



## Comparação entre os métodos de aprendizado de máquina Random Forest e Support Vector Machine para classificação de imagem do satélite CBERS4a

### Comparison between the machine learning methods Random Forest and Support Vector Machine for classifying CBERS4a satellite imagery

Gabriele Samaris de Souza <sup>1</sup>, Ligia Flávia Antunes Batista <sup>2</sup>

#### RESUMO

O sensoriamento remoto desempenha um papel crucial na compreensão e monitoramento das mudanças ambientais, permitindo a coleta de dados sem intervenção direta. Neste estudo, focado na região de Ourinhos, São Paulo, foram utilizadas imagens do satélite CBERS 4A, combinando dados multiespectrais e pancromáticos, para classificar o uso e cobertura do solo. Dois métodos de aprendizado de máquina, *Random Forest* e *Support Vector Machine* (SVM) com diferentes Kernels, foram empregados para essa classificação. O processo seguiu a técnica de classificação baseada em objetos (OBIA), que envolveu fusão de imagens, segmentação, extração de atributos, amostragem, classificação e análise de acurácia. Os resultados indicaram que o método *Random Forest* teve um desempenho superior na delimitação de áreas de vegetação rasteira. Já o método *Support Vector Machine* com *Kernel Linear* se destacou na classificação das áreas de água com sedimentos. Com relação às métricas, acurácia global e índice Kappa, o classificador SVM com *Kernel Linear*, apresentou resultados superiores em comparação com o *Random Forest*.

**PALAVRAS-CHAVE:** sensoriamento remoto; uso e cobertura do solo; análise de imagens baseada em objetos.

#### ABSTRACT

Remote sensing plays a crucial role in understanding and monitoring environmental changes, allowing data to be collected without direct intervention. In this study, focused on the region of Ourinhos, São Paulo, images from the CBERS 4A satellite, combining multispectral and panchromatic data, were used to classify land use and land cover. Two machine learning methods, Random Forest and Support Vector Machine with different Kernels, were used for this classification. The process followed the object-based classification technique (OBIA), which involved image fusion, segmentation, attribute extraction, sampling, classification and accuracy analysis. The results indicated that the Random Forest method performed better in delimiting areas of undergrowth/cultivation. The Support Vector Machine with Linear Kernel method stood out when classifying areas of water with sediment. Visually, both had errors and successes. With regard to the metrics, global accuracy and Kappa index, the SVM classifier with Linear Kernel showed superior results compared to Random Forest.

**KEYWORDS:** remote sensing; land use and cover; object-based image analysis.

## INTRODUÇÃO

O sensoriamento remoto é essencial para entender e monitorar mudanças ambientais, destacando-se por sua capacidade de coletar dados sem intervenção direta e classificar o uso e cobertura do solo, desempenhando um papel crucial em estudos ambientais.

<sup>1</sup> Universidade Tecnológica Federal do Paraná - UTFPR, Londrina, Paraná, Brasil. E-mail: gabrielesamaris@alunos.utfpr.edu.br. ID Lattes: 5032515350282876.

<sup>2</sup> Docente no Curso de Engenharia Ambiental. Universidade Tecnológica Federal do Paraná, Londrina, Paraná, Brasil. E-mail: ligia@utfpr.edu.br. ID Lattes: 9956346241888896.

Os métodos de aprendizado de máquina como o algoritmo *Random Forest* e *Support Vector Machine* auxiliam no processo de classificação de imagem de sensoriamento remoto, em que o algoritmo é treinado para reconhecer padrões a partir de dados de entrada significativos para o estudo, gerando um novo conjunto de dados de saída, que representará a classificação das áreas ou objetos na imagem (ANJOS; ALMEIDA e GALVÃO, 2015).

