

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

Jonilson Michel Fontes Galvão¹
Antônio Cesar Germano Martins²
Darllan Collins da Cunha e Silva³
Tharidu Lokupathirage⁴

Recursos Hídricos e Qualidade da Água

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 ganhou destaque nos últimos anos. Aplicar tais recursos na gestão de reservatórios é de suma importância, principalmente para a região do sudeste brasileiro. Dessa forma, o presente estudo visa apresentar resultados da classificação de regiões em imagens do satélite Sentinel-2 da Área de Preservação Ambiental de Itupararanga, especificamente no entorno do reservatório de Itupararanga, na região metropolitana de Sorocaba, Estado de São Paulo, Brasil. Através do pré-processamento e utilização de índices espectrais (NDVI, GNDVI, NDWI, MNDWI e NDBI) como *input* no sistema de classificação, o conjunto de imagens de uso com comprimentos de bandas diferentes foram avaliados por meio da técnica de *Support Vector Machine* (SVM). Os resultados foram derivados de uma matriz de confusão, sendo registrado um valor de 99,27% de acurácia global para a técnica, o qual precede de resultados similares ao da literatura. Derivadas da SVM, o cálculo quanto uso do solo registra, em termos de porcentagem, 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. A gestão do recurso hídrico para a região metropolitana é fundamental e o uso de inteligência artificial para maximizar a tomada de decisão é crucial.

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

¹Doutorando em Ciências Ambientais – PPGCA, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Programa de Pós-graduação em Ciências Ambientais, jonilson.galvao@unesp.br

²Prof. Dr. Universidade Estadual Paulista “Júlio de Mesquita Filho” – Campus Sorocaba-SP, Instituto de Ciências e Tecnologia de Sorocaba - ICTS, antonio.martins@unesp.br

³Prof. Dr. Universidade Estadual Paulista “Júlio de Mesquita Filho” – Campus Sorocaba-SP, Instituto de Ciências e Tecnologia de Sorocaba - ICTS, darllan.collins@unesp.br

⁴Aluno de Estágio Internacional, École D'ingénieur-E·S D'un Numérique Utile, ESIEA. tharidu.lokupathirage@unesp.br



INTRODUÇÃO

A cobertura do solo é resultante de propriedades biológicas e físicas da superfície do planeta Terra e compreende uma diversidade de vegetações, corpos d'água, coberturas de neve, rochas, solos expostos e ambientes construídos (XIE e NICULESCU, 2021; PRASAD *et al.*, 2022), além de ser um elemento chave para a ciência do sistema terrestre, objeto de estudo na compreensão das interações entre atividades humanas e mudanças superficiais (EBRAHIMY *et al.*, 2021).

O uso da terra é o resultado da interação homem-natureza e a cobertura da terra representa a condição física da superfície (PRASAD *et al.*, 2022) e são influenciados por ações humanas e condicionam as características e processo ambientais a estes dois fatores (XIE; NICULESCU, 2021).

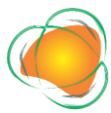
O avanço nas técnicas de sensoriamento remoto (SR) e o sistema de informações geográficas (SIG) fornecem ferramentas para o apoio no estudo da dinâmica ambiental (GHAYOUR *et al.*, 2021). Dessa forma, a informação sobre o uso e cobertura do solo desempenha um papel fundamental na análise da informação da terra para o desenvolvimento e criação de políticas públicas (RANA; SURYANARAYANA, 2020).

Para tanto, a 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; NICULESCU, 2021; EBRAHIMY *et al.*, 2021).

Apoiado a esta questão, 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 *et al.*, 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) para a classificação de imagens do S-2 tem obtido

Realização



resultados com alta acurácia (GHAYOUR *et al.*, 2021; LUCA *et al.*, 2022). Neste contexto, o presente trabalho aborda a aplicação de SVM 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.

METODOLOGIA

ÁREA DE ESTUDO

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). A figura 1 apresenta o reservatório e a região entorno do mesmo.

