

ANÁLISE COMPARATIVA DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA O MAPEAMENTO MULTIESCALA DE UMA SUB-BACIA NO SUL DE MINAS GERAIS

Mariana B. de Vasconcelos¹ (IC), Samara C. de Azevedo (PQ)¹

¹Universidade Federal de Itajubá.

Palavras-chave: Aprendizagem de máquina. Classificação de imagens. Random forest. Support vector machine.

Introdução

Atualmente, são muitas as alterações ambientais decorrentes das atividades humanas que afetam negativamente o âmbito social, ambiental e econômico. O processo de urbanização desordenado tem convertido grande parte da paisagem natural em diferentes usos do solo, acarretando degradação nos ecossistemas naturais, como é o caso da Mata Atlântica, importante bioma brasileiro (SANTOS et al., 2020; SOUZA, et. al., 2021). Entre essas degradações, têm-se o aumento do desmatamento para a implantação de atividades como a agricultura e pastagem, alteração da qualidade da água, assoreamento de rios e nascentes visto as ocupações irregulares em áreas de preservação ambiental. Logo, é evidente a urgência de estudos que se objetivam caracterizar e analisar a relação entre aspectos de expansão urbana e as condições ambientais sob bacias hidrográficas, visando uma gestão urbana integrada com os recursos naturais.

De acordo com Yanfei Zhong et al. (2018), o sensoriamento remoto vem sendo largamente utilizado em muitos campos, mas o sucesso de qualquer classificação de imagem depende de vários fatores, incluindo na escolha de um procedimento de classificação adequado. O estudo de uma abordagem multiescala beneficiada pelas técnicas fornecidas pelo Geoprocessamento tem sido uma eficaz ferramenta para lidar com as complexidades e incertezas em análise de imagem de sensoriamento remoto, o que permite melhores diagnósticos.

Nesse sentido, o presente trabalho apresenta uma comparação multitemporal de uso e cobertura do solo utilizando classificadores de Aprendizado de Máquina conhecidos como *Support Vector Machine* (SVM) e *Random Forest* (RF) para o mapeamento e identificação dos principais vetores de mudanças e seus impactos na sub-bacia hidrográfica do Ribeirão José Pereira, localizada em Itajubá – MG. A escolha da área se justifica pelas suas especificidades, sendo observado a jusante do ribeirão um processo acelerado de expansão urbana e, a montante, a Reserva Biológica Municipal da Serra dos Toledos (RBMST), um importante

remanescente florestal do bioma Mata Atlântica. Ademais, trata-se de uma sub-bacia com frequentes inundações, agravada por ocupação habitacional próxima ao curso d'água em razão da topografia local, que é caracterizada em sua maior parte como ondulada-montanhosa.

O mapeamento e verificação desses aspectos foi realizado com o auxílio do *Orfeo Toolbox* (OTB) no software livre QGIS, por meio da classificação de imagens multiespectrais dos satélites LANDSAT-8 e CBERS 04A em dois períodos distintos. Com isso, além da identificação do classificador com melhor acurácia na distinção das classes, foi possível realizar a identificação das classes com predomínio na bacia e quantificar os percentuais dos vetores de mudança para cada classe para a obtenção de resultados mais próximos da interpretação em escala de detalhes. Essas análises permitem orientar a tomada de decisões no sentido de ações para preservação ambiental na área, incentivando o subsídio do desenvolvimento e implementação de políticas públicas e tornando a população mais resiliente frente aos desastres naturais que ocorrem na área.

Metodologia

Primeiramente, realizou-se pesquisas relacionadas ao tema a fim de compreender a demanda do estudo e também, do local de estudo (Figura 1).

