituto de Ciência e Tecnologia, Sorocaba
Orientador: Antônio Cesar Germano Martins

1. Classificação Supervisionada. 2. Geoprocessamento. 3.
Machine Learning. 4. Sentinel. I. Título.

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

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: Classificação de regiões em imagens do Sentinel 2 através de inteligência artificial

AUTOR: JONILSON MICHEL FONTES GALVÃO

ORIENTADOR: ANTONIO CESAR GERMANO MARTINS

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

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

Prof. Dr. DARLLAN COLLINS DA CUNHA E SILVA (Participação Virtual)
Departamento de Engenharia Ambiental / Unesp - ICT Sorocaba

Prof. Dr. JOSÉ CARLOS DE SOUZA (Participação Virtual)
Departamento de Geografia / Universidade Estadual de Goiás - Câmpus Minaçu

Sorocaba, 28 de fevereiro de 2023

AGRADECIMENTOS

Aos que direta ou indiretamente proporcionaram a confecção deste material.

Ao Prof. Dr. Antônio Cesar Germano Martins, que por sua competência e experiência, orientou-me na concretização da dissertação.

Ao Prof. Dr. Darllan Collins da Cunha e Silva, por orientar, retirar dúvidas e no incentivo à produção técnica na pesquisa acadêmica.

A Universidade Estadual “Júlio de Mesquita Filho”, Campus de Sorocaba, pela base no desenvolvimento da dissertação, aos professores, técnicos e funcionários terceirizados.

Aos amigos dos Laboratório do GASI, em especial aos colegas frequentadores da sala 89 do prédio H.

A CAPES, pela parceria no apoio financeiro, o qual foi necessário para o desenvolvimento deste estudo.

Em especial à minha Família, aos meus avós maternos e aos meus sobrinhos e, até o momento, única sobrinha, que estiveram presentes, ainda que à distância, em momentos de alegria e nas dificuldades.

RESUMO

A disponibilidade de recursos para estudo e compreensão da dinâmica do uso e cobertura do solo através de sensoriamento remoto por meio de técnicas de classificação com auxílio de aprendizado de máquina tem ganho destaque nos últimos anos. Este trabalho visa apresentar resultados da classificação de regiões em imagens do satélite Sentinel-2 da Área de Preservação Ambiental de Itupararanga, especificamente no entorno da represa de Itupararanga, no Estado de São Paulo. Através do pré-processamento e utilização de índices espectrais (NDVI, GNDVI, NDWI, MNDWI e NDBI) foi feita a classificação da cobertura do solo da região usando as técnicas de inteligência artificial: Random Forest (RF) e Support Vector Machine (SVM). As métricas de avaliação de desempenho registraram excelentes resultados com 99,27% de acurácia global para SVM com kernel radial e 99,15% para RF. O cálculo da cobertura do solo realizado por meio do SVM com kernel radial resultou em 3,4% de agricultura, 7,0% de água, 24,0% de vegetação arbórea, 4,2% de eucalipto, 27,7% de vegetação rasteira, 14,9% de solo exposto, 7,2% de interação solo e arbusto, 14,0% de área urbana e 0,2% de nuvem nas imagens analisadas.

Palavras-chave: Classificação-Supervisionada; Geoprocessamento; Machine-Learning; Sentinel.

ABSTRACT

The availability of resources for studying and understanding of land cover dynamics with remote sensing and classification techniques based on machine learning has gained prominence in recent years. This work aims to present results of the classification of regions in Sentinel-2 satellite images from Itupararanga Environmental Preservation Area, specifically around Itupararanga dam, in the State of São Paulo. Through the pre-processing and use of spectral indexes (NDVI, GNDVI, NDWI, MNDWI and NDBI), soil cover of the region was classification using the following artificial intelligence techniques: Random Forest (RF) and Support Vector Machine (SVM). Performance evaluation metrics showed excellent results with 99,27% overall accuracy for SVM with radial kernel and 99,15% for RF. The calculation of land cover performed by means of SVM with radial kernel resulted in 3,4% of agriculture, 7,0% of water, 24,0% of tree vegetation, 4,2% of eucalyptus, 27,7% of undergrowth, 14,9% of exposed soil, 7,2% of soil and shrub interaction, 14,0% of urban area and 0,2% of cloud in the analysed images.

Keywords: Supervised Classification; Geoprocessing; Machine-learning; Sentinel

Es difícil saber a ciencia cierta si una máquina realmente "sabe", "piensa", etc., porque sigue siendo muy difícil definir estas cosas. Entendemos el proceso mental humano sólo un poco mejor de lo que un pez entiende el nadar

JOHN MCCARTHY

Tradução Livre

É difícil saber ao certo se uma máquina "sabe", "pensa", etc., porque ainda é muito difícil definir essas coisas. Compreendemos o processo mental humano apenas um pouco melhor do que um peixe entende de natação.

JOHN MCCARTHY

LISTA DE ABREVIATURA E SIGLAS

AM	Aprendizado de Máquina
CS	Cobertura do Solo
IA	Inteligência Artificial
REM	Radiação Eletromagnética
RF	Random Forest
S2	Sentinel-2
SIG	Sistema de informação geográfica
SR	Sensoriamento Remoto
SVM	Support Vector Machine
SWIR	Infravermelho de Ondas Curtas
US	Uso do Solo

LISTA DE FIGURAS

Figura 1. Os principais constituintes do SR	15
Figura 2. Bandas espectrais do Sentinel 2 versus resolução espacial do Sentinel 2.	18
Figura 3. Modelo básico de uma árvore de decisão	20
Figura 4. N-árvores para escolha da decisão final através de indicação (voto) ..	21
Figura 5. Máquina de vetor de suporte.	22
Figura 6. Vista geométrica dos kernels	23
Figura 7. Sistemática nas bases de dados	26
Figura 8. Número de artigos resultantes dos filtros aplicados pelo ano de publicação	27
Figura 9. Área de Estudo.....	31
Figura 10. Localização da obtenção das amostras, sendo A (agricultura), B (água), C (arbórea), D (eucalipto), E (nuvem), F (rasteira), G (Solo e arbusto), H (solo exposto) e I (Urbano).....	34
Figura 11. Agrupamento dos Matrizes de Confusão para os modelos RF e SVM, sendo (a) RF, (b) SVM Linear, (c) SVM Radial e, (d) SVM Polinomial.....	37
Figura 12. Variáveis de importância na divisão de nós e folhas para a RF	39
Figura 13. Mapa de cobertura do solo da área de estudo usando SVM radial. ...	40

LISTA DE TABELAS

Tabela 1. Divisão do espectro eletromagnético	16
Tabela 2. Características da SVM, equação e descrição para cada modelo de aprendizado de máquina.....	23
Tabela 3. Matriz de confusão com elementos p_{ij} representando a proporção da categoria i da classificação de uma determinada classe e da categoria j da classe de referência.	24
Tabela 4. Medidas de acurácias por categoria em que i representa a cobertura do solo da área mapeada e j representa a cobertura do solo de referência	25
Tabela 5. Informações do Produto adquirido	32
Tabela 6. Índices e características abordadas no estudo	33
Tabela 7. Amostras da área de estudo com o identificador, a classe e sua característica para o estudo.	35
Tabela 8. Bibliotecas utilizadas no Rstudio e respectivas funções	35
Tabela 9. Desempenho dos modelos de aprendizado de máquina	38
Tabela 10. Hectares e porcentagem referente à classe de estudo.....	41

SUMÁRIO

1. INTRODUÇÃO E JUSTIFICATIVAS	12
2. OBJETIVOS.....	14
2.1. OBJETIVO GERAL	14
2.2. OBJETIVOS ESPECÍFICOS	14
3. REFERECIAL TEÓRICO	15
3.1. SENSORIAMENTO REMOTO	15
3.2. SATÉLITE SENTINEL-2.....	17
3.3. INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA.....	18
3.4. ÁRVORE DE DECISÃO E RANDOM FOREST	19
3.5. MÁQUINA DE VETOR SUPORTE	22
3.6. MÉTRICAS DE VALIDAÇÃO	24
3.7. PESQUISA ORDENADA: IA E SR.....	25
3.7.1.Inteligência Artificial e sua aplicação ao Sensoriamento Remoto	27
4. METODOLOGIA	31
4.1. ÁREA DE ESTUDO.....	31
4.2. AQUISIÇÃO DA IMAGEM E PRÉ-PROCESSAMENTO	32
4.3. INDICES ESPECTRAIS	32
4.4. OBTENÇÃO DAS AMOSTRAS	33
4.5. FERRAMENTAS COMPUTACIONAL E APLICAÇÃO DOS MÉTODOS	35
4.5.1.Treinamento dos Modelos, Ajustes e Métrica de Validação	36
5. RESULTADOS E DISCUSSÃO	37
6. CONSIDERAÇÕES FINAIS	42
REFERÊNCIAS BIBLIOGRÁFICAS.....	43

1. INTRODUÇÃO E JUSTIFICATIVAS

A cobertura do solo resultante de propriedades biológicas e físicas da superfície do planeta Terra compreende uma grande diversidade, incluindo vegetações, corpos d'água, coberturas de neve, rochas, solos expostos e ambientes construídos (FORQUOR et al., 2017), sendo um elemento chave para a ciência do sistema terrestre e objeto de estudo na compreensão das complexas interações entre atividades humanas e mudanças superficiais (TU et al., 2020).

Os termos cobertura do solo (CS) e uso do solo (US), relacionados com a superfície do solo, ao longo dos anos, tornaram-se muito utilizados. Enquanto o uso da terra é o resultado da interação homem-natureza, a cobertura da terra representa a condição física da superfície (PRASAD et al., 2022). A CS e US são influenciados por ações humanas e condicionam as características e processo ambientais a estes dois fatores (XIE e NICULESCU, 2021).

Associado ao planejamento, gestão e programas de monitoramento em nível local, regional e nacional, a informação sobre o uso e cobertura do solo desempenha um papel fundamental na análise da informação da terra para o desenvolvimento, criação de políticas públicas e progresso necessário à dualidade homem-ambiente (RANA e SURYANARAYANA, 2020).

O avanço nas técnicas de sensoriamento remoto (SR) e o sistema de informações geográficas (SIG) fornecem ferramentas para o apoio no estudo da dinâmica ambiental (MUSHTAQ et al., 2021).

No entanto, a crescente disponibilidade de conjuntos de dados temporais arquivados, junto a necessidade de métodos para analisar quantidades gigantescas de arquivos são desafios para o cenário do SR, mas também uma solução (XIE e NICULESCU, 2021; EBRAHIMY et al., 2021).

O desenvolvimento da tecnologia de satélites proporciona armazenamento de dados em domínio público para fins de pesquisa (AGUILERA, 2020). O Sentinel 2 (S-2) é parte do Programa de observação da Terra “Copernicus” e registra, com imagens ópticas em alta resolução espacial, informações sobre corpos terrestre e aquático (RAIYANI et al., 2021). A disponibilidade das imagens de maior resolução espacial agrega melhoria na precisão de mapas de classificação de US e CS (LUCA et al., 2022).

O rápido crescimento das abordagens computacionais, a evolução das características dos sensores e a disponibilidade de dados de satélite impulsionaram o desenvolvimento de novos métodos de classificação de imagens (VASILAKOS; KAVROUDAKIS; GEORGANTA, 2020). Cada classificador tem seu processo de operação específico e, dependendo do classificador e das características do software, os resultados geralmente variam (GHAYOUR et al., 2021).

Os métodos mais utilizados atualmente para classificação de imagens são os supervisionados através de classificadores de aprendizado de máquina, especialmente os baseados em análise pixel a pixel (VASILAKOS; KAVROUDAKIS; GEORGANTA, 2020). Estes algoritmos usam amostras de locais de treinamento para classificação, através de assinaturas espectrais únicas atribuídas a cada classe pelo usuário (GHAYOUR et al., 2021).

A escolha do algoritmo ideal depende de parâmetros como condições do local, dados existentes e similaridade espectral das classes (GHAYOUR et al., 2021). O uso de *Support Vector Machine* (SVM) e *Random Forest* (RF) para a classificação de imagens do S-2 tem obtido resultados com alta acurácia (VASILAKOS; KAVROUDAKIS; GEORGANTA, 2020; GHAYOUR et al., 2021; LUCA et al., 2022).

As ferramentas do geoprocessamento são importantes na gestão e ordenamento territorial, principalmente em um país com uma gigantesca área. Aliado ao desenvolvimento dos grandes bancos de dados, o sensoriamento remoto junto ao campo da inteligência artificial tem recebido grande destaque na linha de modelagem ambiental, principalmente com satélites de alta resolução, como o Sentinel 2.

Aproveitar o uso de imagens de alta resolução espacial e temporal se faz necessário para manter a observação frente aos desafios de um país detentor de uma grande porção de vegetação e com o recurso hídrico abundantes.

Neste contexto, o presente trabalho aborda a avaliação de algoritmos de aprendizagem de máquina (SVM e RF) para a classificação de regiões em imagens, na área de proteção ambiental de Itupararanga, que é uma unidade de uso sustentável e detém a bacia hidrográfica da represa Itupararanga.

2. OBJETIVOS

2.1.OBJETIVO GERAL

- Empregar e avaliar a eficácia de métodos de Inteligência Artificial para classificar regiões em imagens do satélite Sentinel 2 no entorno da represa de Itupararanga.