O método de classificação *Support Vector Machine* se baseia em uma separação ótima entre classes, por meio de um hiperplano de separação (NASCIMENTO et al., 2009; MOREIRA, FERNANDES e NERY, 2014). Através de uma dada função, o método irá maximizar os resultados. Tais funções são chamadas de *Kernels* que, segundo Sousa, Teixeira e Silva (2009), são usadas para projetar os dados de entrada para o espaço com mais dimensões. Os principais *Kernels* são: linear, polinomial, radial, sigmoidal.

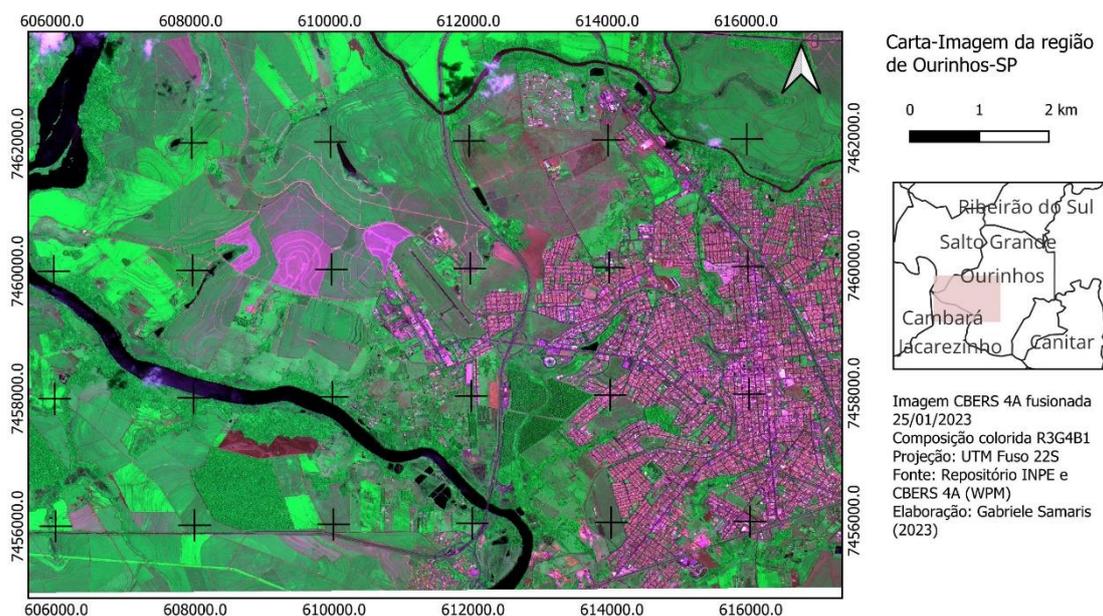
O classificador *Random Forest* usa várias árvores de decisão treinadas e validadas com sub-amostras aleatórias e repetidas. A classe mais votada pelas árvores é a escolhida pelo classificador (CHO et al., 2021).

Este trabalho tem como objetivo a classificação do uso e cobertura do solo da área de Ourinhos- São Paulo, utilizando imagens fusionadas obtidas pelo satélite CBERS 4A com a câmera Multiespectral e Pancromática de Ampla Varredura (WPM), com dois métodos de classificação de aprendizado de máquina, *Random Forest* e *Support Vector Machine*.

## METODOLOGIA

A área estudada se encontra em Ourinhos- SP entre as coordenadas UTM 605506E a 617378E e 7454956N a 7464492N. A área é de 113, 295 km<sup>2</sup>, conforme mostra a figura 1.

Figura 1 – Mapa da área de estudo (Imagem CBERS 4<sup>a</sup> WPM com composição colorida R3G4B1)



Fonte: Autoria própria (2023).



As imagens foram adquiridas pelo satélite sino-brasileiro CBERS 4A, desenvolvido em parceria do Brasil com a China, pela câmera Multiespectral e Pancromática de Ampla Varredura (WPM), referente ao dia 25 de janeiro de 2023. Foi utilizada a composição colorida RGB 341, com Sistema de Referências de Coordenadas WGS 84 e projeção cartográfica/ UTM, fuso 22 e hemisfério Sul.

