

CLASSIFICAÇÃO E SEGMENTAÇÃO DE PLANTAS DANINHAS EM PASTAGENS CULTIVADAS UTILIZANDO O ALGORITMO RANDOM FOREST

Amanda Teixeira Moreira Capacia, Paula Alberti Bonadiman, Thiago Alexandre da Silva Oliveira, Maria Eduarda Carolo Freitas, Caroline Guedes Rezende Nunes, Samuel de Assis Silva

Universidade Federal do Espírito Santo, Alto Universitário, Cx.P 16, Guararema – 29.500-000 – Alegre-ES, Brasil, amandacapacia@gmail.com, bonadimanpaula@gmail.com, thiagoalexandre984@gmail.com, meduardacarolo@hotmail.com, carolinegrnunes@gmail.com, samuel.silva@ufes.br

Resumo

Com o crescimento populacional, a produção de alimentos precisa ser aumentada. As plantas daninhas são um dos principais fatores bióticos que competem com as culturas agrícolas por recursos essenciais. O sensoriamento remoto é uma tecnologia amplamente utilizada para detectar e mapear plantas daninhas, visando reduzir perdas e promover um manejo mais sustentável. Neste estudo, foram utilizadas técnicas de sensoriamento remoto e aprendizado de máquina para identificar e classificar automaticamente a incidência de plantas daninhas em áreas de pastagem. Foram coletadas imagens aéreas com sensores óticos RGB e multiespectral e aplicado o algoritmo de aprendizado supervisionado *Random Forest* (RF) para a classificação. O algoritmo RF reportou bons resultados no processo de classificação, sendo capaz de segmentar as áreas infestadas com plantas daninhas, separando-as da pastagem cultivada.

Palavras-chave: Inteligência artificial. Sensoriamento remoto. Classificação supervisionada. Agricultura de Precisão.

Área do Conhecimento: Engenharia Agrônômica.

Introdução

O crescimento populacional mundial segue acelerado e com ele o aumento na demanda por alimentos e energia, o que impulsiona a necessidade do aumento da produção agrícola (YOUNG e PIERCE, 2014). Dentre essa necessidade, atrela-se a minimização de perdas que são geradas por diversos fatores.

Dentre os geradores de perdas, estão as plantas daninhas. Quando presentes, exercem uma competição por recursos como água, luz, espaço e nutrientes devido à sua alta rusticidade além de servirem como hospedeiras alternativas de pragas e doenças (GHANIZADEH e HARRINGTON, 2019; ZIMDAHL, 2018; JHA et al., 2017). Para aumentar a eficiência das medidas de controle das plantas daninhas é necessário que se conheça, dentre outros fatores, as espécies presentes no campo e, bem como, sua densidade, distribuição e abrangência ao longo do espaço (FERREIRA et al., 2019; SOUZA et al., 2020).

O sensoriamento remoto, especialmente o aéreo, tem sido amplamente utilizado para detectar e mapear plantas daninhas. Essa técnica tem se destacado nos últimos anos (REEDHA et al., 2022; ROSLIM et al., 2021; LAN et al., 2021) devido à sua segurança, baixo custo e alta resolução espacial. Diversos autores tem reportado a maior efetividade em classificar plantas daninhas associando sensoriamento remoto aéreo e inteligência computacional (HASAN et al. 2021; ASSAD et al., 2020; SHAWKY et al., 2020).

Apesar das diversas pesquisas realizadas nos últimos anos com essa temática, poucos são os estudos que, associando o uso de *machine learning* e sensoriamento remoto aéreo, tem proposto a classificação e segmentação de plantas daninhas em ambientes produtivos e complexos. Objetivou-se com este trabalho utilizar técnicas de sensoriamento remoto e *machine learning*, através do algoritmo *Random Forest*, para a identificação e classificação automatizada da incidência de plantas daninhas em áreas cultivadas com pastagem.

A era digital e suas implicações sociais: Desafios e contribuições

Metodologia

O estudo foi realizado no município de Alegre, localizado no estado do Espírito Santo. A região é classificada como "Cwa", com clima tropical quente-úmido, temperatura média anual de 23,1°C, inverno frio e seco e precipitação média anual de 1341mm (Köppen e Geiger, 1928). O experimento foi conduzido em uma área plana, cultivada a cinco anos com capim-tanzânia (*Panicum maximum* cv. Tanzânia) sob diferentes manejos.