2.2.OBJETIVOS ESPECÍFICOS

- Fazer uma pesquisa bibliográfica dos últimos cinco anos de trabalhos relacionadas aos estudos que abordem o Sentinel 2 e o uso de inteligência artificial em SR.
- Utilizar métodos supervisionados para a classificação de regiões em imagens do Sentinel 2.
- Desenvolver o Mapa de cobertura do solo no entorno da represa de Itupararanga.

3. REFERECIAL TEÓRICO

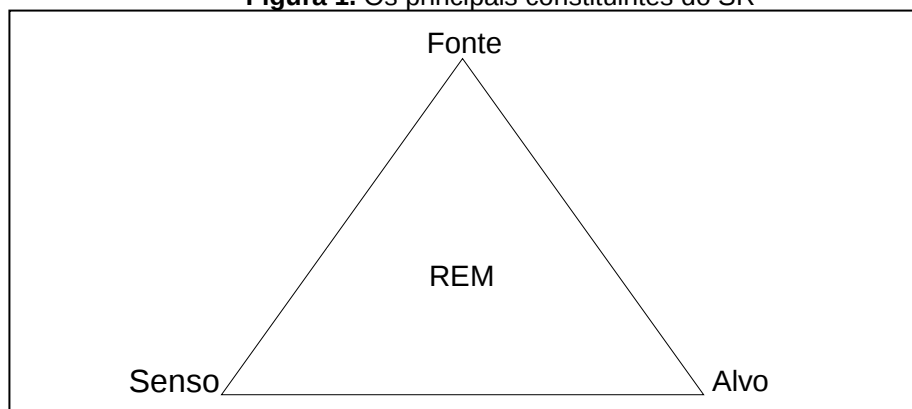
As próximas seções estão dedicadas ao referencial teórico sobre sensoriamento remoto e subáreas de uso da Inteligência Artificial.

3.1. SENSORIAMENTO REMOTO

O Sensoriamento Remoto (SR) é o conjunto de técnicas destinado à obtenção de informação sobre objetos sem que haja contato físico entre eles e o sistema de coleta de dados (NOVO; PONZONI, 2001), e que procuram o desenvolvimento da aquisição de imagens da superfície terrestre por meio da detecção e medição quantitativa das respostas das interações da radiação eletromagnética com os materiais terrestres (MENEZES et al., 2012). O objeto imageado é registrado pelo sensor por meio de medições da radiação eletromagnética, tal como a luz solar refletida da superfície de qualquer objeto (MENEZES et al., 2012).

No centro da Figura 1 está representada a Radiação Eletromagnética (REM) que é a ligação entre os demais constituintes do SR. Em cada vértice do triângulo há um elemento: a fonte, que para o SR advém do sol ou de um emissor artificial; o sensor, instrumento capaz de coletar e registrar a REM refletida ou emitida pelo objeto e o alvo (também denominado de objeto), que representa o elemento do qual se pretende extrair a informação (NOVO; PONZONI, 2001).

Figura 1. Os principais constituintes do SR



Fonte: NOVO; PONZONI (2001).

A REM representa um canal de comunicação eficiente e em alta velocidade entre o sensor e o fenômeno remoto (JENSEN, 2009). Devido ao elemento sensor, o SR pode fornecer informação nova e fundamental para diversas categorias das ciências, conforme descrito por JENSEN (2009, p. 8)

A ciência do sensoriamento remoto pode fornecer nova e fundamental informação científica. Sob condições controladas, o sensoriamento remoto pode fornecer informações biofísica básica, incluindo localização em x e y, elevação ou profundidade em z; biomassa; temperatura; e teor da água. (JENSEN, 2009, p.8)

O espectro eletromagnético representa a distribuição da radiação eletromagnética, por regiões, segundo o comprimento de onda ou a frequência (GUEDES; SILVA, 2018). A faixa de comprimento da REM é praticamente ilimitada, por isso, deve ser vista como um espectro contínuo (MENEZES et al., 2012).

A Tabela 1 apresenta uma divisão dos comprimentos de ondas em grupos, porém não existem limites rigorosamente precisos e, de acordo com a literatura, é comum encontrar diferenças nas definições dos mesmos (MENEZES et al., 2012). O intervalo de comprimentos de onda mais familiar é a faixa do visível da radiação solar, que é a que olho humano detecta (MENEZES et al., 2012).

Tabela 1. Divisão do espectro eletromagnético

Intervalo Espectral	Comprimento de Onda
Raios cósmicos	menor que $0,01 \text{ A}^\circ$
Raios gama	$0,01 - 0,1 \text{ A}^\circ$
Raios X	$0,1 - 10 \text{ A}^\circ$
Ultravioleta	$100\text{nm} - 0,38\mu\text{m}$
Visível	$0,38 - 0,76 \mu\text{m}$
Infravermelho próximo	$0,76 - 1,2 \mu\text{m}$
Infravermelho de ondas curtas	$1,2 - 3,0 \mu\text{m}$
Infravermelho médio	$3,0 - 5,0 \mu\text{m}$
Infravermelho termal	$5,0 \mu\text{m} - 1 \text{ mm}$
Micro-ondas	$1 \text{ mm} - 100 \text{ cm}$
Rádio	$1 \text{ m} - 10 \text{ km}$

Obs: $\text{A}^\circ = 10^{-10} \text{ m}$; $\text{nm} = 10^{-9} \text{ m}$; $\mu\text{m} = 10^{-6} \text{ m}$

Fonte: ESA, 2022 (adaptado).

O sensor é o instrumento capaz de mensurar a radiação eletromagnética proveniente de um objeto, transformá-las em um sinal elétrico e registrá-las de tal forma que essa possa ser armazenada ou transmitida (MORAES, 2002). As

variações de energia eletromagnética da área observada podem ser coletadas por sistemas imageadores ou não-imageadores (MORAES, 2002).

Entende-se por um sistema imageador aquele que obtém a imagem observada através de *scanners*, câmeras fotográficas e demais instrumentos dessa categoria, ao passo que sistemas não-imageadores obtém a imagem por radiômetros ou espectroradiômetros (MORAES, 2002).

Sensores podem ser classificados como ativos e passivos, de modo que, a principal diferença entre eles é que os sensores passivos não possuem fonte própria de energia eletromagnética ao passo que os ativos sim (MORAES, 2002). A principal vantagem no sensor ativo é o grande comprimento de ondas produzido pelos radares que não é barrado ou absorvido pelas micrométricas partículas ou gases da atmosfera e, dessa forma, pode ser operado sob diversas condições atmosféricas (FIGUEIREDO, 2005; MENEZES et al., 2012).

Os sensores passivos dependem da variação na condição de iluminação solar (ângulos de elevação e azimute solar), como aqueles embarcados no CBERS, Landsat TM e Sentinel 2. Na próxima seção são apresentados detalhes do o Satélite Sentinel-2 (S-2), cujos produtos serão utilizados nesta dissertação.

3.2. SATÉLITE SENTINEL-2

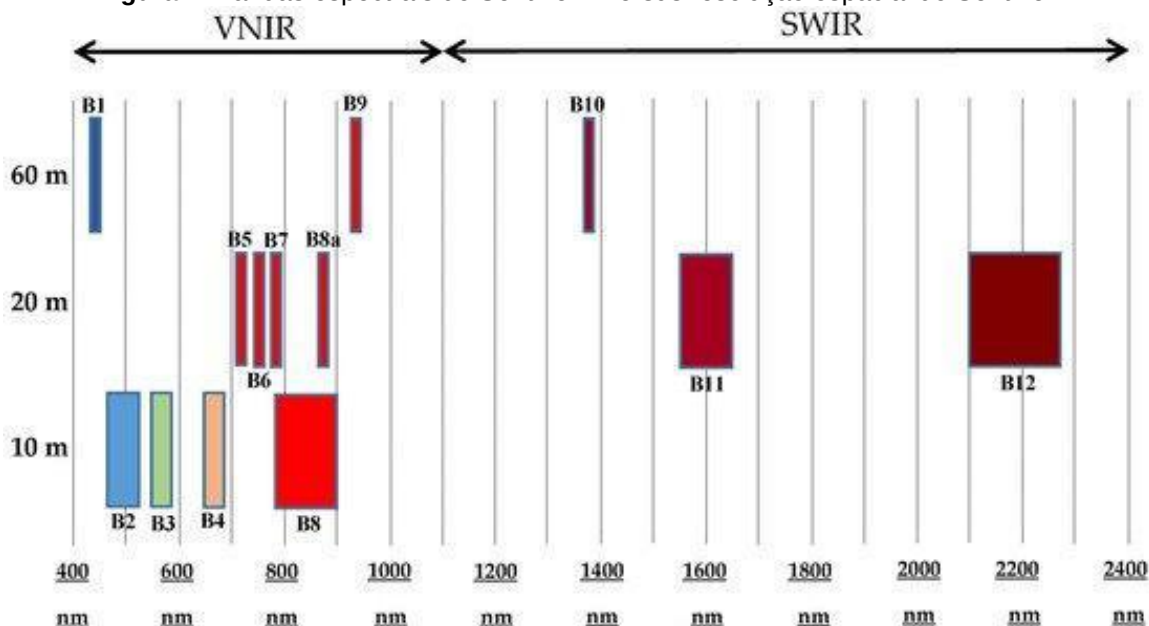
Sob jurisdição da Agência Espacial Europeia (sigla em inglês, ESA), o programa Copernicus de observação da Terra possui as missões Sentinel (1, 2, 3 e 5) dos quais o Sentinel-2 resulta em maior significância devido a presença do instrumento multiespectral (PHIRI et al., 2020).

O S-2 é composto por dois satélites denominados gêmeos (Sentinel-2A e Sentinel-2B), sendo o sistema de aquisição das bandas apresentado na Figura 2. A principal característica de uso deste satélite é a utilização das bandas de Infravermelho de borda que possuem comprimento de onda de 705 nm, 740 nm, 783 nm e 865 nm, referente às denominadas bandas B05, B06, B07 e B8A, respectivamente.

Ainda na Figura 2, pode-se observar as bandas de Infravermelho de Ondas Curtas (SWIR) com resolução espacial de 20m, que são ideais para a categorizar materiais artificiais e naturais, assim como penetrar névoa. Três bandas de resolução espacial de 60m são aplicadas em triagem de nuvens e correções atmosféricas, sendo a banda de 443 nm para aerossóis, 945 nm para vapor de

água e 1375 nm para detecção de cirros (GASCONT et al., 2017). Para mais detalhes acessar (<https://sentinel.esa.int/web/sentinel/missions/sentinel-2>).

Figura 2. Bandas espectrais do Sentinel 2 versus resolução espacial do Sentinel 2.



Fonte: GASCON et al. (2017).

3.3. INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA

A Inteligência Artificial (IA) é um campo multidisciplinar formada por uma enorme variedade de áreas, tais como: lógica matemática, ciência da computação, linguística, filosofia, processamento de linguagem natural, planejamento, sistemas nebulosos, visão, aprendizado de máquina, redes neurais, robótica, sistema baseado em conhecimento, psicologia e biologia. Do geral (aprendizagem e percepção), até tarefas específicas como demonstração de teoremas matemáticos e condução autônoma de carros, são inúmeras as aplicações (RUSSEL; NORVIG, 2013; GOMES, 2013). Segundo Russel e Norvig (2013), trata-se verdadeiramente de um campo universal.

Dentro da IA, o Aprendizado de Máquina (AM) é a área que se dedica a implementar algoritmos que permitam a análise de dados de forma automática, sendo as melhorias e as técnicas para construí-lo derivadas do conhecimento prévio, da representação dos dados e do *feedback* da aprendizagem (CHO, 2020; WEI et al., 2019; RUSSEL; NORVIG 2013).

De acordo com Russel e Norvig (2013) existem três tipos de *feedback* que determinam os três tipos de aprendizagem:

Na aprendizagem não supervisionada, o agente aprende padrões na entrada, embora não seja fornecido nenhum feedback explícito; em aprendizagem por reforço, o agente aprende a partir de uma série de reforços – recompensas ou punições; na aprendizagem supervisionada, o agente observa alguns exemplos de pares de entrada e saída, e aprende uma função que faz o mapeamento de entrada na saída (RUSSEL; NORVIG, 2013).

As técnicas de AM empregam um princípio de inferência denominado indução, no qual se obtém conclusões genéricas a partir de um conjunto particular de exemplos sendo o aprendizado indutivo (LORENA; CARVALHO; 2007).

Os recursos e ferramentas do AM atualmente estão amplamente baseados em software livre, além da disponibilidade de acesos e aperfeiçoamento de modelos (WEI et al., 2019). O AM se baseia em programas que melhoram com a experiência utilizando medidas de desempenho, o que permite realizar tarefas cognitivas, tal como detecção de objetos ou tradução de linguagem natural (JANIESCH; ZSCHECH; HEINRICH; 2021).

As aplicações da AM nas áreas relacionadas à alta dimensão de dados como classificação, regressão e *clustering*, apresentam bons resultados relativos à decisão confiável e reproduzível tendo sido utilizado em detecções de fraudes, pontuação de crédito, análise de melhor oferta, reconhecimento de voz e imagem (JANIESCH; ZSCHECH; HEINRICH; 2021).

