

EXPLORAÇÃO DE DIFERENTES COMPOSIÇÕES E CLASSIFICAÇÃO DE IMAGENS DE COBERTURA DO SOLO ATRAVÉS DA PLATAFORMA *GOOGLE EARTH ENGINE*

ELISANDRA HERNANDES DA FONSECA¹; KAREN GULARTE PERES MENDES²; DANIELE MARTIN SAMPAIO³; CAROLINE DE OLIVEIRA NUNES⁴; VITOR SOUZA MARTINS⁵; HUGO ALEXANDRE SOARES GUEDES⁶

¹Universidade Federal de Pelotas – elisandrah.fonseca@gmail.com

²Universidade Federal de Pelotas, Instituto Federal Sul-rio-grandense – karperes@gmail.com

³Universidade Federal de Pelotas – dmartinsampaio@gmail.com

⁴Universidade Federal de Pelotas – nunescarolineoliveira@gmail.com

⁵Michigan State University – vitorstmartins@gmail.com

⁶Universidade Federal de Pelotas – hugo.guedes@ufpel.edu.br

1. INTRODUÇÃO

O recente alinhamento da ciência de dados e sensoriamento remoto é de grande importância, particularmente, no subcampo do mapeamento de uso e cobertura do solo, já que é considerada uma ferramenta de gestão indispensável para diversos setores. Plataformas como *Google Earth Engine* (GEE) que aceitam diferentes linguagens, como *Javascript* e *Python*, fornecem ao usuário ferramentas de programação amigáveis, facilitando o desenvolvimento de diversos métodos de mapeamento de uso e cobertura. Neste sentido, nas últimas décadas os métodos convencionais têm sido gradativamente substituídos por métodos que requerem menor tempo de processamento e maior poder computacional. Para isso, uma solução possível é a implementação de algoritmos de Aprendizado de Máquina que não apenas fornecem desempenho de computação eficiente em termos de potência e tempo, mas também fornecem resultados mais precisos do que os métodos convencionais (ABDI, 2020).

Em razão da popularização do uso de aprendizado de máquina e de plataformas de computação em nuvem de alto desempenho, as comunidades de sensoriamento remoto e observação da Terra têm se dedicado a integração e comparação de diferentes técnicas, múltiplos algoritmos de classificação e manipulação de diversas composições temporais. No GEE, é possível aplicar diretamente diferentes classificadores como *Random Forest* (RF), *Classification and Regression Trees* (CART) e *Support Vector Machine* (SVM) em uma determinada imagem, ou ainda fazer estudos multitemporais selecionando o máximo de imagens possível para preencher as lacunas, sem se preocupar em como as diferentes estações do ano podem afetar a precisão da classificação (PIMPLE *et al.* 2018).

Neste sentido, este trabalho visa analisar o efeito de diferentes métodos de composição nos resultados da classificação. A partir de 9 conjuntos de dados distintos de *Land use and cover* (LULC), pretende-se verificar os desempenhos obtidos por meio de integração de múltiplas técnicas de aprendizado de máquina e composição temporal no GEE, aplicando-os em séries temporais das imagens Sentinel-2 nível 2A. Esses conjuntos de dados serão treinados e validados pelas mesmas amostras de alta resolução do *Google Earth* e aplicados em uma cena com heterogeneidade de classes de uso e cobertura do solo para o ano de 2020.

2. METODOLOGIA

Para realização desta pesquisa foram gerados 9 conjuntos de dados de LULC através do *Google Earth Engine* (GEE). Primeiramente, escolheu-se

trabalhar com dados multiespectrais do satélite Sentinel-2 nível 2A do sensor MSI, devido à qualidade das suas imagens ópticas, corrigidas atmosféricamente e com resolução espacial média (10, 20 e 60m). As imagens foram adquiridas e filtradas na plataforma GEE. Selecionou-se então as bandas (“B2”, “B3”, “B4”, “B5”, “B8”, “B8A” e “B11”), dispostas em posições estratégicas do espectro, a fim de identificar padrões de comportamento espectral de diferentes alvos como vegetação, solo e água, dentre outros. Selecionou-se uma cena em uma região de bastante heterogeneidade consultando o MapBiomas, e, em seguida, todas as imagens disponíveis para 2020 foram filtradas para essa região (141 cenas) e para porcentagem de cobertura de nuvens menor que 10%, resultando em 39 cenas.