Os levantamentos aéreos foram realizados utilizando: a) um sensor óptico RGB CMOS equipado com uma lente FOV 84° 8,8mm/24mm que captura os segmentos dos comprimentos de onda do espectro eletromagnético visível RGB, onde: R – Red (vermelho); G – Green (verde) e B – Blue (azul), e; b) um sensor óptico multiespectral RedEdge MX com lentes de 47,9° HFOV de 5,4mm que capturam segmentos dos comprimentos de onda do espectro eletromagnético visível, rededge e infravermelho próximo (*Near Infrared* - NIR). Os sensores foram embarcados em veículos aéreos não tripulados (VANT's) multirotores, sendo um do modelo Phantom 4 Pro com o sensor RGB por padrão e outro do modelo Inspire 2, embarcado com o sensor multiespectral. Os voos foram realizados entre as 11h às 13h e a uma altura máxima de 80m, número variável de imagens, a velocidade (10 m.s⁻¹) e a sobreposições laterais (80%) e horizontais (80%) foram definidos após testes preliminares.

Foram realizadas calibrações radiométricas das imagens a fim de anular quaisquer distorções causadas por aspectos internos da câmera. As imagens foram processadas para a obtenção de ortomosaicos usando o software OpenDroneMap. Foram utilizados 15 pontos de controle fotointerpretáveis distribuídos por toda a extensão da área. Os pontos foram georreferenciados utilizando um receptor GNSS com precisão subcentimétrica.

Para a identificação das plantas daninhas *in loco*, foram determinadas a incidência e a área de abrangência de cada planta. Em seguida, elas foram identificadas e foram georreferenciadas para a obtenção de suas coordenadas planimétricas de latitude e longitude, utilizando-se um receptor GNSS de precisão subcentimétrica. Ao todo, foram 48 pontos coletados. Posteriormente, foram coletadas amostras de cada planta daninha identificada e estas foram classificadas com o auxílio do Manual de Identificação e Controle de Plantas Daninhas (LORENZI et al., 2014).

Para a automatização, a identificação das plantas daninhas foi obtida por sensoriamento remoto através da resposta espectral do dossel. Foi utilizado o modelo de *machine learning Random Forest* (RF) aplicado para a reflectância das plantas em bandas individualizadas.

O algoritmo foi executado utilizando as ferramentas do diretório Orfeo Toolbox, implementados em ambiente Sistema de Informação Geográfica (SIG) utilizando o QGIS. Para tal, a sequência constituiu na construção do *dataset*, selecionando polígonos característicos de cada planta daninha, da pastagem e de solo exposto na área – *features*. Os modelos foram treinados, respeitando a proporção de 75% das amostras para treinamento e os demais 25% para teste.

Após o treinamento, as imagens foram classificadas para identificação das plantas daninhas e sua segmentação em folha estreita e folha larga. As *features* foram divididas em quatro classes distintas, onde: C1 – Plantas daninhas de folha larga; C2 – Plantas daninhas de folha estreita; C3 – Pastagem cultivada com *P. maximum* cv. Tanzânia e C4 – Solo exposto. Ao total, foram feitas 39 amostras para cada classe.

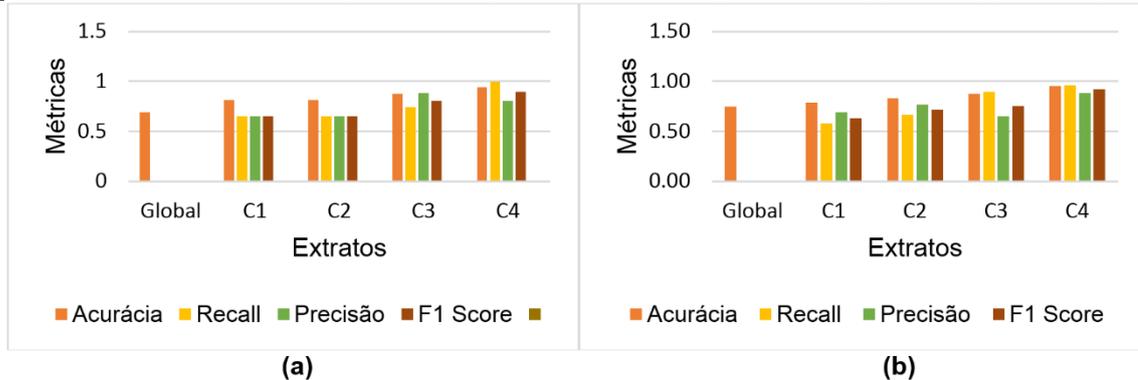
A determinação do desempenho da classificação foi realizada a partir do cálculo de diferentes métricas. A adoção de mais de uma métrica para avaliar o algoritmo torna-se necessária, pois diversas vezes ele pode expressar bons valores para uma determinada métrica e para outras não (LUQUE et al., 2019). Devido ao impacto gerado no desequilíbrio dos valores resultantes nos cálculos, há a necessidade da visualização do desempenho da classificação sob diferentes pontos de vista. A partir da matriz de confusão gerada na classificação, foram calculadas as seguintes métricas: acurácia, precisão, recall e F1-score. A matriz de confusão é uma matriz bidimensional que retrata o desempenho do classificador ao correlacionar o rótulo real e o rótulo previsto (ALI et al., 2022).