As técnicas de AM empregam um princípio de inferência denominado indução, no qual se obtém conclusões genéricas a partir de um conjunto particular de exemplos (LORENA; CARVALHO; 2007).

A aprendizagem supervisionada relaciona uma saída com uma entrada com base em dados rotulados. Esses dados são “alimentados” pelo usuário com pares de entrada e saída conhecidos, de modo que a cada saída é atribuído um rótulo que pode ser um valor numérico ou uma classe.

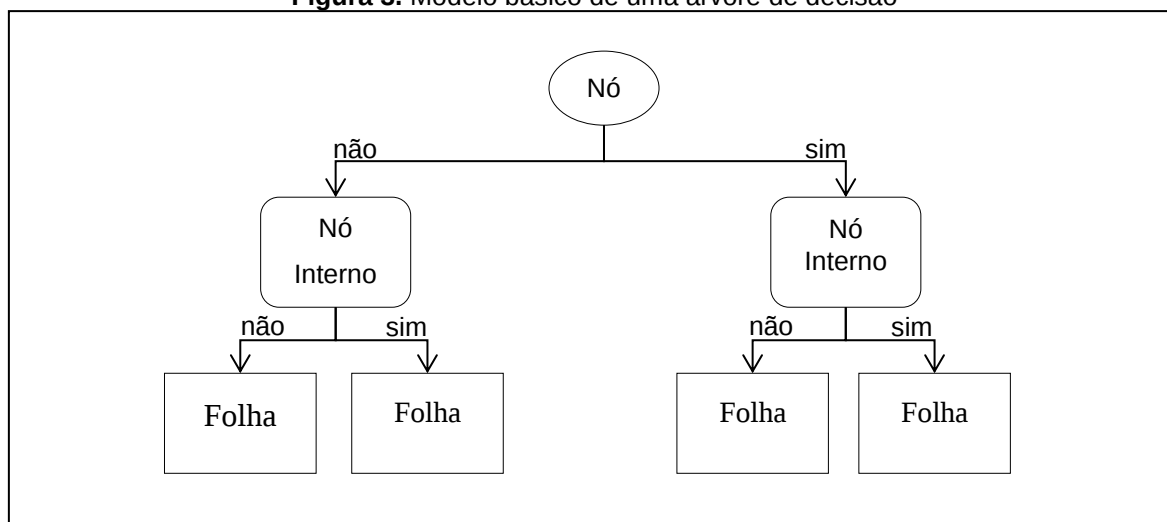
Dentre os modelos de aprendizagem supervisionados, destacam-se a árvore de decisão e Random Forrest (RF), Supporting Vector Machine (SVM) e as redes neurais artificiais. Nas próximas seções são apresentadas revisões dos modelos de aprendizado de máquina utilizados na dissertação.

3.4.ÁRVORE DE DECISÃO E RANDOM FOREST

Árvores de decisão são métodos de aprendizado de máquina para construir modelos de previsão a partir do particionamento recursivos dos dados em subconjuntos, de acordo com avaliações sobre os descritores utilizados. Este

particionamento pode ser representado graficamente como um grafo, como mostrado na Figura 3 (LOH, 2011), sendo que a partição de início recebe o nome de nó, que divide o conjunto de entrada em nós internos a partir da avaliação do ganho de informação medido através de uma função, como, por exemplo, a entropia e no final têm-se as folhas que estão associadas a uma classe (RUSSEL; NORVIG, 2013).

Figura 3. Modelo básico de uma árvore de decisão



Fonte: o Autor

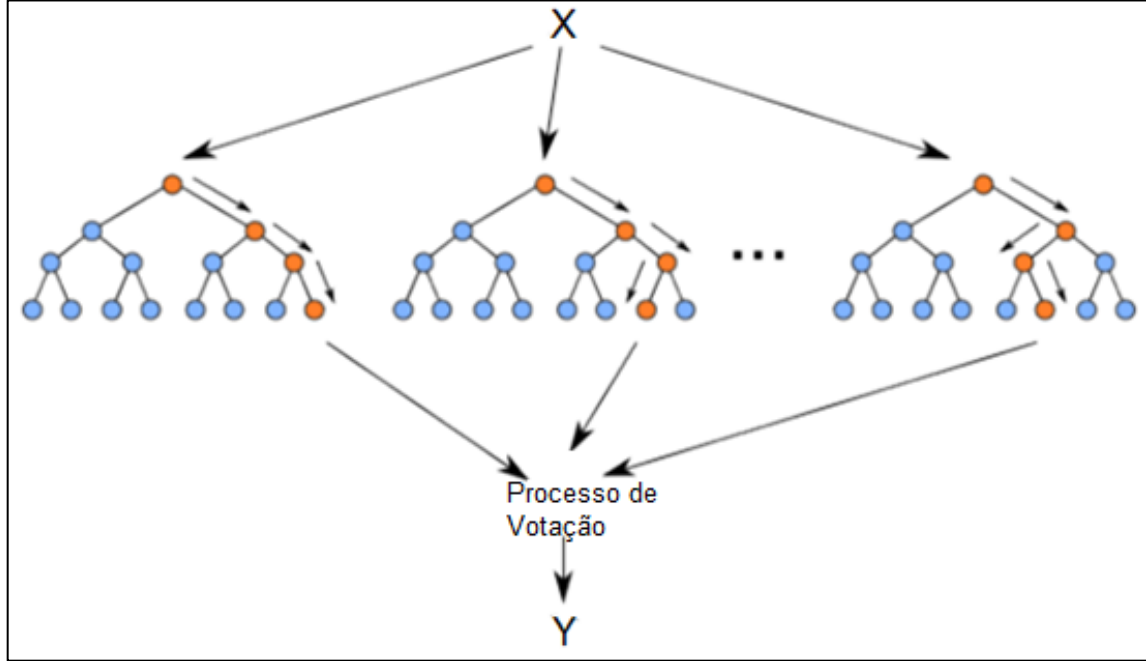
A partir da abordagem de árvore de decisão foi proposta uma extensão conhecida por *Random Forest* (RF) que é um mecanismo que agrega de forma aleatória várias árvores com versatilidade para lidar com tarefas de classificação e regressão supervisionada (BIAU; SCORNET, 2016).

A RF consiste na combinação de diferentes componentes dentre os quais incidem critérios de divisão das árvores (SCORNET, BIAU; VERT, 2015). O *Bagging* (*bootstrap-aggregating*) representa um esquema geral de agregação que procede de submostras (processo de seleção das variáveis do conjunto de dados original) de modo a utilizar amostragem com substituições para tornar o procedimento completamente aleatório (SCORNET, BIAU; VERT, 2015; BIAU; SCORNET, 2016).

De acordo com Lorenzett e Telocken (2016) a RF possui característica de “dividir para conquistar”. Para os autores, a técnica apresenta um algoritmo mais poderoso do que comparado somente a uma árvore de decisão, evitam sobre ajuste e são menos sensíveis a ruídos. A Figura 4 apresenta o funcionamento do método de classificação RF, em que o elemento X (base de dados) gera diversas árvores com suas respectivas regras e novos padrões para a tomada de decisão.

Dessa forma, é realizada a votação para cada resultado predito, sendo considerado como o resultado da predição Y , aquele com mais saídas semelhantes.

Figura 4. N-árvores para escolha da decisão final através de indicação (voto)



Fonte: Lorenzetti e Telöcken (2016), adaptado.

Na RF, além do número de árvores, os parâmetros mais importantes são o tamanho do subconjunto aleatório de variáveis consideradas do decorrer da divisão e o tamanho das árvores (NEBRINI, KÖNIG; WRIGHT, 2018).

No critério de divisão de árvores de classificação, o índice de Gini é comumente utilizado. Dessa forma, a RF realiza a seleção de características usando subconjuntos de “variáveis mais importantes” para a classificação e este resultado é avaliado pela importância, considerando o índice de Gini, que também pode ser usado como indicador geral de relevância de características (MENZE et al., 2009; WRIGHT; ZIEGLER, 2017).

O critério de Gini utiliza um índice de dispersão que considera c classes, sendo o $gini_{index}$ em um nó definido por:

$$gini_{index}(\text{nó}) = 1 - \sum_{i=1}^c p\left(\frac{i}{\text{nó}}\right) \quad (1)$$

em que $p(i/\text{nó})$ é a proporção da classe i no nó (Filho, 2014).

A medida Gini é obtida pela diferença entre $gini_{index}$ antes e após a divisão da seguinte forma:

$$Gini = gini_{index}(pai) - \sum_{j=1}^n \left[\frac{N(v_j)}{N} gini_{index}(v_j) \right] \quad (2)$$

em que n se refere ao número de nós-filhos, N é o número total de observações do nó-pai e $N(v_j)$ é o número de observações associados ao nó-filho v_j .

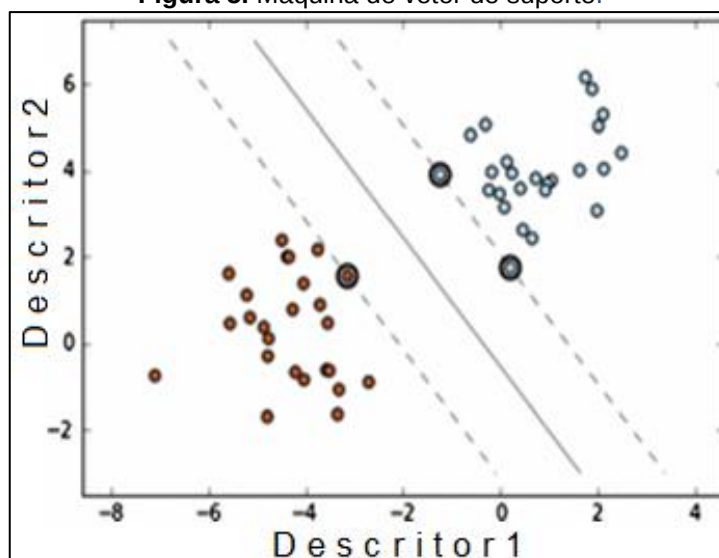
No processo de crescimento da árvore, as variáveis podem ser utilizadas diversas vezes e quando a melhor divisão é encontrada, o processo é repetido para cada nó filho até que a divisão não ocorra mais, chegando a decisão, representada pela folha (Filho, 2014).

3.5. MÁQUINA DE VETOR SUPORTE

A *Support Vector Machine* (SVM) é um modelo de aprendizado de máquina com abordagem supervisionada que constrói um separador de margem máxima, criando uma separação linear em hiperplano (RUSSEL; NORVIG, 2013)

Dessa forma, SVM's constituem uma técnica voltada para classificações (LORENA; CARVALHO; 2007). Conforme pode ser observado na Figura 5, o SVM separa as classes com margem máxima (linha contínua) entre as instâncias de linha de fronteiras (conhecidas como Vetores de Suporte) (CHAUHAN, DAHIYA; SHARMA; 2018).

Figura 5. Máquina de vetor de suporte.

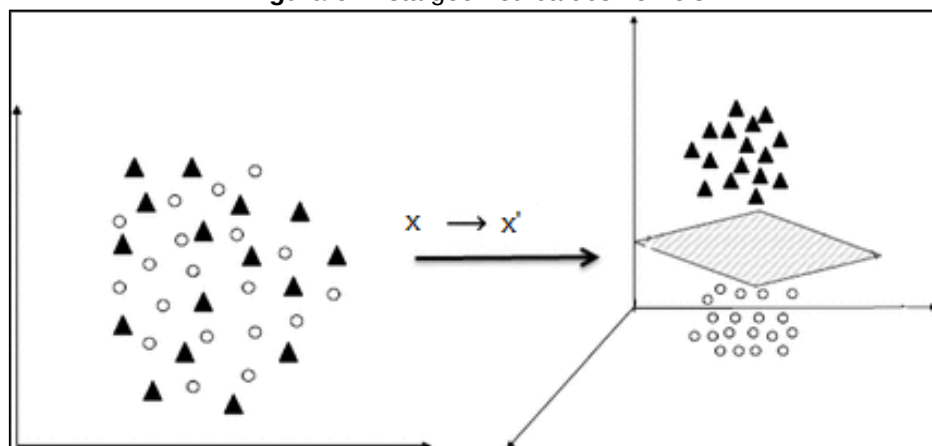


Fonte: CHAUHAN, DAHIYA; SHARMA; 2018.

De acordo com Chauhan, Dahiya e Sharma (2018), a aplicação do SVM estende-se para problemas não linearmente separáveis utilizando Funções de Kernel que transformam dados de um determinado espaço (denominado *Input Space*) para um novo espaço de alta dimensão (*Feature Space*) onde os dados são separáveis com uma superfície linear (denominado de Hiperplano). Na Figura

6, a função de kernel é aplicada na estrutura dos dados (x) transformando o espaço de entrada bi-dimensional em um espaço tridimensional (x'), permitindo que os dados sejam separados pelo hiperplano.

Figura 6. Vista geométrica dos kernels



Fonte: RUSSEL; NORVIG, 2013, adaptado.

As SVM's estão entre as principais representantes de métodos de aprendizado de máquina baseado em maximização de margens (FACELI et al., 2022). No estudo sobre classificação de padrões lineares e não lineares baseadas em SVM, GHOSH et al. (2019) apresentam as principais características de variações deste modelo que estão resumidos, junto as informações em (FACELI et al., 2022), na Tabela 2.