Foi aplicada a técnica de classificação baseada em objetos (*Object Based Image Analysis*- OBIA) com a seguinte sequência de etapas: fusão, segmentação, extração de atributos, amostragem, classificação e análise de acurácia (JUNIOR et al.,2018).

Foi feita a fusão das bandas multiespectrais com a banda pancromática e com isso, a imagem passou a ter 2 m de resolução espacial. Este processo envolve a combinação das bandas pancromáticas e multiespectrais para produzir uma imagem sintética. Foi aplicado o método IHS (*intensity, hue, saturation*), baseado na transformação do sistema de cores RGB (Fonseca et al., 2011).

Foram empregados os softwares livres QGIS e Orfeo Toolbox para os variados processamentos da imagem e elaboração dos mapas.

A escolha dos parâmetros de segmentação foi fundamentada em uma análise visual, que considerou a agregação de regiões com características similares, além de uma avaliação quantitativa baseada no número total de segmentos resultantes.

Em seguida, foi feita a extração de atributos estatísticos das 4 bandas da camada, por meio de modelo gráfico, na qual foram geradas as estatísticas dos pixels por segmento, sendo: máximo, mínimo, desvio padrão, variância, média, mediana, contagem, soma.

Posteriormente foi realizado o treinamento do classificador, que consiste em definir polígonos de amostras. Foram coletadas 50 amostras de cada uma das seguintes classes: água limpa, água com sedimentos, área urbana, vegetação densa e vegetação rasteira/cultivo. Não foi incluída a classe solo exposto devido ao fato da imagem conter poucos trechos com este tipo de cobertura. Do total de 250 amostras, 50 delas foram utilizadas para validação.

Finalizado o treinamento, iniciou-se a etapa de aplicação dos métodos de classificação *Random Forest* e *Support Vector Machine* variando os *Kernels*, sendo eles: linear, polinomial, radial e sigmoidal. Para a avaliação dos resultados, foram calculadas as métricas de acurácia global e índice Kappa.

## RESULTADOS E DISCUSSÃO

Os melhores parâmetros para segmentação foram: raio espacial 5, raio de alcance 30 e tamanho mínimo do segmento 100. Estas configurações resultaram um total de 105.608 segmentos para a imagem.

De modo geral, o método *Random Forest* e o classificador SVM com *Kernel Linear* apresentaram bons resultados. As demais funções de *Kernel* geraram resultados muito inferiores, atribuindo uma única classe em grande parte da área processada.

A partir da análise visual e dos resultados de acurácia global e índice Kappa da Tabela 1, pode-se afirmar que os melhores resultados foram dos classificadores *Random Forest* e *Support Vector Machine* com *Kernel Linear*. Ressalta-se que a acurácia global de 100% atingida pelo SVM com *Kernel Linear* refere-se aos resultados corretos obtidos com as amostras de validação e não indica que o resultado da imagem classificada como um todo é totalmente correto.