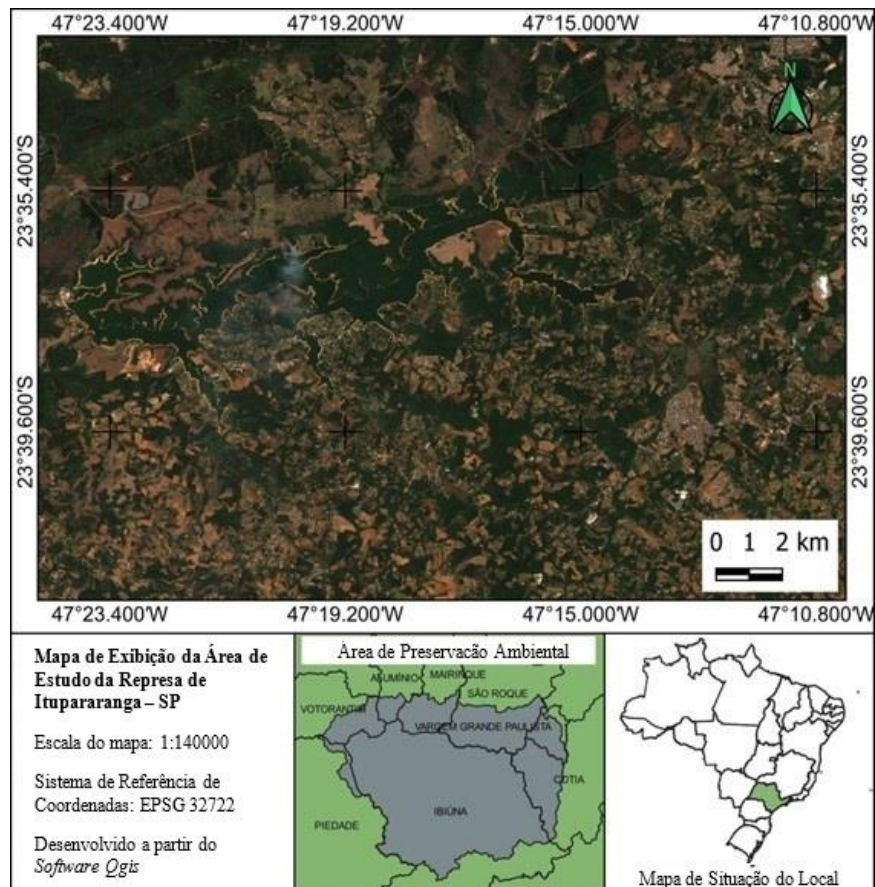


Figura1. Área de Estudo

Fonte: Elaborado pelos autores (2023)

Realização



AQUISIÇÃO DA IMAGEM E PRÉ-PROCESSAMENTO

A aquisição do produto segue processo de cadastro, escolha da área de estudo, definição do produto através de filtros e *download* do material. 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 1 contém as características do produto utilizado nesta pesquisa.

Tabela 1. 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).

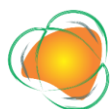
INDICES ESPECTRAIS E AMOSTRAGEM

O monitoramento da mudança na cobertura do solo, corpos d'água, estágios da planta são atividades, quando utilizados imagens satelitais, que tomam como base os índices espectrais (XIE; NICULESCU, 2021; ABDI, 2019). Esta pesquisa faz uso dessa abordagem a Tabela 2, contém organização dos índices usados, sendo as bandas do satélite resumidas em: Banda verde (B03), Banda vermelha (B04), Infravermelho (B08) e SWIR (B11).

Tabela 2. 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 –	Gao (1996)	$[B08 - B11] / [B08 + B11]$	Corpo hídrico

Realização



NDWI1

Índice de Água de Diferença Normalizada – NDWI2	Mcfeters (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: Elaborado pelos autores (2023)

O procedimento de aquisição das informações de *input* para a SVM segue amostragem com base na comparação das regiões de interesse através de supervisão (PRASAD et al., 2022; XIE e NICULESCU, 2021), feita no *Software QGIS*. Os polígonos foram armazenados em arquivos vetoriais, em formato *shapefile*, pois asseguram o sistema de referência das coordenadas para posterior uso do *Software de Modelagem* (LUCA et al., 2022). A Tabela 3 contém o resumo da quantidade de

Todos os arquivos foram armazenados em um banco de dados, com vista a utilização posterior por meio do software de estatística Rstudio. 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 3, bem como o identificador usado em todo o procedimento, desde o *Qgis* ao *Rstudio*.