Tabela 2. Características da SVM, equação e descrição para cada modelo de aprendizado de máquina

SVM	Equação	Descrição
Linear	$h(x) = W \cdot X + b$	São fornecidos n conjuntos de dados de treinamento X , W é o vetor normal ao hiperplano e pertence X e, $\frac{b}{\ W\ }$ é a distância do hiperplano em relação à origem.
Radial	$\exp(-\sigma \ X_i - X_j\ ^2)$	Toma como base a distância euclidiana entre dois pontos de referência com um parâmetro livre γ , sendo $\sigma = \frac{1}{2\gamma^2}$
Polinomial	$(\delta(X_i \cdot X_j) + K)^d$	K (não é negativo) é o produto interno do espaço de recursos com base em um mapeamento, δ é um parâmetro que visa reduzir a diferença entre as ordens superiores e as ordens inferiores do polinômio e d o grau do polinômio.

Fonte: Faceli et al., 2022; Ghosh et al., 2019 (adaptado)

A escolha de um classificador por meio do SVM envolve a escolha de uma função Kernel e os parâmetros da função (FACELLI et al., 2022). A escolha do kernel implica no desempenho do classificador obtido, pois são eles que definem a fronteira de decisão (FACELLI et al., 2022; GHOSH et al., 2019).

3.6. MÉTRICAS DE VALIDAÇÃO

Dentre as métricas de validação estão: a acurácia do usuário (AU), acurácia do produtor (AP) e acurácia global (AG), todas derivadas da matriz de confusão (LUCA et al., 2022; PRASAD et al., 2022; RANA e SURYANARAYANA, 2020; STEINHAUSEN et al., 2018). A Tabela 3 apresenta um exemplo genérico de uma matriz de confusão com a guia de referência e a guia classificada (PANTALEÃO; SCOFIELD; 2009, apud CONGALTON, 1991).

Tabela 3. Matriz de confusão com elementos p_{ij} representando a proporção da categoria i da classificação de uma determinada classe e da categoria j da classe de referência.

Classificada	Referência				
	1	2	...	n	Total
1	p_{11}	p_{12}	...	p_{1n}	p_{1+}
2	p_{21}	p_{22}	...	p_{2n}	p_{2+}
...	...	...	...	...	...
n	p_{n1}	p_{n2}	...	p_{nn}	p_{n+}
Total	p_{+1}	p_{+2}	...	p_{+n}	

Fonte: PANTALEÃO; SCOFIELD; 2009, adaptado.

A Tabela 4 apresenta as métricas de AG, AU e AP calculadas a partir da matriz de confusão, sendo AG a medida calculada pela soma total dos valores corretamente classificados dividida pelo número total de valores da matriz de confusão (valores da diagonal principal), enquanto AP é a probabilidade de um valor de referência ter sido corretamente classificado e AU é a probabilidade de um valor classificado representar a categoria (PANTALEÃO; SCOFIELD; 2009, apud CONGALTON, 1991).

Tabela 4. Medidas de acurácias por categoria em que i representa a cobertura do solo da área mapeada e j representa a cobertura do solo de referência

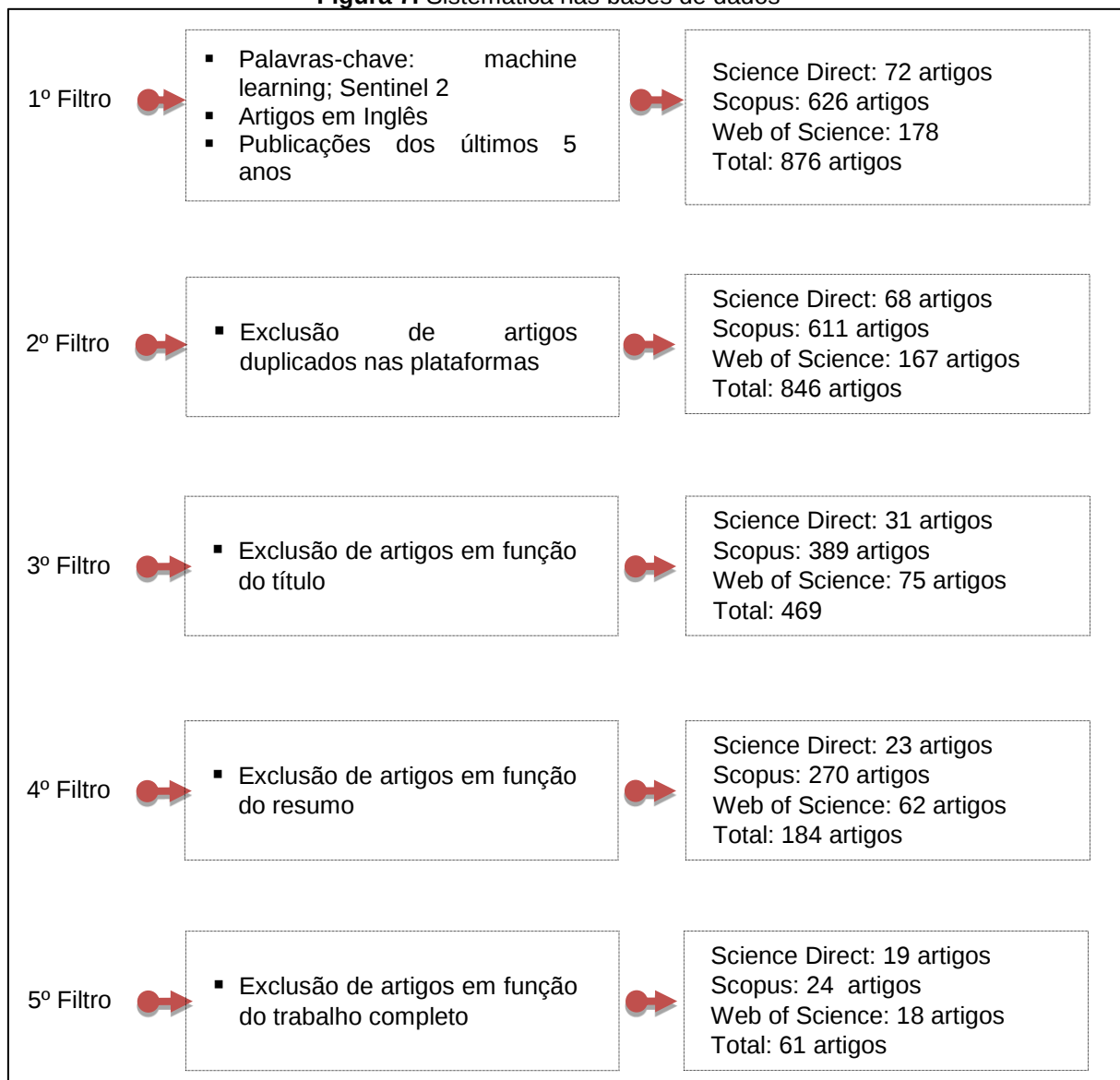
Nome	Fórmula
Acurácia Global	$AG = \sum_{i=1}^m p_{ii}$
Acurácia Produtor	$AP = \frac{p_{ii}}{p_{+i}}$
Acurácia do Usuário	$AU = \frac{p_{ii}}{p_{i+}}$

Fonte: PANTALEÃO; SCOFIELD; 2009

3.7. PESQUISA ORDENADA: IA E SR

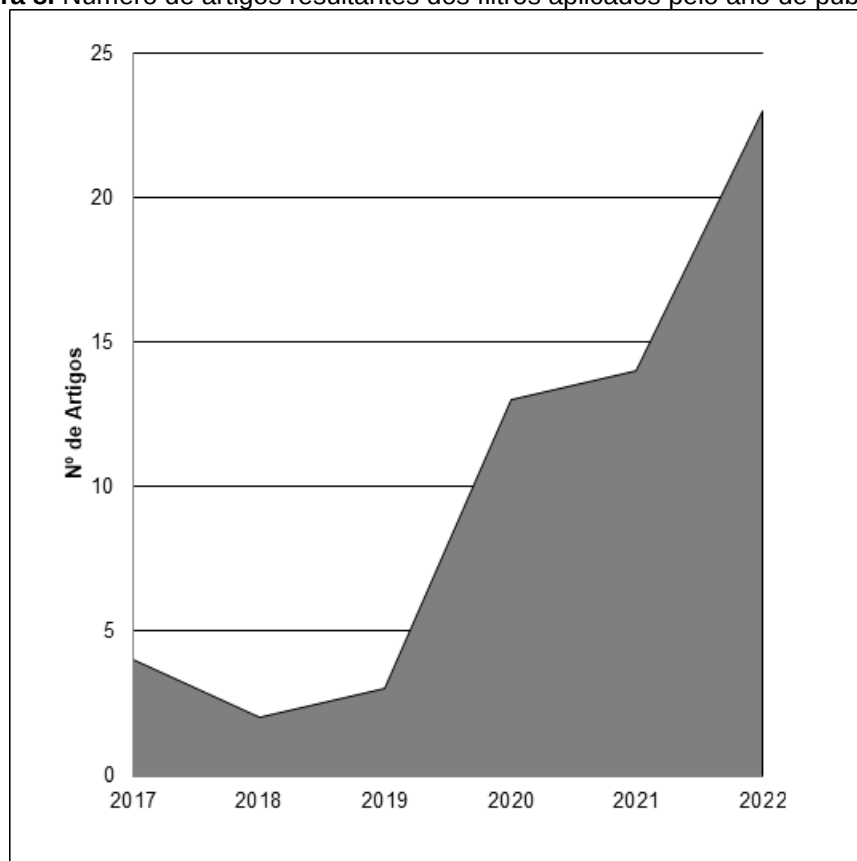
Estudos na área de IA e SR advêm de diferentes áreas das ciências, bem como o tipo de aplicação de estudo.

De modo a buscar localizar esta dissertação, foi realizada uma pesquisa ordenada. Os dados dos artigos encontrados em três bases (*Science Direct*, *Scopus* e *Web of Science*) foram tabulados e selecionados por meio de critérios de exclusões proposto por França et al. (2018). Na Figura 7 constam os critérios de exclusão e a sequência de avaliações realizadas considerando as similaridades com o presente trabalho.

Figura 7. Sistemática nas bases de dados

Fonte: o Autor

A exclusão multicritério baseada em França et al. (2018) reduziu de 876 artigos indexados para 61 que estão correlacionados com o assunto da dissertação. A Figura 8 apresenta a evolução temporal dessas publicações dos últimos 5 anos, evidenciando a tendência da utilização de métodos supervisionados junto ao sensoriamento remoto.

Figura 8. Número de artigos resultantes dos filtros aplicados pelo ano de publicação

Fonte: o Autor

Na próxima seção, considerações dos artigos filtrados serão detalhadas para apresentar aspectos encontrados pelos autores que utilizam da inteligência artificial e sensoriamento remoto para a classificação da cobertura do solo.

3.7.1. Inteligência Artificial e sua aplicação ao Sensoriamento Remoto

As técnicas e ferramentas do SR são muito eficientes, principalmente com o advento de soluções de fácil acesso e código aberto (LUCA et al., 2022; RAHMAN et al., 2020). O progresso na tecnologia do SR e a observação da Terra destacam a importância dos métodos automatizados/semiautomatizados (NAJAFI et al., 2021). Esse progresso no SR possibilitou a classificação de cenas usando imagens de satélite e tem sido responsável no desenvolvimento de novas abordagens, especialmente em métodos de aprendizado de máquina (RAIYANI et al., 2021).

Estudos que comparam algoritmos de aprendizado de máquina devem fazê-lo de maneira consistente para não introduzir viés (ABDI, 2019). O uso do S-2 em análises temporais ou estágios sucessionais diferem e podem passar por correções geométricas e/ou radiométricas e, em boa parte, atmosférica,

principalmente no nível 1C, que não apresenta correção alguma (LUCA et al., 2022; DUPUY et al., 2020; ABUDI, 2019). Não apenas as correções atmosféricas, mas a reamostragem de outras bandas, principalmente as de 20m, para imagens de resolução de 10 m são comumente realizadas aplicando o algoritmo do vizinho mais próximo tendo em vista a qualidade das bandas de infravermelho de ondas curtas e SWIR para múltiplos usos (PRASAD et al., 2022; EBRAHIMY et al., 2021; DUPUY et al., 2020; NGUYEN et al., 2020; ABUDI, 2019; STEINHAUSEN et al., 2018; FORKUOR et al., 2017). Outra forma de reamostragem é a bilinear (DABIJA et al., 2021; ZHANG et al., 2021; LENCO et al., 2019; TAVARES et al., 2019).

Uma alternativa de reamostragem é a utilização de buffer baseado em distância dos pixels para minimizar o impacto negativo da borda nos padrões espectrais de cobertura do solo (Dabija et al., 2021) e as regras baseadas na abundância de cobertura da terra (CLARK, 2017).

A reamostragem permite que todas as bandas tenham mesma resolução espacial de modo a possibilitar o uso de índices espectrais como entrada de dados nos algoritmos de aprendizagem de máquina (MALINOWSKI et al., 2020; RAHMAN et al., 2020; PRASAD et al., 2022).