**Tabela 1 – Resultados de Acurácia Global e Índice Kappa**

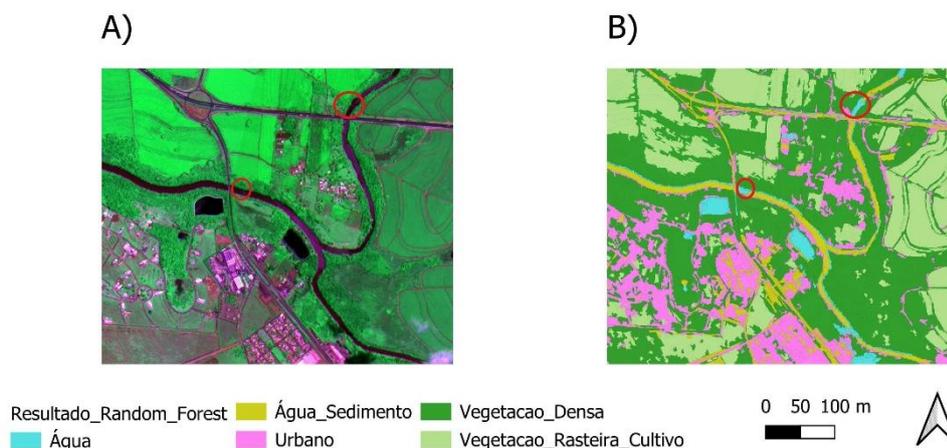
Método	Acurácia Global [%]	Índice Kappa
<i>Random Forest</i>	85,8446	0,8063
<i>SVM Kernel Linear</i>	100	1
<i>SVM Kernel Sigmoidal</i>	1,2004	0
<i>SVM Kernel Polinomial</i>	24,9909	0,1467
<i>SVM Kernel Radial</i>	1,1639	0,0004

Fonte: Autoria própria (2023).

Em geral, a classe vegetação rasteira/cultivo foi bem classificada pelo método *Random Forest*. A classe água com sedimentos foi melhor delimitada pelo método SVM com *Kernel Linear*. Em ambos os classificadores, algumas áreas de solo exposto foram erroneamente categorizadas como área urbana ou como água com sedimentos, justamente por não ter feito parte da amostragem. Isso se deve também ao comportamento espectral parecido entres essas classes.

Conforme ilustrado na Figura 2b, um recorte do resultado gerado por meio da aplicação do método *Random Forest* na região da imagem original (Figura 2a), é possível observar algumas confusões na classificação entre água com sedimentos e água limpa, em que a primeira foi erroneamente classificada como a segunda nas áreas destacadas com o círculo em vermelho.

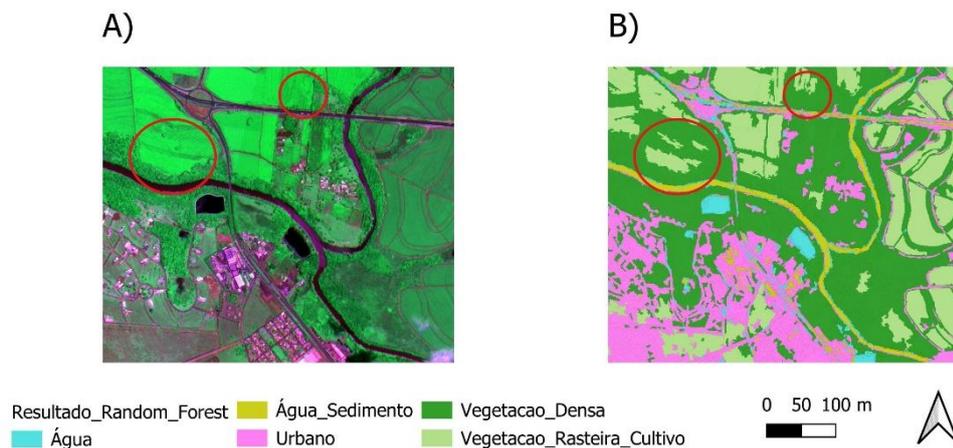
**Figura 2- Resultado da classificação pelo método *Random Forest***



Fonte: Autoria própria (2023).

Como mostram as Figuras 3a e 3b, podemos ver confusões também na classe vegetação rasteira/cultivo, a qual foi classificada em alguns trechos como vegetação densa. O classificador SVM com *Kernel* linear apresentou um desempenho inferior na delimitação da classe vegetação rasteira/cultivo, em comparação com o classificador *Random Forest*.

Figura 3- Resultado do método *Support Vector Machine* com *Kernel* Linear



Fonte: Autoria própria (2023).

## CONCLUSÃO

As técnicas de aprendizado de máquina empregadas geraram resultados muito bons, sendo que para o SVM apenas o *Kernel* linear mostrou-se adequado para os dados avaliados.