Figura 1 - Mapa de localização da área de estudo

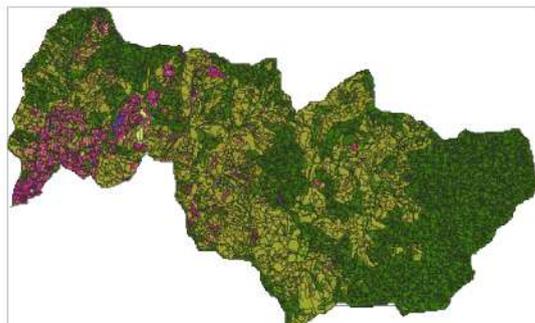


Logo após, deu-se início à etapa prática, a qual envolveu

a produção de mapas para estudo dos aspectos físicos e característicos da área por meio do software de geoprocessamento de acesso livre, QGIS versão 3.16.11. Os mapas são referentes à localização da área de estudo, à declividade, à hipsometria, ao Índice de Vegetação por Diferença Normalizada (NDVI) e, principalmente, ao uso e cobertura do solo, fazendo uma comparação temporal entre os anos de 2018 e 2022 por meio da classificação com os algoritmos de Aprendizado de Máquina, sendo eles: *Random Forest* (RF) e *Support Vector Machine* (SVM). Além disso, mediu-se a acurácia desses processos pelo Índice Kappa, calculado por meio do plugin “*Semi-Automatic Classification*” (SCP), utilizando a ferramenta “*Postprocessing*” e, posteriormente, “*Accuracy*”. Esse índice tem o intuito de quantificar o quão um método se sobressai ao outro, observando qual classificador retornou valor mais próximo de 1, sendo que, quanto mais próximo, melhor o resultado.

Para o mapeamento multitemporal do uso e cobertura da terra, utilizou-se imagens multiespectrais do satélite CBERS 04A, datadas de agosto de 2022, mais especificamente da Câmera Multiespectral e Pancromática de Ampla Varredura (WPM), devido sua alta resolução espacial, sendo esta igual a 8 metros. Para o ano de 2018, devido a impossibilidade de obter conjunto de imagens do mesmo satélite, optou-se pelas imagens do sensor OLI do satélite LANDSAT-8, obtidas em julho de 2018. Essas imagens possuem resolução espacial de 30 metros, porém, realizou-se o método de fusão com a banda pancromática, o que aumentou sua resolução para 15 metros. Com a mesclagem das bandas multiespectrais dessas imagens, gerou-se uma camada vetorial de composição de cor natural, o que facilitou o processo de classificação do uso e cobertura do solo. Dessa forma, com posse das imagens multiespectrais, tornou-se viável realizar a classificação com os algoritmos de Aprendizado de Máquina. O processo para realizar o mapeamento do uso e cobertura do solo com a técnica de RF e SVM são semelhantes, alterando apenas algumas funções específicas. Essas técnicas envolveram, basicamente, as seguintes etapas: mesclagem das imagens multiespectrais; segmentação; cálculo de estatísticas zonais, caracterização das classes, amostragem, extração de atributos, treinamento do classificador, classificação, validação e pós-classificação. Na etapa de segmentação, utilizou-se a ferramenta “*Segmentation*”, que se encontra no complemento OTB e, mais especificamente, na seção *Learning*. Essa etapa retornou um arquivo vetorial, o qual divide a área de estudo em diversos segmentos, o que facilita a classificação visto que a imagem encontra-se fragmentada em objetos de acordo com os

critérios de segmentação, como mostrado na Figura 2. Figura 2 - Exemplo do resultado da segmentação realizada pela ferramenta “*segmentation*”



A partir dos segmentos, extraiu-se estatísticas como média, máximo, mínimo e desvio padrão para cada um dos segmentos por meio da ferramenta *Zonal Statistics*, também do OTB. Dessa forma, os cálculos foram feitos de acordo com os valores de reflectância das bandas espectrais da composição colorida natural.

Na etapa de extração de atributos e caracterização das classes, existem dois modos para a extração de atributos: criando uma camada vetorial de polígonos para as amostras ou selecionando os polígonos gerados por meio da segmentação.

Na etapa de treinamento da imagem, utilizou-se a ferramenta *TrainImageClassifier* do OTB. Adotou-se, separadamente, ambos os classificadores: SVM e RF; Esses algoritmos possuem a função de realizar uma determinação ótima de limites de separação das classes, minimizando os erros. Por fim, para realizar a classificação do uso e cobertura do solo propriamente dita, utilizou-se a ferramenta *VectorClassifier*, inserindo o modelo e as estatísticas que foram criadas.