O uso de índices espectrais para o aprendizado de máquina é abordado em diferentes cenários. Os índices de GNDVI (Vegetação de Diferença Normalizada Verde), MNDWI (Índice modificado da Diferença Normalizada da Água) e NDBI (Índice de Acumulação de Diferença Normalizada) são utilizados pelos autores Prasad et al. (2022) com o intuito de mapeamento da cobertura do solo na costa centro-oeste da Índia.

O uso de RF e SVM mostra alta precisão devido ao produto com alta resolução espacial e espectral no estudo dos autores Prasad et al. (2022), para os quais, o SVM oferece melhor precisão na produção de mapas de uso e cobertura da terra, com resultados de 95,82%. Igualmente, Vasilakos et al. (2020) mostram que os resultados da SVM apresentam superioridade ao modelo RF, tendo uma acurácia de 93%.

Na avaliação de Prasad et al. (2022) entre as treze variáveis de imagens do S-2, o SWIR-2, Banda Vermelha, NDBI foram as mais importantes para a classificação, seguidas por SWIR-1, Infravermelho médio (B05), NIR, GNDVI,

MNDWI, Infravermelho médio (B07), Infravermelho médio (B06), Banda verde e Banda azul.

A depender da natureza do estudo, mais índices podem ser utilizados. O NDVI (Índice de Vegetação com diferença Normalizada), NBR (Taxa de Queima Normalizada) e NDRE (Índice de Borda Vermelha de Diferença Normalizada) são utilizadas por Luca et al. (2022) com uma RF que realiza a classificação supervisionada de uma região. Para os autores, a arquitetura da RF registrou 97,0% na classificação da área de estudo e, de acordo com os mesmos, os dados de treinamento para a classificação do uso da terra podem ser baseados em imagens de alta resolução em vez da observação em campo, sendo mais econômico e tendo representatividade adequada de uma gama de características ambientais presentes em grandes porções de territórios.

WAŚNIEWSKI et al. (2022) realizam uma classificação por meio da RF em uma área localizada no centro da Polônia junto com a província de Łódź,. No estudo, o maior valor de AU foi obtido para coníferas florestais (94% a 99%) e herbáceas periódicas (94% a 96%), enquanto que AP variou de 92% a 98% e de 90% a 94%, respectivamente. O valor da AG para todas as classes de cobertura da terra foi superior a 92%.

Áreas com vegetação utilizam não apenas o NDVI ou GNDVI, mas também o EVI (Índice de Vegetação Aprimorado) que agrega peso aos estudos, como na análise realizada por Xie e Niculescu (2021) do monitoramento dos tipos e taxa de mudanças plurianuais da cobertura e uso do solo, bem como as consequências dessas mudanças ocorridas na Península de Crozon. Os índices espectrais NDVI, GNDVI e EVI junto de classificadores SVM e RF obtiveram resultados satisfatórios no estudo.

Os índices de vegetação apresentam inúmeras aplicações em variados cenários de análise. Abdi (2019) avaliou paisagens boreais, na determinação da cobertura vegetal através de SVM e RF com auxílio de NDVI, MNDWI e NDVI.

São nítidas as combinações dos índices espectrais em séries temporais com vista ao mapeamento de cobertura de terra (DUPUY, GAETANO e MÉZO, 2020; DUPUY et al., 2020; RAHMAN et al., 2020, LENCO et al., 2019).

Em um estudo nacional, Tavares et al. (2019) utilizaram dados do S-2 para a classificação e cobertura do solo na área de Belém, na Amazônia Oriental

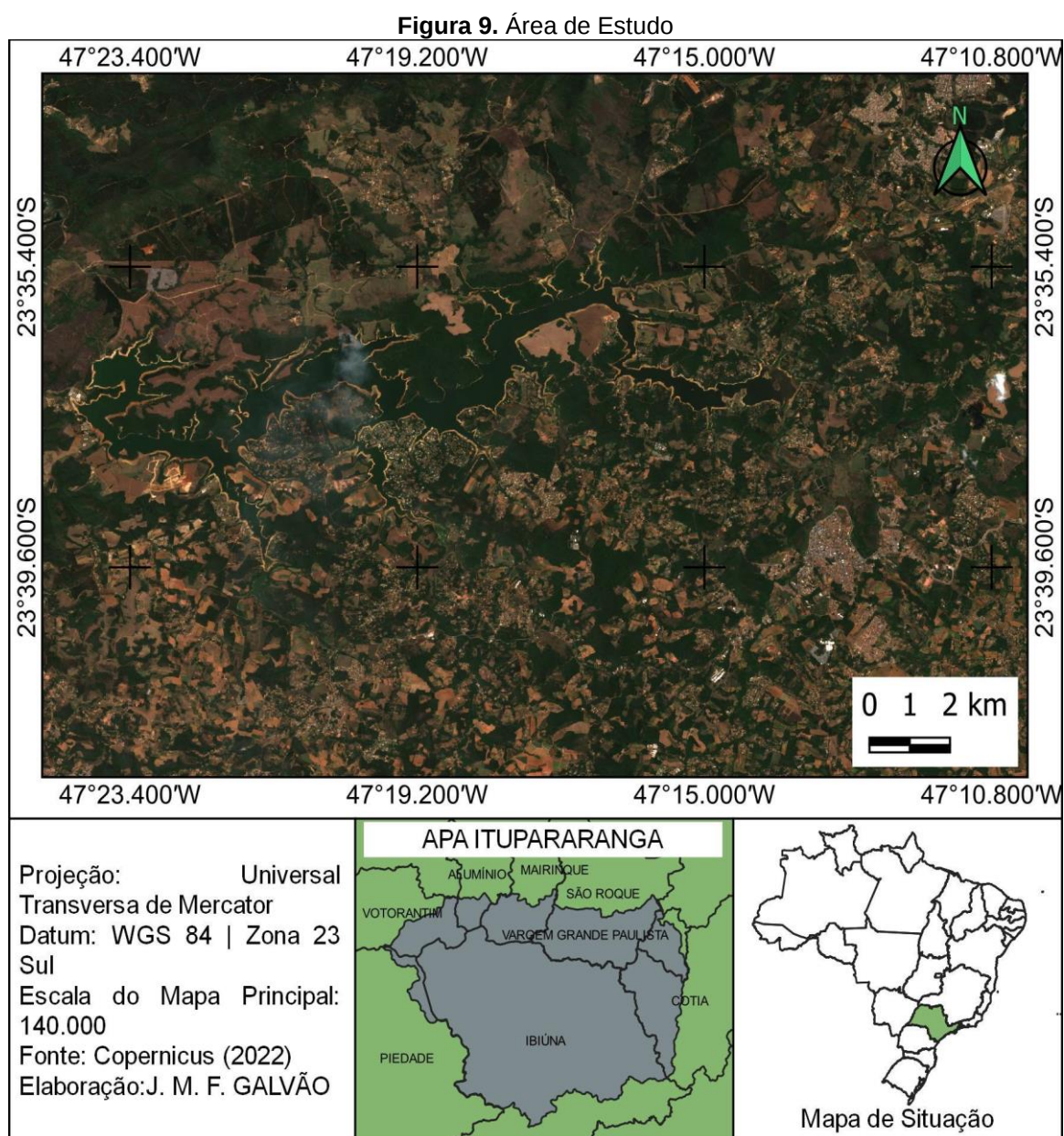
Brasileira. Este estudo aborda dois índices espectrais NDVI e NDWI com uma RF para realizar a classificação da região e o mapeamento de US e CS.

4. METODOLOGIA

Nesta seção são abordados os tópicos referentes ao processo metodológico seguido na estrutura da dissertação.

4.1. ÁREA DE ESTUDO

A região de estudo está situada na APA Itupararanga, que abrange um total de oito municípios do estado de São Paulo: Alumínio, Cotia, Ibiúna, Mairinque, Piedade, São Roque, Vargem Grande Paulista e Votorantim (Figura 9).



Fonte: o Autor.

Construída em 1911, o reservatório de ITUPARARANGA é formado por uma barragem que represa as águas do Rio Sorocaba, no município de

Votorantim (ROSA et al., 2015), com capacidade de atender 800.000 habitantes, sendo responsável por atender as demandas dos municípios de Sorocaba, Votorantim, Mairinque, Alumínio, Ibiúna e São Roque (HARKOT, 2019).

4.2. AQUISIÇÃO DA IMAGEM E PRÉ-PROCESSAMENTO

A plataforma de dados de imagens de satélites Copernicus (<https://scihub.copernicus.eu/dhus/#/home>) conta com interface de fácil utilização para a aquisição de dados. Com tudo, os produtos são identificados por meio de faixa de dados folhas com metadados para análises.

Dessa forma, a análise que este estudo realizou conta como um filtro aplicado com limite de até 10% de cobertura de nuvens para a região da represa Itupararanga – SP. O nível de produto é o 2A, que conta com o processamento para a correção atmosférica por meio da plataforma Copernicus. A Tabela 5 contém as características do produto utilizado nesta pesquisa.

Tabela 5. Informações do Produto adquirido

Descrição	Dados
Cobertura de nuvens (%)	0,004233
Sombra de nuvem (%)	0,0
Recursos escuros (%)	0,204019
Órbita relativa	38
Nível de processamento	Nível-2A
Tipo de Produto	S2MSI2A
Tamanho do produto	1,04 GB

Fonte: ESA, 2022 (adaptado).

4.3. INDICES ESPECTRAIS

A utilidade dos índices espectrais depende do objetivo da pesquisa. No desenvolvimento em questão, há forte presença de vegetação, corpo hídrico e área construída. Por essas características, e por meio da revisão da literatura, seis índices espectrais foram utilizados.

Esses índices espectrais são amplamente aplicados para monitoramento de mudanças na cobertura do solo, corpos d'água, estágios multi-sazonal e sucessional (XIE e NICULESCU, 2021; ABDI, 2019; CLARK, 2017).

Por conseguinte, nesta pesquisa, os índices aplicados constam organizados na Tabela 6, sendo as bandas do satélite S-2 resumidas em: Banda verde (B03), Banda vermelha (B04), Infravermelho (B08) e SWIR (B11).

Tabela 6. Índices e características abordadas no estudo

Informações	Autoria	Índice espectral	Utilidade
Índice de vegetação por Diferença normalizada – NDVI	Rouse et al. (1973)	$[B08 - B04] / [B08 + B04]$	Vegetação
Índice de Vegetação de Diferença Normalizada Verde – GNDVI	Gitelson et al. (1996)	$[B08 - B03] / [B08 + B03]$	Corpo hídrico ou solo exposto
Índice de Água de Diferença Normalizada – NDWI1	Gao (1996)	$[B08 - B11] / [B08 + B11]$	Corpo hídrico
Índice de Água de Diferença Normalizada – NDWI2	Mcfeeters (1996)	$[B03 - B08] / [B03 + B08]$	Corpo hídrico
Índice da diferença normalizada da água modificado – MNDWI	Xu (2006)	$[B03 - B11] / [B03 + B11]$	Corpo hídrico
Índice Normalizado da Diferença Construída –NDBI	Zha et al. (2003)	$[B03 - B11] / [B03 + B11]$	Áreas construídas

Fonte: o Autor

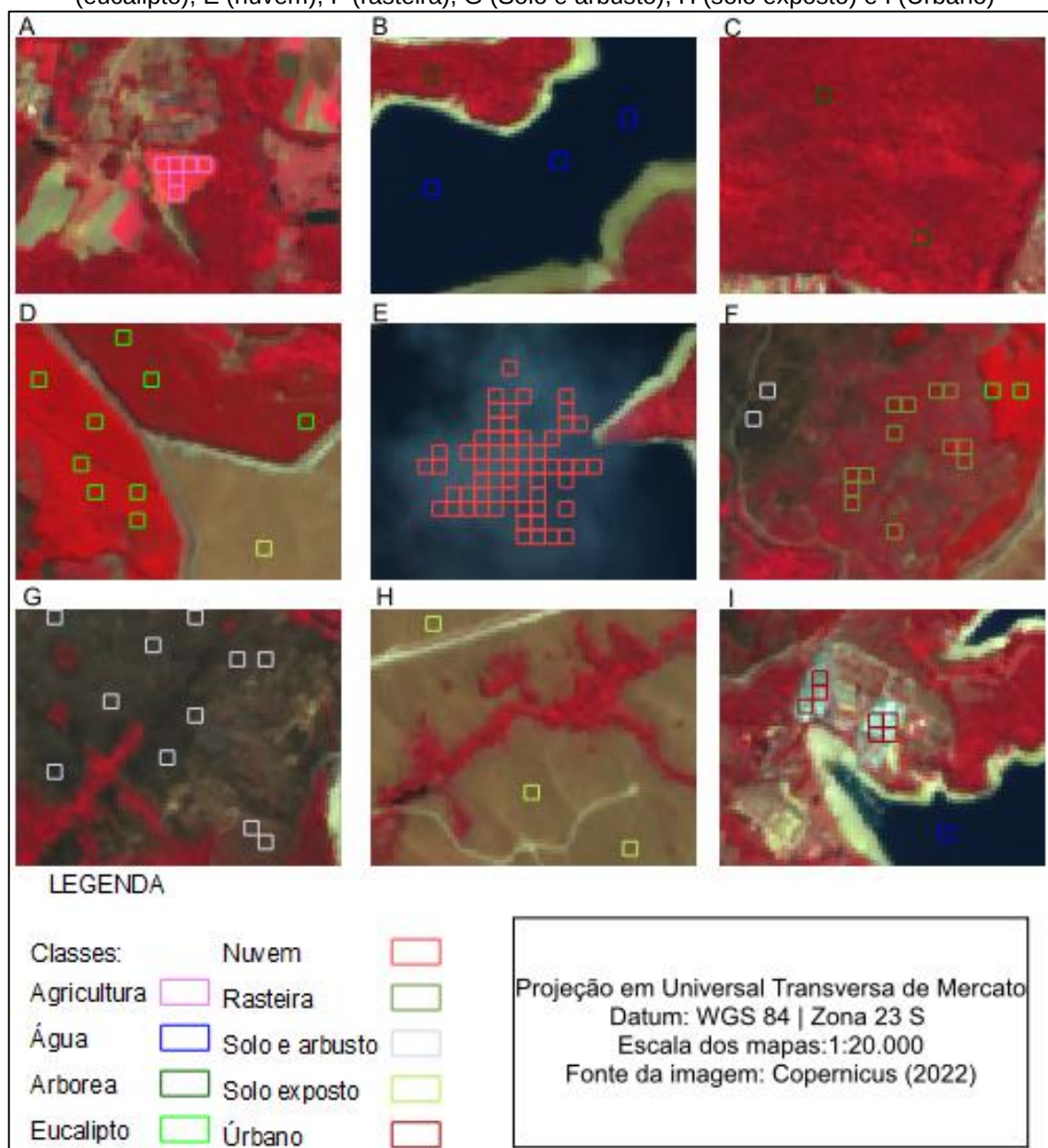
4.4. OBTENÇÃO DAS AMOSTRAS

As aquisições das amostras foram realizadas através de coleta por procedimento manual, de modo a identificar por meio da imagem de satélite as regiões de interesse (PRASAD et al., 2022; XIE e NICULESCU, 2021).