Em ambos os classificadores foi possível observar confusões com relação a classificação de categorias que possuem comportamento espectral semelhantes, o que é esperado neste tipo de aplicação. Analisando as métricas de índice Kappa e acurácia global, o SVM com *Kernel* Linear apresentou melhores resultados. Como a abordagem de aprendizado de máquina é altamente dependente dos volumes de dados de entrada, são necessários novos testes com áreas maiores e quantidade superior de amostras de treinamento.

Trabalhos futuros envolvem avaliação de dados de diferentes sensores, áreas e resoluções espaciais.

## Agradecimentos

Agradeço a instituição de ensino UTFPR – Londrina pela oportunidade de aprendizado enriquecedora.

## Conflito de interesse

Não há conflito de interesse.



## REFERÊNCIAS

ANJOS, C. S.; ALMEIDA, C. M.; GALVÃO, L. S. Identificação de materiais urbanos por meio de métodos inovadores de classificação de imagens. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 17. (SBSR), 2015, João Pessoa. **Anais...** São José dos Campos: INPE, 2015. p. 4377-4384. Internet. ISBN 978-85-17-0076-8. IBI: <8JMKD3MGP6W34M/3JM4CN2>. Disponível em: <<http://urlib.net/ibi/8JMKD3MGP6W34M/3JM4CN2>>. Acesso em: 28 ago. 2023.

CHO, D. F. et al. Desempenho do algoritmo de classificação de imagens Random Forest para mapeamento do uso e cobertura do solo no cerrado brasileiro. **Anuário do Instituto de Geociências**, v. 44, 2021. ISSN 0101-9759. Disponível em: <<https://revistas.ufrj.br/index.php/aigeo/article/view/37979>>. Acesso em: 26 ago. 2023.

FONSECA, L.; NAMIKAWA, L.; CASTEJON, E.; CARVALHO, L.; PINHO, C.; PAGAMISSE, A. (2011). Image Fusion for Remote Sensing Applications. In: Zheng, Y. **Image Fusion and Its Applications**. Mississippi: Tech, 2011, p. 153-178.

JUNIOR, E. L. et al. Comparação entre abordagens de classificação por pixel e baseada em objeto para o monitoramento de recursos hídricos. **Revista Brasileira de Cartografia**, Uberlândia, v.70, n.4, p. 1289 - 1309, out/dez, 2018. Disponível em: <<https://seer.ufu.br/index.php/revistabrasileiracartografia/article/view/46142>>. Acesso em: 5 set. 2023.

MOREIRA, A. A.; FERNANDES, F. H. S.; NERY, C.V.M. Aplicação do algoritmo Support Vector Machine na análise espaço-temporal do uso e ocupação do solo na Bacia do Rio Vieira. **Caminhos de Geografia, Uberlândia**, v.15, n.50, p. 152 - 163, jun, 2014.

NASCIMENTO, R. F. F.; et al. O algoritmo Support Vector Machines (SVM): avaliação da separação ótima de classes em imagens CCD-CBERS-2. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14. (SBSR), 2009, Natal. **Anais...** São José dos Campos: INPE, 2009. p. 2079-2086. DVD, On-line. ISBN 978-85-17-00044-7. (INPE-15957-PRE/10566). Disponível em: <<http://urlib.net/ibi/dpi.inpe.br/sbsr@80/2008/10.20.10.59>>. Acesso em: 25 ago. 2023.

SOUSA, B. F. S.; TEIXEIRA, A. S.; SILVA, F. A. T. F. Classificação de bioma caatinga usando Support Vector Machines (SVM). In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14. (SBSR), 2009, Natal. **Anais...** São José dos Campos: INPE, 2009. p. 7917-7924. DVD, On-line. ISBN 978-85-17-00044-7. (INPE-16045-PRE/10654). Disponível em: <<http://urlib.net/ibi/dpi.inpe.br/sbsr@80/2008/11.17.14.03>>. Acesso em: 25 ago. 2023.