A próxima etapa consistiu na separação dos conjuntos, através do filtro por data, para os períodos anual (janeiro-dezembro), sazonal (período com menor porcentagem de nuvens – junho-outubro) e melhor cena do ano (22-08-2020). Aplicou-se então o método de composição mediana, para os conjuntos anual e sazonal, gerando pixels de valor mediano, e livres de nuvens da coleção de imagens (máscara de nuvens). Além do método da mediana, no GEE também existem outros métodos disponíveis como média, mínimo, máximo, desvio padrão (PHAN et al., 2020). Ao aplicar o método média não houve mudanças significativas em comparação a mediana sendo assim, optou-se por retirá-la desta pesquisa.

Os compostos sazonais e anuais, obtidos a partir da mediana das bandas, e também a melhor cena, foram utilizados como dados de entrada para os algoritmos RF, CART e SVM. Esses algoritmos de aprendizado de máquina são do tipo supervisionado e não paramétrico, e podem ser acessados por meio do pacote *ee.Classifier* no GEE. O RF faz previsões a partir de múltiplas árvores de decisão, onde cada uma é gerada a partir dos valores de um vetor aleatório (BREIMAN, 2001). Neste trabalho selecionou-se 100 árvores (*ntree* = 100), enquanto *mtry* foi definido com o valor padrão (raiz quadrada do número total de recursos). O CART deriva-se de decisões binárias simples e funciona relacionando os atributos de entrada (variáveis preditoras) e os atributos de destino (BREIMAN et al., 1984). Ele gera árvores de classificação e árvores de regressão, dependendo se o atributo é nominal ou contínuo, respectivamente. Neste trabalho, o número máximo de nós folha em cada árvore foi definido como 100 (*maxNodes* =100). Finalmente, o SVM possui vantagens em gerenciar a escassez de amostras; ele usa hiperplanos para dividir vetores de suporte para classificar os pontos de dados de log claramente com o objetivo de encontrar os dois tipos de vetores de suporte independentes com a maior margem (distância máxima) (YAO et al., 2008). O SVM é composto por vários hiperparâmetros: tipo de kernel, gama e valor de penalidade. Esses hiperparâmetros podem ser ajustados para melhorar o desempenho do SVM. Neste trabalho foi utilizado (*kernelType*=Função Polinomial ‘POLY’, grau=3).

As classificações resultaram nas composições expostas na Tabela 1. O uso do solo foi dividido em 5 classes: vegetação, agricultura, área construída, corpos d'água e solo exposto. As amostras de treinamento e validação foram coletadas de 625 pontos, sendo destinados 75% para treinamento e 25% para validação. A amostragem foi realizada através da plataforma GEE, selecionando visualmente os alvos em imagens de alta resolução do *Google Earth*; a validação da classificação foi dada por matriz de confusão e índice Kappa, calculados também através do GEE (CONGALTON, 1991).

Tabela 1- Composição do Conjunto de classificação.

Composição do Conjunto	RF	SVM	CART
Melhor imagem 2020	conjunto 1	conjunto 2	conjunto 3
Mediana Anual (Jan-Dez)	conjunto 4	conjunto 5	conjunto 6
Mediana Sazonal (Maio-Out)	conjunto 7	conjunto 8	conjunto 9

3. RESULTADOS E DISCUSSÃO

A classificação do uso e cobertura do solo foi realizada dos 9 conjuntos descritos, apresentando precisão geral entre 86,07% e 95,25% e índice Kappa entre 81,98 % e 98,86%. Dentre os conjuntos, os classificados com algoritmo RF foram os que obtiveram maiores valores de precisão e índice Kappa, já os classificados com SVM, obtiveram os valores mais baixos. Na Figura 1 selecionou-se uma porção da imagem a fim de comparar as diferentes composições; os conjuntos 1,2,3 referentes a cena única na primeira coluna demonstram bastante confusão em relação às classes urbano e solo exposto. Dentre os 3 conjuntos o SVM foi o que demonstrou pior desempenho com 86% de precisão geral classificando pixels de solo como área construída. Já nos conjuntos 4,5,6 a classificação foi satisfatória, tendo menor porcentagem de pixels sendo classificados erroneamente, já nos conjuntos 7,8,9 da mediana sazonal (talvez em razão da menor quantidade de cenas), as precisões baixaram um pouco.