As áreas escolhidas são organizadas por meio de arquivos vetoriais em *shapefile*, sendo que esse tipo de arquivo mantém referência das coordenadas dos pontos sobre a paisagem analisada (RAHMAN et al., 2020).

Dessa forma, a organização na etapa de obtenção das amostras segue as premissas supracitadas, sendo a Figura 10 a representação do procedimento adotada para cada amostragem sob a imagem da área de estudo com um enquadramento que permita a visualização do conjunto de dados coletados.

Figura 10. Localização da obtenção das amostras, sendo A (agricultura), B (água), C (arbórea), D (eucalipto), E (nuvem), F (rasteira), G (Solo e arbusto), H (solo exposto) e I (Urbano)



Fonte: o Autor.

A coleta dos dados foi feita utilizando uma composição de falsa cor, por meio da ferramenta Google Earth, e o processo de gradeamento das amostras foi realizado por meio do software Qgis.

Todos os arquivos foram armazenados em um banco de dados, com vista a utilização posterior por meio do software de estatística Rstúdio. Também é importante ressaltar a integração que há entre ambos os softwares, porém a operação é frequentemente individual.

As características observadas na área de estudo foram definidas em 9 classes e os atributos de cada item é discriminado na Tabela 7, bem como o identificador usado em todo o procedimento, desde o Qgis ao Rstúdio.

Tabela 7. Amostras da área de estudo com o identificador, a classe e sua característica para o estudo.

Identificador	Classe	Características
1	Agricultura	Campos de usos múltiplos para agricultura
2	Água	Corpo hídrico
3	Arbórea	Vegetação densa
4	Eucalipto	Campos de plantio de eucalipto
5	Nuvem	Cobertura de nuvem
6	Rasteira	Vegetação rasteira
7	Solo e Arbusto	Interação entre vegetação rasteira arbustiva e solo exposto
8	Solo Exposto	Área de solo em exposição
9	Urbano	Área alterada com presença de construção

Fonte: o Autor

4.5.FERRAMENTAS COMPUTACIONAL E APLICAÇÃO DOS MÉTODOS

A avaliação do desempenho de diferentes algoritmos de aprendizado de máquina na classificação de imagens de satélites foi realizado através de software de código aberto, o Rstúdio. O Hardware computacional utilizado nesta pesquisa foi uma WorkStation com processador Intel (R) Xeon (R) com 16 GB de memória, velocidade de 2.53 GHz e com capacidade operacional de 64 Bits. Os métodos RF e SVM foram executados em linguagem R, através da interface Rstúdio por meio de 10 pacotes descritos na Tabela 8.

Tabela 8. Bibliotecas utilizadas no Rstudio e respectivas funções

Biblioteca	Função
Caret	Usado no processo de divisão de dados, pré-processamento, seleção de recursos, ajuste de modelo usando reamostragem e estimativa de importância de variável
e1071	Funções para a utilização de Support Vector Machine, com argumentos

	para agrupamento de dados e os tipos de classificação e regressão
randomForest	Funções para utilização da Random Forest, com argumentos para divisão de dados, tipos de classificação e regressão.
Raster	Funções para análise com arquivo raster, tal como a criação de arquivos em raster, sobreposição, conversão de vetor para raster e diversas funcionalidades, assim como a integração com o sistema de informação geográfica.
reshape2	Reestruturação e agregação de dados.
rgdal	Importação e exportação de dados raster e dados vetoriais.
rgeos	Operações de topologia em geometrias
sp	Avaliação de Classes e métodos para dados espaciais

Fonte: o Autor.

4.5.1. Treinamento dos Modelos, Ajustes e Métrica de Validação

As bandas espectrais e os índices espectrais (B02, B03, B04, B05, B06, B07, B08, B8A, B11, B12, NDVI, GNDVI, NDWI 1, NDWI 2, MNDWI, NDBI) foram utilizados como descritores. As amostras foram divididas em 70% para treino e 30% para validação. A métrica de validação utilizada no estudo é a matriz de confusão devido a maior utilização desta métrica entre os estudos que utilizam aprendizado de máquina para a classificação das imagens de satélite S-2 (EBRAHIMY et al., 2021; GHAYOUR et al., 2021; LUCA et al., 2022).

O código utilizado na dissertação pode ser conferido por meio da plataforma github através do link <https://github.com/GALVAOJMF/MESTRADO.git>.

Para a RF ocorreu uma seleção do número de árvores e o número de entradas para os valores avaliados (BELGIU; DRĂGUT, 2016), enquanto para a SVM foi feita a escolha do kernel, o custo e o gama (MOUNTRAKIS et al., 2011).

5. RESULTADOS E DISCUSSÃO

A Figura 11 contém o agrupamento dos resultados das matrizes de confusão para os modelos Random Forest, SVM linear, SVM Radial e SVM Polinomial exportados por meio do software Rstúdio.

Figura 11. Agrupamento dos Matrizes de Confusão para os modelos RF e SVM, sendo (a) RF, (b) SVM Linear, (c) SVM Radial e, (d) SVM Polinomial

(a)									
	AGRI	AGUA	ARB	EUC	NUV	RAS	SOLARB	SOLEXP	URB
AGRI	287	0	0	0	0	0	0	0	4
AGUA	0	288	0	0	0	0	0	0	0
ARB	0	0	283	1	0	4	0	0	0
EUC	0	0	5	288		0	0	0	0
NUV	0	0	0	0	289	0	0	0	0
RAS	0	0	0	0	0	285	0	1	3
SOLARB	0	0	0	0	0	0	287	0	3
SOLEXP	0	0	0	0	0	0	0	288	0
URB	0	0	0	0	0	0	1	0	279

(b)									
	AGRI	AGUA	ARB	EUC	NUV	RAS	SOLARB	SOLEXP	URB
AGRI	287	0	0	0	0	0	0	0	2
AGUA	0	288	0	0	0	0	0	0	0
ARB	0	0	285	2	0	6	0	0	0
EUC	0	0	3	287	0	0	0	0	0
NUV	0	0	0	0	289	0	0	0	0
RAS	0	0	0	0	0	282	0	2	0
SOLARB	0	0	0	0	0	1	276	4	4
SOLEXP	0	0	0	0	0	0	4	283	4
URB	0	0	0	0	0	0	8	0	279

(c)									
	AGRI	AGUA	ARB	EUC	NUV	RAS	SOLARB	SOLEXP	URB
AGRI	287	0	0	0	0	0	0	0	3
AGUA	0	288	0	0	0	0	0	0	0
ARB	0	0	287	1	0	3	0	0	0
EUC	0	0	0	288	0	0	0	0	0
NUV	0	0	0	0	289	0	0	0	0
RAS	0	0	1	0	0	285	0	1	3
SOLARB	0	0	0	0	0	0	286	0	4
SOLEXP	0	0	0	0	0	1	0	288	0
URB	0	0	0	0	0	0	2	0	279

(d)									
	AGRI	AGUA	ARB	EUC	NUV	RAS	SOLARB	SOLEXP	URB
AGRI	287	0	0	0	0	0	0	0	4
AGUA	0	288	0	0	0	0	0	0	0
ARB	0	0	282	1	0	3	0	0	0
EUC	0	0	0	288	0	0	0	0	0
NUV	0	0	0	0	289	0	0	0	0
RAS	0	0	6	0	0	285	0	3	3
SOLARB	0	0	0	0	0	0	288	0	1
SOLEXP	0	0	0	0	0	1	0	286	1
URB	0	0	0	0	0	0	0	0	280

Fonte: o Autor.

A partir das matrizes de confusão da Figura 11, foi construída a Tabela 9, onde pode ser observado que o modelo SVM com Kernel radial apresenta maior

desempenho, com AG de 99,27%, enquanto o modelo RF aparece na segunda posição com AG de 99,15%. O modelo que apresentou menor desempenho resultou em uma AG de 98,46%, sendo o SVM Linear.

Ainda analisando a Tabela 9, as classes de água e nuvem apresentam 100% de precisão entre as AP e AU, em conformidade com os estudos de Ebrahimi et al. (2021) e Dabija et al. (2021). A classe de agricultura apresenta AP de 100% nos modelos, no entanto, a AU da classe de agricultura apresenta maiores sobreposições com valores de solo exposto e vegetação rasteira. As classes arbóreas e rasteira apresentam maiores erros nas AP's e AU's. Os resultados das classes de eucalipto, solo exposto, interação solo e arbusto e área urbana apresentam confusões entre si, como pode ser observado pelos valores de suas métricas.

Tabela 9. Desempenho dos modelos de aprendizado de máquina

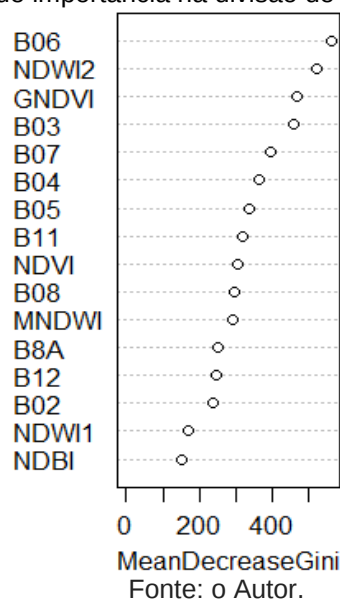
CLASSES	RF		SVM LINEAR		SVM RADIAL		SVM POLINOMIAL	
	AP	AU	AP	AU	AP	AU	AP	AU
Agricultura	100%	98,63%	100%	99,31%	100%	98,97%	100%	98,63%
Água	100%	100%	100%	100%	100%	100%	100%	100%
Arbórea	98,26%	98,26%	98,96%	97,27%	99,65%	98,63%	97,92%	98,60%
Eucalipto	99,65%	98,29%	99,31%	98,27%	99,65%	100%	99,65%	100%
Nuvem	100%	100%	100%	100%	100%	100%	100%	100%
Rasteira	98,62%	98,62%	97,58%	99,30%	98,62%	98,28%	98,62%	95,96%
Solo e Arbusto	99,65%	98,97%	95,83%	96,84%	99,31%	98,62%	100%	99,65%
Solo exposto	99,65%	100%	97,92%	97,25%	99,65%	99,65%	98,96%	99,31%
Urbano	99,54%	99,64%	96,54%	97,21%	96,54%	99,29%	96,89%	100%
AG	99,15%		98,46%		99,27%		99,11%	

Fonte: o Autor.

Na importância de Gini (Figura 12) das 16 variáveis, 31,25% da construção das árvores apresenta 4 variáveis mais relevantes na RF, são elas: infravermelho médio 06 (B06); o índice de corpos d'água (NDWI2); o índice destinado a corpo hídrico e solo exposto (GNDVI) e; a Banda 03 (B03). Ao passo que 12,5%

apresentam baixa relevância na RF: índice de corpos d'água (NDWI1) e índice para áreas construídas (NDBI).

Figura 12. Variáveis de importância na divisão de nós e folhas para a RF



A abordagem através de SVM e RF apresentou resultados satisfatórios. A boa precisão alcançada pelo conjunto de dados S-2, nesta abordagem, apresenta AG de 99,15% para RF, 98,46% para o SVM Linear, 99,27% para o SVM Radial e 99,11% para o SVM Polinomial, resultando que o SVM com kernel radial apresenta melhor desempenho para a classificação do S-2, que está de acordo com trabalhos que comparam os SVM's e RF em estudos desenvolvidos em Luca et al. (2022), WAŚNIEWSKI et al. (2022), Dabija et al. (2021), Ebrahimi et al. (2021), Ghayour et al. (2021) e Vasilakos et al. (2020).