No processo de validação, utilizou-se a ferramenta “*Seleção aleatória por subconjuntos*”. Essa função permite fazer uma seleção de um número igual de amostras por classe, que se separam em grupos de treino e validação. Adotou-se 70% dessas amostras para o treinamento do algoritmo e 30% para a validação da classificação. Esse processo é crucial na utilização do Aprendizado de Máquina, pois permite que as classes fiquem balanceadas.

A etapa de pós-classificação é necessária para corrigir eventuais confusões de classes, dada a proximidade espectral de algumas classes.

Resultados e discussão

A partir dos resultados expostos no mapa do Índice de Vegetação por Diferença Normalizada (NDVI), para o ano de 2022 (Figura 3), nota-se que a sub-bacia em estudo possui maior densidade de vegetação ao leste, explicado pela presença de mata nativa referente à

RBMST.. Em 2018 (Figura 4), nota-se uma densidade de vegetação na mesma área do resultado de 2022, porém em uma escala menor. Essa variação pode ter ocorrido devido a diferença da resolução espacial entre os satélites CBERS 04A e LANDSAT-8, visto que apresentam uma resolução de 8 metros e 15 metros, respectivamente. No caso de feições antrópicas e solo exposto, ambos os resultados mostraram adequados para sua separação e identificação.

Figura 3 – Mapa NDVI para o ano de 2022

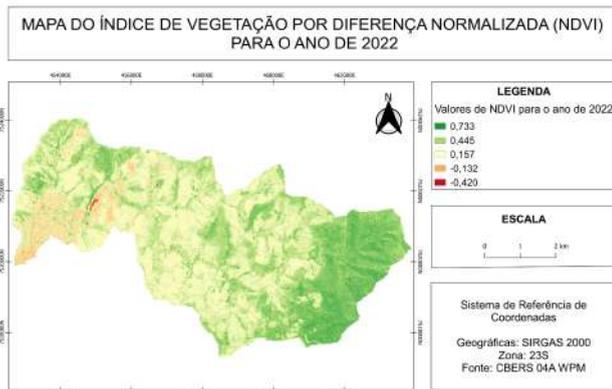
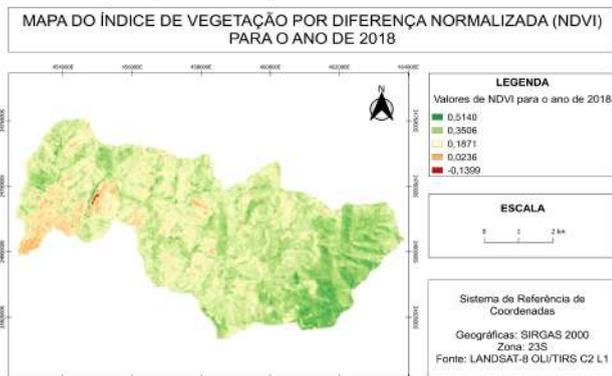


Figura 4 – Mapa NDVI para o ano de 2018



As Figuras 5, 6, 7 e 8 apresentam os resultados da classificação gerados a partir dos classificadores SVM e RF para os anos de 2022 e 2018, respectivamente.

Figura 5 – Mapa do uso e cobertura do solo da área de estudo utilizando o algoritmo SVM para o ano de 2022

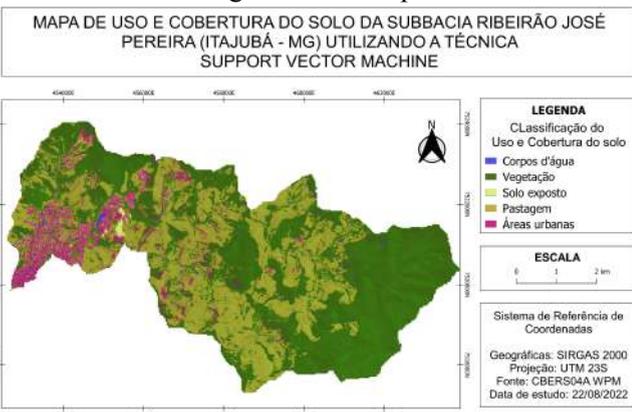


Figura 6 – Mapa do uso e cobertura do solo da área de estudo utilizando o algoritmo RF para o ano de 2022