Tabela 3. 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

Realização



FERRAMENTAS COMPUTACIONAL E APLICAÇÃO DOS MÉTODOS

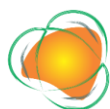
A avaliação do desempenho de diferentes algoritmos de aprendizado de máquina na classificação de imagens de satélites foi realizado através de software de código aberto, o Rstúdio. O Hardware computacional utilizado nesta pesquisa foi uma WorkStation com processador Intel (R) Xeon (R) com 16 GB de memória, velocidade de 2.53 GHz e com capacidade operacional de 64 Bits. A SVM foi executada em linguagem R, através da interface *Rstúdio* e o pacote *e1071*.

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; LUCA *et al.*, 2022).

RESULTADOS E DISCUSSÃO

Análises de algoritmos de classificação de imagens fazem uso de matriz de confusão para compreensão das acurácias registradas pelo modelo (LUCA *et al.*, 2022; PRASAD *et al.*, 2022). Dessa forma, o primeiro resultado apresentado é a matriz de confusão derivado do *Software* estatístico, conforme observado na Figura 2, sendo a diagonal principal o resultado que mostra o comportamento dos resultados atingidos pelo modelo.



	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

Figura 1. Matriz de Confusão do SVM Radial

Fonte: Elaborado pelos autores (2023)

A Tabela 4 é desenvolvida com base na matriz de confusão, sendo que as classes de água e nuvem apresentam 100% de precisão entre as AP e AU, em conformidade com os estudos de Ebrahimy 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 de AP e AU. 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 4. Tabelas das acurácias atingidas pela SVM

Classe	SVM RADIAL	
	Acurácia do produtor	Acurácia do Usuário
Agricultura	100%	98,97%
Água	100%	100%
Arbórea	99,65%	98,63%
Eucalipto	99,65%	100%
Nuvem	100%	100%
Rasteira	98,62%	98,28%
Solo e Arbusto	99,31%	98,62%
Solo exposto	99,65%	99,65%
Urbano	96,54%	99,29%
Acurácia Global	99,27%	

Fonte: Elaborado pelos autores (2023)

A abordagem através de SVM apresentou resultados satisfatórios. A boa precisão alcançada

Realização



pelo conjunto de dados S-2, registrou pontuação de 99,27 de acurácia global, resultados similares são registrados nos trabalhos em Luca et al. (2022), Waśniewski et al. (2022), Dabija et al. (2021), Ebrahimy et al. (2021), Ghayour et al. (2021) e Vasilakos et al. (2020).

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 trabalhos de Rana e Suryanarayana (2020) e Ebrahimy et al. (2021).

Dessa forma, o algoritmo treinado com os conjuntos de *Input* por meio dos índices espectrais foi aplicado no *stack*, ou conjunto, de dados pertencentes às imagens satelitais do Sentinel-2. Conforme se observa na Figura 3, o mapa de cobertura de uso solo encontra-se classificado pelo método da SVM.