Os índices espectrais para vegetação (NDVI) e detecção de áreas construídas (NDBI) não se mostrou tão importante para a construção das árvores aleatórias, assim como nos resultados de WAŚNIEWSKI et al. (2022), Prasad et al. (2022).

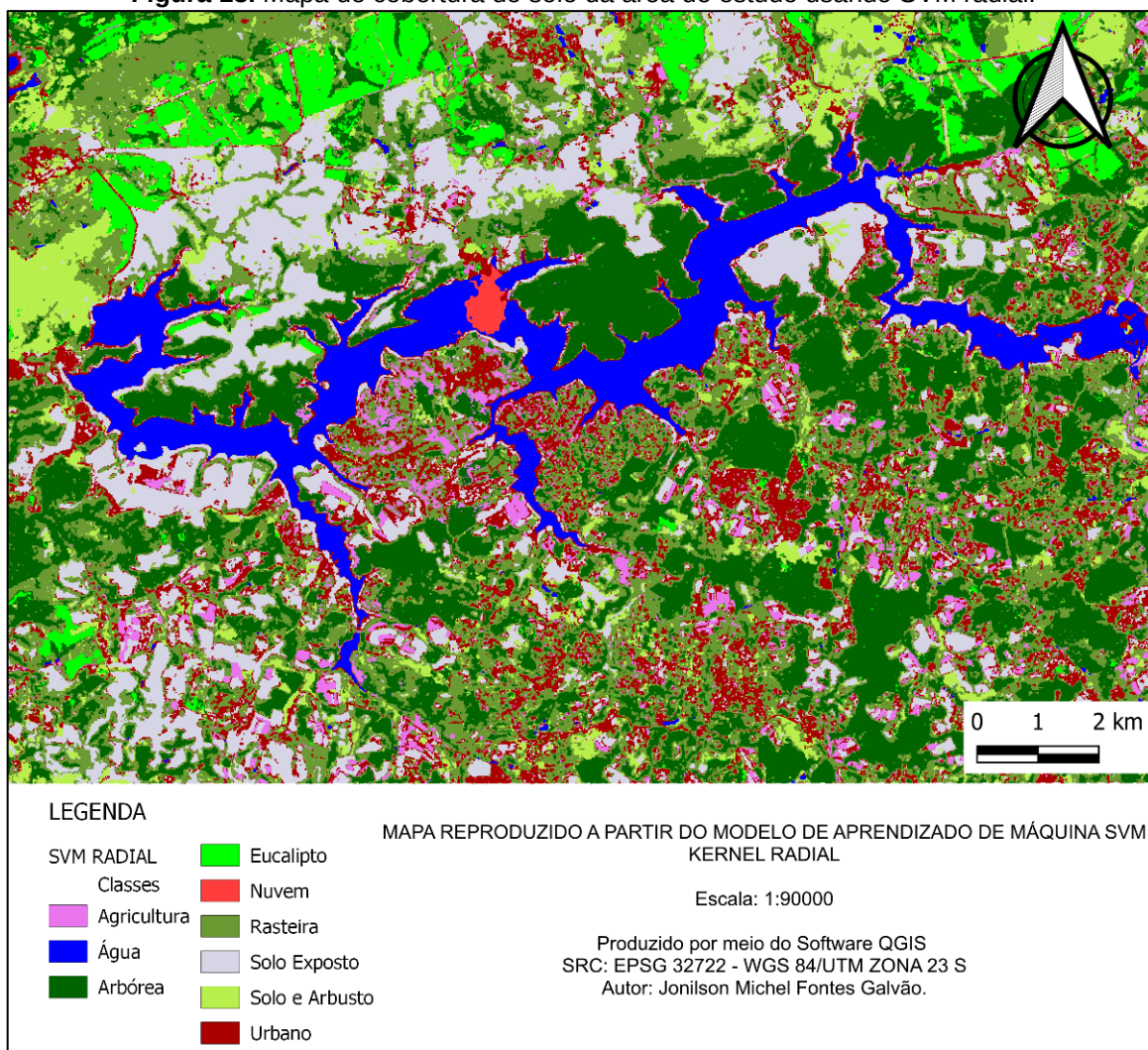
A pouca diferença observada nas acurácias de classificação obtidas pelos classificadores RF e SVM são semelhantes aos estudos relatados por Prasad et al. (2022), Dabija et al. (2021), Ebrahimi et al. (2021) e Rana & Suryanarayana (2020). Vale ressaltar que, em todos esses, o SVM apresentou superioridade em relação a RF.

Diante do analisado, a construção do mapa de cobertura do solo foi feito com base no modelo de aprendizado de máquina com melhor acurácia. Dessa

forma, a Figura 13 apresenta o resultado da aplicação do modelo treinado para a região do entorno da represa Itupararanga – SP.

Em se tratando da classificação e confecção do mapa de cobertura do solo, a alta precisão e menor confusão com o algoritmo SVM com kernel radial, produz resultados mais precisos, conforme encontrado em trabalho de Rana & Suryanarayana (2020) e Chatziantoniou et al. (2017).

Figura 13. Mapa de cobertura do solo da área de estudo usando SVM radial.



Fonte: o Autor.

O resultado apresentado na Tabela 10 mostra a porcentagem da área para todas as classes da região de estudo, com base no modelo com melhor acurácia (SVM Radial). Além disso, é importante observar que a área representa mais de 50% com área vegetação rasteira e vegetação arbórea, tendo em vista que trata-se de uma área de preservação ambiental.

Tabela 10. Hectares e porcentagem referente à classe de estudo

Identificador	Classe	Hectares	Área ocupada (%)
1	Agricultura	830.010	3,4
2	Água	1.693.960	7,0
3	Vegetação Arbórea	5.830.470	24,0
4	Eucalipto	1.025.730	4,2
5	Nuvem	47.440	0,2
6	Vegetação Rasteira	6.739.610	27,7
7	Solo Exposto	3.618.700	14,9
	Interação Solo e		
8	Arbusto	1.762.860	7,2
9	Área urbana	2.779.220	11,4
Total		24.328.000	100,0

Fonte: o Autor

6. CONSIDERAÇÕES FINAIS

Neste trabalho foram testadas duas abordagens para a classificação de regiões em imagens do satélite Sentinel-2 através de aprendizado de máquina em uma área no entorno da represa Itupararanga no Estado de São Paulo. Deve-se enfatizar que o número de trabalhos nesta linha de pesquisa apresenta crescimento significativo nos últimos 4 anos de acordo com o levantamento sistemático das bibliografias levantadas nesta dissertação.

A utilização de métodos supervisionados para a classificação das regiões em imagens do Sentinel-2 apresentou valores dentro da literatura e se mostram muito eficientes no ordenamento e planejamento territorial.

O modelo de SVM com kernel radial apresentou superioridade frente aos outros modelos, porém, vale ressaltar que a simplicidade na construção da RF é fator significativo na avaliação de custo e desempenho da máquina a ser utilizada durante a execução dos modelos.

O mapa desenvolvido por meio da aplicação do modelo SVM com kernel radial para o entorno da bacia hídrica de Itupararanga pode servir para gestores ambientais e na avaliação temporal para diversas finalidades.

Com efeito, o emprego dos métodos de inteligência artificial para classificar regiões em imagens do satélite Sentinel – 2 no entorno da represa Itupararanga mostra resultados satisfatórios.

Para trabalhos futuros, é interessante comparar os resultados obtidos com os de redes neurais convolucionais que têm demonstrado robustez na modelagem ambiental e geoprocessamento.

REFERÊNCIAS BIBLIOGRÁFICAS

ABDI, A. M. **Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data.** *GIScience & Remote Sensing*, v. 57, n. 1, p. 1-20, 22 ago. 2019.

ABIDA, Khoulood; BARBOUCHI, Meriem; BOUDABBOUS, Khaoula; TOUKABRI, Wael; SAAD, Karem; BOUSNINA, Habib; CHAHED, Thouraya Sahli. **Sentinel-2 Data for Land Use Mapping: comparing different supervised classifications in semi-arid areas.** *Agriculture*, v. 12, n. 9, p. 1429, 9 set. 2022.

AGUILERA, M. A. Z.. **Classification Of Land-Cover Through Machine Learning Algorithms For Fusion Of Sentinel-2a And PlanetScope Imagery.** 2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS), pp. 246-253, 2020.

BELGIU, M.; DRĂGUȚ, L. **Random forest in remote sensing: A review of applications and future directions.** *ISPRS J. Photogramm. Remote Sens.* v. 114, p. 24–31, 2016.

BIAU, G., SCORNET, E. **A random forest guided tour.** *TEST*, v. 25, p. 197–227, 2016.

BITTENCOURT, H. R.; CLARKE, R. T. **Use of Classification and Regression Trees (CART) to classify remotely-sensed digital images.** In: 2003 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM. Anais... Toulouse: IEEE, 2003.

BREIMAN, L., FRIEDMAN, J., STONE, C. J.; OLSHEN, R. A. **Classification and regression trees regression trees.** Wadsworth: Belmont, CA v. 14, 1984.

CHATZIANTONIOU, A.; PSOMIADIS, E.; PETROPOULOS, G. **Co-Orbital Sentinel 1 and 2 for LULC Mapping with Emphasis on Wetlands in a Mediterranean Setting Based on Machine Learning.** *Remote Sensing*, v. 9, n. 12, p. 1259, 4 dez. 2017.

CHAUHAN, N. S. **Model Evaluation Metrics in Machine Learning.** 2020. Disponível em < <https://www.kdnuggets.com/2020/05/model-evaluation-metrics-machine-learning.html> > Acesso em 27 de junho de 2022.

CHAUHAN, V. K.; DAHIYA, K.; SHARMA, A. **Problem formulations and solvers in linear SVM: a review.** *Artif Intell Ver*, v. 52, p. 803-855, 2019.

CHAUURASIA, K.; NEERAJ, B; BURLE, D; MISHRA, V. K.. **Topographical Feature Extraction Using Machine Learning Techniques from Sentinel-2A Imagery.** IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, pp. 1659-1662, 2020

CHO, S.; VASARHELYI, M. A.; SUN, T. S.; ZHANG, C. A. **Learning from Machine Learning in Accounting and Assurance.** *Journal Of Emerging Technologies In Accounting*, v. 17, n. 1, p. 1-10, 1 mar. 2020.

CLARK, M. L. **Comparison of simulated hyperspectral HypsIRI and multispectral Landsat 8 and Sentinel-2 imagery for multi-seasonal, regional land-cover mapping.** Remote Sensing Of Environment, v. 200, p. 311-325, out. 2017.

CONGALTON, R. G. **A review of assessing the accuracy of classifications of remotely sensed data.** Remote sensing of environment, v. 37, n. 1, p. 35-46, 1991.

DABIJA, A.; KLUCZEK, M.; ZAGAJEWSKI, B.; RACZKO, E.; KYCKO, M.; AL-SULTTANI, A. H.; TARDÀ, A.; PINEDA, L.; CORBERA, J.. **Comparison of Support Vector Machines and Random Forests for Corine Land Cover Mapping.** Remote Sensing, v. 13, n. 4, p. 777, 20 fev. 2021.

DEMIRKAN, D. Ç.; KOZ, A.; DÜZGÜN, H. Ş. **Hierarchical classification of Sentinel 2-a images for land use and land cover mapping and its use for the CORINE system.** Journal Of Applied Remote Sensing, v. 14, n. 02, p. 1, 27 jun. 2020.

DUPUY, S.; GAETANO, R.; MÉZO, L. L. **Mapping land cover on Reunion Island in 2017 using satellite imagery and geospatial ground data.** Data In Brief, v. 28, p. 104934, fev. 2020.

DUPUY, S.; LAURENCE, D.; RAFFAELE, G.; VALÉRIE, A.; ELOISE, R.. **Land cover maps of Antananarivo (capital of Madagascar) produced by processing multisource satellite imagery and geospatial reference data.** Data In Brief, v. 31, p. 105952, ago. 2020.

EBRAHIMY, H.; NABOUREH, A.; FEIZIZADEH, B.; ARYAL, J.; GHORBANZADEH, O. **Integration of Sentinel-1 and Sentinel-2 Data with the G-SMOTE Technique for Boosting Land Cover Classification Accuracy.** Applied Sciences, v. 11, n. 21, p. 10309, 3 nov. 2021.

ESA, European Space Agency. **Sentinel online.** Disponível em <<https://sentinel.esa.int/web/sentinel/missions/sentinel-2>>. Acesso em: 8 de fevereiro de 2022.

FACELI, K.; LORENA, A. C.; GAMA, J., ALMEIDA, T. A.; CARVA, A. C. P. L. F. **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina.** 2 ed. Rio de Janeiro: LTC, p. 304, 2022.

FIGUEREDO, D., 2005. **Conceitos básicos de sensoriamento remoto.** Conab. Disponível em < https://www.clickgeo.com.br/wp-content/uploads/2013/04/conceitos_sm.pdf > Acesso: 10 jun. 2022

FILHO, J. P. **Capacidade preditiva de Modelos Credit Scoring em inferência dos rejeitados.** Dissertação (Mestrado em Estatística) – Centro de Ciências Exatas e de Tecnologia, Universidade federal de São Carlos, São Carlos, p. 95. 2014.