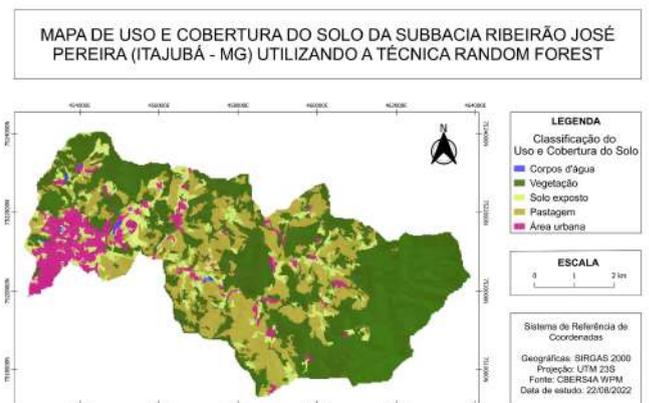


Figura 7 – Mapa do uso e cobertura do solo da área de estudo utilizando o algoritmo SVM para o ano de 2018

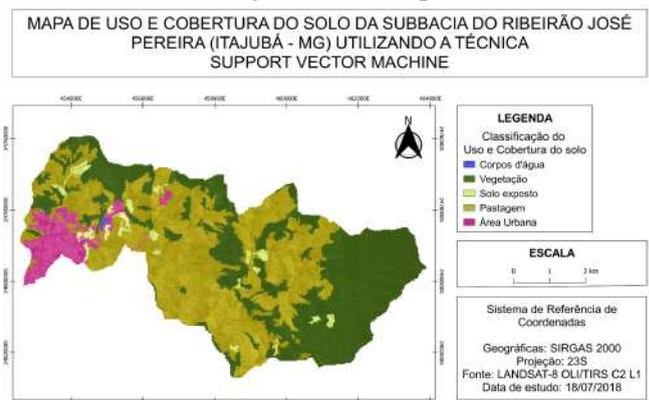
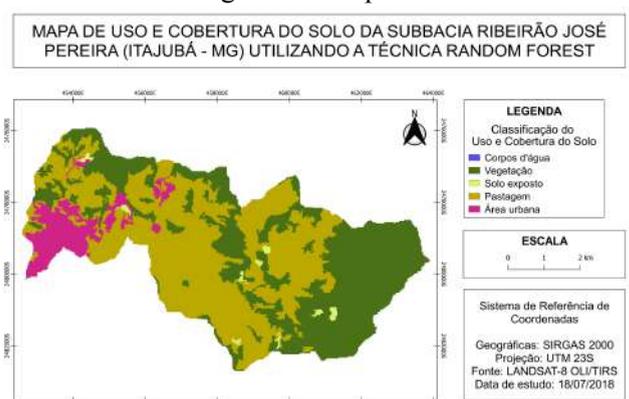


Figura 8 – Mapa do uso e cobertura do solo da área de estudo utilizando o algoritmo RF para o ano de 2018



A partir da análise visual, é possível verificar que o classificador SVM conseguiu distinguir melhor as classes e tornou o resultado mais nítido e próximo da realidade da área de estudo, ressaltando detalhes como, por exemplo, pequenas construções em área de pastagem, a passagem do Ribeirão José Pereira e o solo exposto presente em determinadas áreas. Já a classificação resultante do classificador RF, nota-se que a imagem possui menos especificidade, o que mostra

que ocorreu menor distribuição e distinção das classes. Além disso, as análises foram confirmadas pela avaliação quantitativa por meio do cálculo do Índice Kappa. O classificador SVM obteve o maior valor do índice Kappa com 0,9406 em 2022 e 0,9656 em 2018, o que indica uma excelente classificação (LANDIS & KOCH, 1977), enquanto o classificador pelo RF alcançou um Índice Kappa de 0,7774 em 2022 e 0,9660 em 2018, o que mesmo sendo uma classificação muito boa, acabou sendo inferior ao SVM, que proporcionou diagnósticos mais eficazes.