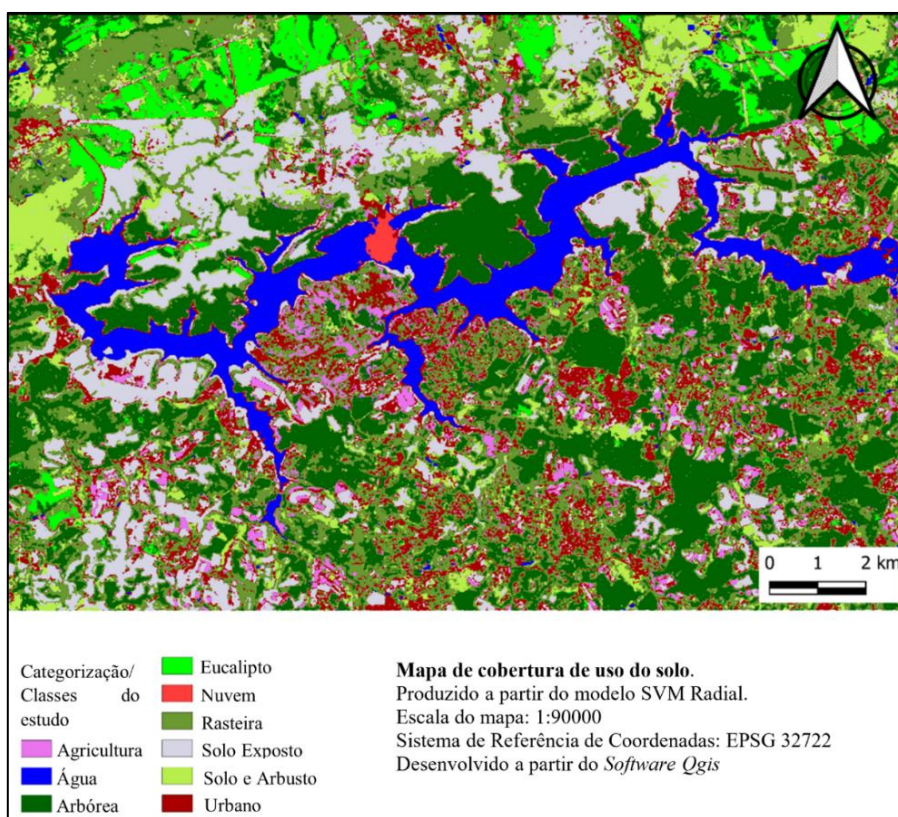


Figura 3. Mapa de cobertura do solo da área de estudo usando SVM radial.

Fonte: Elaborado pelos autores (2023).

O cálculo do uso da terra é fator fundamental para a gestão territorial, dessa forma, a partir do modelo SVM, a porcentagem das áreas foi calculada para observação do uso da



terra na região de estudo. A quantidade de hectares encontradas foi da ordem de 24.328.000 e o maior uso da área de estudo encontra-se para a porção de vegetação por meio vegetação rasteira, da interação solo e arbusto e a vegetação arbórea, respectivamente, conforme observado na Tabela 5.

Tabela 5. 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
8	Interação Solo e Arbusto	1.762.860	7,2
9	Área urbana	2.779.220	11,4
Total		24.328.000	100,0

Fonte: Fonte: Elaborado pelos autores (2023).

CONCLUSÕES

A abordagem da classificação de imagens do satélite Sentinel-2 através da SVM no entorno do reservatório de Itupararanga apresentou resultados satisfatórios, conforme observado nos valores da literatura.

A espacialização da informação providas de modelos de Inteligência Artificial é rápida e o produto testado contém resultados para a tomada de decisão de gestores ambientais, principalmente por trata-se de uma região de captação de recurso hídrico para a região metropolitana de Sorocaba.

Diferentes abordagens são fundamentais no que refere monitoramento territorial e o impacto que causa na gestão dos recursos naturais, principalmente o hídrico. Dessa forma, o uso das Redes Neurais Convolucionais para classificação de imagens do Sentinel-2 está

Realização



recebendo maior notoriedade, logo é uma abordagem considerada para trabalhos futuros.

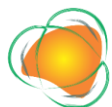
AGRADECIMENTOS

A UNESP-ICTS pela estrutura e provimento de todo material necessário para o desenvolvimento deste estudo.

REFERÊNCIAS

- 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.
- 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.
- 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.
- 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.
- 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.
- LUCA, G.; SILVA, J. M. N.; FAZIO, S.; 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.
- 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.
- 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.

Realização



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: POMPEO, 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.

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, A.; HOŚCİHO, A.; 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.

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.

Realização