FIORIN, D. V.; SCHUCJ, N. J.; MARTINS, F. R.; PEREIRA, P. **Forecast of solar energy resource by using neural network methods**. Revista Brasileira de Ensino de Física, v. 33, p. 01-20, 2011.

FONTANA, E. **Introdução aos algoritmos de aprendizagem supervisionada**. Universidade Federal do Paraná: Departamento de Engenharia Química, 2020. Disponível em: https://fontana.paginas.ufsc.br/files/2018/03/apostila_ML.pdf Acesso em: 17 ago. 2022.

FORKUOR, G.; DIMOBE, K.; SERME, I.; TONDOH, J. E. **Landsat-8 vs. Sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in burkina faso**. Giscience & Remote Sensing, v. 55, n. 3, p. 331-354, 31 ago. 2017.

FRANÇA, J.F.; OLIVEIRA, R.A.; GERENUTTI, M.; JOZALA, A.F.; GROTO, D. **High performance liquid chromatography applied in hormone contaminations detection: A scoping review in ecotoxicology**. Journal of Liquid Chromatography & Related Technologies, v. 41, n. 7, p. 377-383, 2018.

GAO, B. C. 1996. **NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space**. Remote Sensing of Environment, 58(3), 257- 266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)

GASCON, F.; BOUZINAC, C.; THÉPAUT, O.; JUNG, M.; FRANCESCONI, B.; LOUIS, J.; LONJOU, V.; LAFRANCE, B.; MASSERA, S.; GAUDEL-VACARESSE, A. **Copernicus Sentinel-2A Calibration and Products Validation Status**. Remote Sensing, v. 9, n. 6, p. 584, 10 jun. 2017.

GHAYOUR, L.; NESHAT, A.; PARYANI, S.; SHAHABI, H.; SHIRZADI, A.; CHEN, W.; AL-ANSARI, N.; GEERTSEMA, M.; AMIRI, M. P.; GHOLAMNIA, M. **Performance Evaluation of Sentinel-2 and Landsat 8 OLI Data for Land Cover/Use Classification Using a Comparison between Machine Learning Algorithms**. Remote Sensing, v. 13, n. 7, p. 1349, 1 abr. 2021.

GHOSH, S.; DASGUPTA, A.; SWETAPADMA. A. **A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification**. 2019 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, pp. 24-28, 2019.

GITELSON, A. A., KAUFMAN, Y. J., MERZLYAK, M. N.. **Use of a green channel in remote sensing of global vegetation from EOS-MODIS**. Remote Sensing of Environment, 58(3), 289-298, 1996.

GOMES, D. dos S. **Inteligência Artificial: conceitos e aplicações**. Olhar Científico. v1, n. 2, p. 234-246, 2010.

GONZALEZ, R. C.; WOODS, R. E. **Processamento digital de imagens**. Tradução Roberto Marcondes Cesar Junior, Luciano da Fontoura Costa. São Paulo: Blucher, 2000. 509 páginas

GUEDES, J. C. F.; SILVA, S. M. P. da. **Sensoriamento remoto no estudo da vegetação: Princípios físicos, sensores e métodos**. ACTA Geográfica, v. 12, n. 29, Boa Vista, 2018

HARKOT, O. A. M. G. **Conflitos socioambientais relacionados ao uso e ocupação do solo nas margens da represa Itupararanga – Ibúna, SP**. Dissertação (Mestrado Profissional em Administração, Gestão Ambiental e Sustentabilidade). Programa de Mestrado em Administração, Gestão Ambiental e Sustentabilidade, Universidade Nove de Julho. São Paulo, p. 76. 2019.

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. **Machine learning and deep learning**. Electron Markets, v. 31, p. 685–695, 2021.

JENSEN, J. R. **Sensoriamento Remoto do Ambiente: Uma perspectiva em Recursos Terrestres**. Tradução: Epiphany, J. C. N. (coordenador). São José dos Campos, SP: Parêntese Editora, 2009. 598p. (Prentice Hall Series in Geographic Information Science). Tradução de: Remote Sensing of the Environment: An Earth Resource Perspective.

LOH, Wei-Yin. **Classification and regression trees**. Wires Data Mining And Knowledge Discovery, v. 1, n. 1, p. 14-23, jan. 2011.

LOH, Wei-Yin. **Fifty Years of Classification and Regression Trees**. International Statistical Review, v. 82, n. 3, p. 329-348, 30 jun. 2014.

LORENA, A. C.; CARVALHO, A. C. P. L. F. de. **Uma Introdução às Support Vector Machines**. Revista De Informática Teórica E Aplicada, 14(2), p. 43–67, 2007. <https://doi.org/10.22456/2175-2745.5690>

LORENZETT, C. D. C; TELOCKEN, A. V. **Estudo Comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de Decisão**. Simpósio de Pesquisa e Desenvolvimento em Computação, v. 2, n. 1, 2016.

LUCA, G. de; SILVA, J. M. N.; FAZIO, S. di; MODICA, G. **Integrated use of Sentinel-1 and Sentinel-2 data and open-source machine learning algorithms for land cover mapping in a Mediterranean region**. European Journal Of Remote Sensing, v. 55, n. 1, p. 52-70, 12 jan. 2022.

MALINOWSKI, R.; LEWIŃSKI, S.; RYBICKI, M.; GROMNY, E.; JENEROWICZ, M.; KRUPIŃSKI, M.; NOWAKOWSKI, A.; WOJTKOWSKI, C.; KRUPIŃSKI, M.; KRÄTZSCHMAR, E. **Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery**. Remote Sensing, v. 12, n. 21, p. 3523, 27 out. 2020.

MCFEETERS, S. K. **The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features**. International Journal of Remote Sensing, 17(7), 1425-1432, 1996.

MENESES, P. R.; ALMEIDA, T.; SANTA ROSA, A. N. C.; SANO, E. E.; SOUZA, E. B.; BAPTISTA, G. M. M.; BRITES, R. S. 2012. **Introdução ao Processamento de Imagens e Sensoriamento Remoto**. Brasília: CNPq, 276p.

MENZE, B. H.; KELM, B. M.; MASUCH, R.; HIMMELREICH, W.; BACHERT, P.; PETRICH, W.; HAMPRECHT, F. A. **A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data.** BMC Bioinformatics 10, 213, 2009.

MORAES, E. C. **Fundamentos de Sensoriamento Remoto.** São José dos Campos –SP: INPE, 2002. Capítulo 1.

MOUNTRAKIS, G.; IM, J.; OGOLE, C. **Support vector machines in remote sensing: A review.** ISPRS J. Photogramm. Remote Sens, v. 66, p. 247–259, 2011.

MUSHTAQ, F.; MAHMOOD, K.; HAMID, M. C.; TUFAIL, R. **A Comparative Study of Support Vector Machine and Maximum Likelihood Classification to Extract Land Cover of Lahore District, Punjab, Pakistan.** Pakistan Journal Of Scientific & Industrial Research Series A: Physical Sciences, v. 64, n. 3, p. 265-274, 29 set. 2021.

NAJAFI, P.; FEIZIZADEH, B.; NAVID, H.. **A Comparative Approach of Fuzzy Object Based Image Analysis and Machine Learning Techniques Which Are Applied to Crop Residue Cover Mapping by Using Sentinel-2 Satellite and UAV Imagery.** Remote Sensing, v. 13, n. 5, p. 937, 3 mar. 2021.

NEMBRINI, S.; KÖNIG, I. R.; WRIGHT, M. N. **The revival of the Gini importance?** Bioinformatics, v. 34, n. 21, p. 3711-3718, 10 maio 2018. Oxford University Press (OUP).

NOVO, E. M. L. de M.; PONZONI, F. J. **Introdução ao Sensoriamento Remoto.** São José dos Campos: INPE, 2001.

PANTALEÃO, E; SCOFIELD, G. B. **Comparação entre medidas de acurácia de classificação para imagens do satélite ALOS.** XIV Simpósio Bras Sensoriamento Remoto, p. 7039-7046, 2009

PHIRI, D.; SIMWANDA, M.; SALEKIN, S.; NYIRENDA, V.; MURAYAMA, Y.; RANAGALAGE, Manjula. **Sentinel-2 Data for Land Cover/Use Mapping: a review.** Remote Sensing, v. 12, n. 14, p. 2291, 16 jul. 2020.

PRASAD, P.; LOVESON, V. J.; CHANDRA, P.; KOTHA, M. **Evaluation and comparison of the earth observing sensors in land cover/land use studies using machine learning algorithms.** Ecological Informatics, v. 68, p. 101522, maio 2022.

RAIYANI, K.; GONÇALVES, T.; RATO, L.; SALGUEIRO, P.; SILVA, J. R. M. da. **Sentinel-2 Image Scene Classification: a comparison between sen2cor and a machine learning approach.** Remote Sensing, v. 13, n. 2, p. 300, 16 jan. 2021.

RANA, V. K.; SURYANARAYANA, T. M. V.. **Performance evaluation of MLE, RF and SVM classification algorithms for watershed scale land use/land cover mapping using sentinel 2 bands.** Remote Sensing Applications: Society and Environment, v. 19, p. 100351, ago. 2020.

ROSA, A. H.; SILVA, Â. A. M. J.; MELO, C. de A., CARLOS, V. M.; GUANDIQUE, M. E. G., FRACETO, L. F.; LOURENÇO, R. W. **Diagnóstico ambiental e avaliação de uso e ocupação do solo visando a sustentabilidade da represa de Itupararanga, importante área da bacia do médio Tietê**. In: POMPÊO, M.; MOSCHINI-CARLOS, V.; NISHIMURA, P. Y.; SILVA, S. C. da; DOVAL, J. C. L.. **Ecologia de reservatórios e interfaces**. São Paulo: Instituto de Biociências da Universidade de São Paulo, ed. xii, p. 212-231, 2015.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. **Monitoring vegetation systems in the great plains with ERTS**. In: Third ERTS Symposium, Proceedings, NASA SP-351, NASA, Washington, DC, v. 1, p. 309-317, 1973

RUSSELL, S.; NORVIG, P. **Inteligência Artificial**. Disponível em: Minha Biblioteca, (3rd edição). Grupo GEN, 2013.

SCORNET, E.; BIAU, G.; VERT, J. P. **Consistency of random forests**. Ann. Statist. 43 (4) 1716 - 1741, August 2015.

STEINHAUSEN, M. J.; WAGNER, P. D.; NARASIMHAN, B.; WASKE, B. **Combining Sentinel-1 and Sentinel-2 data for improved land use and land cover mapping of monsoon regions**. International Journal Of Applied Earth Observation And Geoinformation, v. 73, p. 595-604, dez. 2018.

TU, Y.; LANG, W.; YU, L.; LI, Y.; JIANG, J.; QIN, Y.; WU, J.; CHEN, T.; XU, B. **Improved Mapping Results of 10 m Resolution Land Cover Classification in Guangdong, China Using Multisource Remote Sensing Data With Google Earth Engine**. Ieee Journal Of Selected Topics In Applied Earth Observations And Remote Sensing, v. 13, p. 5384-5397, 2020..

VASILAKOS, C.; KAVROUDAKIS, D.; GEORGANTA, A. **Machine Learning Classification Ensemble of Multitemporal Sentinel-2 Images: the case of a mixed mediterranean ecosystem**. Remote Sensing, v. 12, n. 12, p. 2005, 22 jun. 2020.

WAŚNIEWSKI, Adam; HOŚCŁO, Agata; CHMIELEWSKA, Milena. **Can a Hierarchical Classification of Sentinel-2 Data Improve Land Cover Mapping?** Remote Sensing, v. 14, n. 4, p. 989, 17 fev. 2022.

WEI, J.; CHU, X.; SUN, X. Y.; XU, K.; DENG, H. X.; CHEN, J.; WEI, Z.; LEI, M. **Machine learning in materials science**. Infomat, v. 1, n. 3, p. 338-358, set. 2019.

WRIGHT, Marvin N.; ZIEGLER, Andreas. **ranger: A fast implementation of random forests for high dimensional data in C++ and R**. arXiv preprint arXiv:1508.04409, 2015.

XIE, G.; NICULESCU, S. **Mapping and Monitoring of Land Cover/Land Use (LCLU) Changes in the Crozon Peninsula (Brittany, France) from 2007 to 2018 by Machine Learning Algorithms (Support Vector Machine, Random Forest, and Convolutional Neural Network) and by Post-classification Comparison (PCC)**. Remote Sensing, v. 13, n. 19, p. 3899, 29 set. 2021.

Xu, H. **Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery.** International Journal of Remote Sensing, 27(14), 3025-3033, 2006

YANG, Y.; YANG, D.; WANG, X.; ZHANG, Z.; NAWAZ, Z. **Testing Accuracy of Land Cover Classification Algorithms in the Qilian Mountains Based on GEE Cloud Platform.** Remote Sensing, v. 13, n. 24, p. 5064, 14 dez. 2021.

ZHA, Y.; GAO, J.; NI, S. **Use of normalized difference built-up index in automatically mapping urban areas from TM imagery.** International Journal Of Remote Sensing, v. 24, n. 3, p. 583-594, jan. 2003.

ZHANG, W.; TANG, P.; CORPETTI, T.; ZHAO, L. **WTS: a weakly towards strongly supervised learning framework for remote sensing land cover classification using segmentation models.** Remote Sensing, v. 13, n. 3, p. 394, 23 jan. 2021.