Conclusões

Os resultados apresentados constataram que os métodos utilizados na confecção dos mapas de Uso e Cobertura do solo apresentam um desempenho consideravelmente eficiente, entretanto, é notório que um deles concede um diagnóstico mais próximo da realidade: algoritmo/classificador SVM.

Ademais, sobre os resultados dos mapas de Uso e Cobertura do solo no que se refere à análise temporal entre os anos 2018 e 2022, nota-se um processo de urbanização e degradação ambiental ao longo do tempo, visto que ocorre uma diminuição da área de vegetação e um aumento da área de pastagem. Foi possível comprovar essa afirmação a partir da observação do mapa NDVI.

Seguindo essa linha de raciocínio, ainda que o algoritmo SVM apresente um desempenho mais confiável, é possível minimizar o problema da confusão espectral observada no algoritmo RF, realizando uma alteração manual na extração e classificação das amostras que foram aplicadas no treinamento desse método.

Além disso, ao se comparar os resultados dos NDVI's para os anos de estudo, é importante ressaltar que algumas diferenças ocorreram devido à diferença do satélite que foi usado para as classificações, além de ser necessário levar em consideração que imagens mais atuais possuem maior acurácia nos resultados, visto o avanço tecnológico na área.

Em síntese, é importante ressaltar a importância desse ramo de pesquisa para o processo de aprendizagem acadêmica e profissional, visto que é crucial compreender os impasses que o crescimento urbano pode causar na preservação ambiental de determinada área. Esse tipo de estudo orienta as decisões e incentiva o desenvolvimento de políticas públicas com olhares mais atentos para o desenvolvimento ordenado e sustentável.

Agradecimento

Agradeço o apoio recebido pelo Programa Institucional de Bolsas de Iniciação Científica (PIBIC) e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). À Universidade Federal de Itajubá, agradeço a oportunidade de aprendizado. À minha orientadora, Samara Calçado Azevedo, sou grata pela confiança e por todo apoio no desenvolvimento da pesquisa.

Referências

ABADIAS, I.M.; FONSECA, P.R.B.; BARBOSA, C.H.; SALES, M.C.G. Identificação dos principais impactos ambientais ocasionados pelo manejo da pecuária no município de Humaitá - AM. **Revista Gestão & Sustentabilidade Ambiental**, v. 9, p. 664-682, 2020.

BELGIU, M.; DRAGUT, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, USA, n. 114, p. 24 - 31, 2016.

EXBRAYAT, J.-F.; LIU, Y.Y.; WILLIAMS, M. Impact of deforestation and climate on the Amazon Basin's aboveground biomass during 1993-2012. **Scientific Reports**, v.7, 2017.

LANDIS, J.R.; KOCH, G.G. The measurement of observer agreement for categorical data. **Biometrics**, v.33, n.1, p. 159-174, 1977.

MELLO, K. et al. Multiscale land use impacts on water quality: Assessment, planning, and future perspectives in Brazil. **Journal of Environmental Management**, v. 270, 2020.

SANTOS, L. D.; SCHLINDWEIN, S. L.; FANTINI, A. C.; HENKES, J. A.; BELDERRAIN, M. C. N. Dinâmica do desmatamento da Mata Atlântica: causas e consequências. **R. Gest. Sust. Ambient.**, v. 9, n. 3, p. 378-402, 2020b

SOUZA *et. al.* Long-term ecological trends of small secondary forests of the atlantic forest hotspot: A 30-year study case. **Forest Ecology and Management**. n. 489, 106778. <https://doi.org/10.1016/j.buildenv.2020.106778>

ZHONG, Y. et. al. Computational intelligence in optical remote sensing image processing. **Applied Soft Computing**, França, n. 64, p. 75-93, 2018.