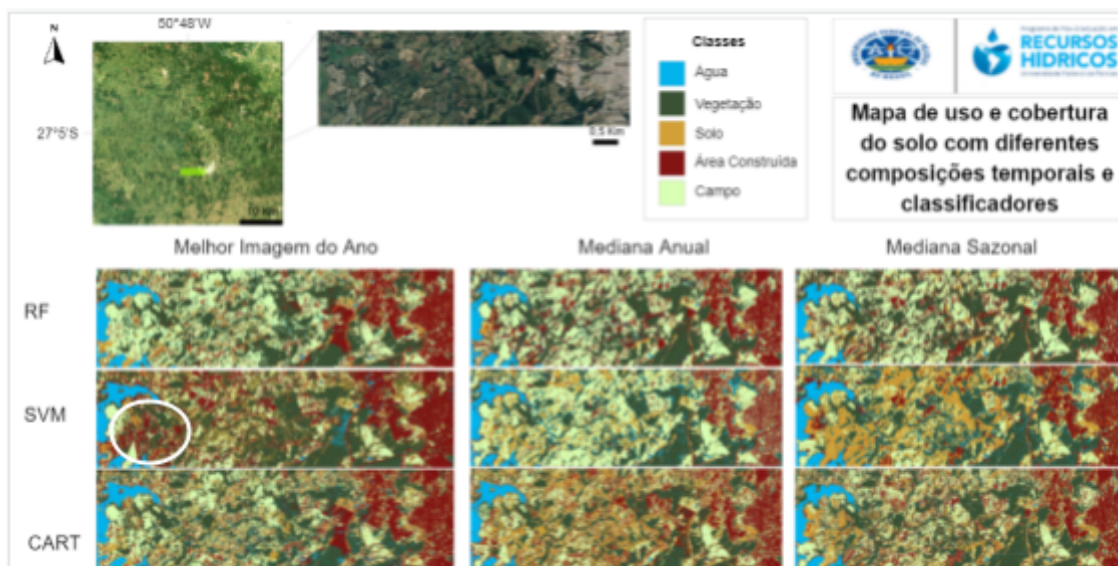


Figura 1 - Mapa das classificações de uso e cobertura da terra para diferentes dados de entrada

Na tabela 2 abaixo, é possível conferir que os maiores valores de precisão geral e índice Kappa foram para os conjuntos 4,5,6 oriundos da mediana anual, sendo o de maior precisão o conjunto 4 e o de menor precisão geral, o conjunto 2. Os principais desafios encontrados foram os pixels de solo sendo classificados frequentemente como área urbana, sendo o maior, erro de omissão 45%, para o conjunto 2 na Figura 1, na qual a área circulada mostra a região onde ocorreu a confusão maior baixando os valores de Precisão e Índice Kappa.

Tabela 2- Precisão e índice Kappa dos conjuntos analisados.

Conjunto	Precisão geral	Índice Kappa
Conjunto 1	99,16%	98,81%
Conjunto 2	86,07%	81,98%
Conjunto 3	96,10%	94,96%
Conjunto 4	99,25%	98,86%
Conjunto 5	88,57%	85,19%
Conjunto 6	98,90%	98,58%
Conjunto 7	98,87%	98,48%
Conjunto 8	86,86%	81,98%
Conjunto 9	98,35%	97,87%

4. CONCLUSÕES

Concluiu-se que o conjunto com maior precisão foi o 4, mediana anual classificada com RF, e o menor foi o conjunto 2, derivado da melhor imagem classificada com SVM. A plataforma do GEE permitiu a aplicação de diferentes métodos, de maneira rápida e eficaz tornando possível essa comparação, sendo um excelente recurso aos gestores, facilitando a tomada de decisão com base em informações e dados confiáveis e precisos.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ABDI, Abdulhakim Mohamed. 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, 2020.

BREIMAN, L., Friedman, L., Olshen, R., and Stone, C. (1984). **Classification and regression trees**. Wadsworth Inc., Belmont, California.

BREIMAN, Leo. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.

PHAN, Thanh Noi; KUCH, Verena; LEHNERT, Lukas W. Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition. **Remote Sensing**, v. 12, n. 15, p. 2411, 2020.

PIMPLE, U., SIMONETTI, D., SITTHI, A., PUNGKUL, S., LEADPRATHOM, K., SKUPEK, H., ... & TOWPRAYOON, S. (2018). Google earth engine based three decadal landsat imagery analysis for mapping of mangrove forests and its surroundings in the trat province of Thailand. **Journal of Computer and Communications**, 6, 247-264, 2018.

YAO, X.; THAM, L. G.; DAI, F. C. Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. **Geomorphology**, v. 101, n. 4, p. 572-582, 2008.