A era digital e suas implicações sociais: Desafios e contribuições

Resultados

As métricas utilizadas para avaliar o desempenho do classificador RF estão representadas na Figura 1.

Figura 1. Métricas da classificação das imagens utilizando o algoritmo RF. (a) RGB; (b) Multiespectral.



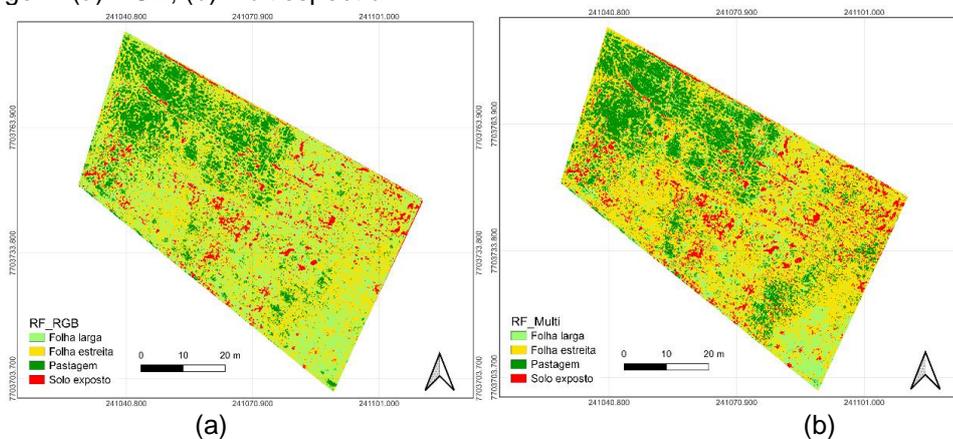
Fonte: os autores.

As figuras representam os valores obtidos a partir do cálculo das métricas pelos valores gerados pela matriz de confusão. A acurácia indica a fração de predições corretas resumindo as *features* que realmente foram classificadas corretamente independentemente da classe (LI et al., 2019). O recall, expressa a proporção de acertos dentro do que de fato era verdadeiro. A precisão considera as variações das amostras em cada classe e não considera os verdadeiros negativos. O F1-score atribui uma igual importância entre a precisão e o recall. Os valores resultantes dos cálculos variam entre 0 e 1, onde 1 considera-se 100% de acerto dentro daquilo que é proposto pela métrica.

Para cada classe, foram calculadas as métricas gerando os extratos representados nas figuras de forma conjunta. Apenas a acurácia foi calculada de forma global, à fim de obter-se melhor discussão sobre os resultados.

Os mapas classificados por RF a partir das classes avaliadas estão representados na Figura 2.

Figura 2. Mapas classificados por RF a partir de quatro classes de cobertura do solo em área cultivada com pastagem. (a) RGB; (b) Multiespectral.



Fonte: os autores.

Ambos os mapas são representados por 4 classes diferentes, cujas cores compreendem cada uma delas. Para C1 (plantas daninhas de folha larga), a cor é verde claro; C2 (plantas daninhas de folha estreita), amarelo; C3 (pastagem), verde escuro e C4 (solo exposto), vermelho. O primeiro mapa (a) foi gerado a partir da imagem RGB e o segundo mapa (b), a partir da imagem Multiespectral.

A era digital e suas implicações sociais: Desafios e contribuições

Discussão

Observando os valores das métricas obtidas (Figura 1) pelo algoritmo em C1, a acurácia é superior na imagem RGB, em contrapartida, a precisão na imagem multiespectral foi maior. O recall e o F1 score da imagem RGB foram ambos de 0,65 e da imagem multiespectral, os valores foram de 0,58 e 0,63 respectivamente.

Em C2, a acurácia de ambas as imagens foi superior à 0,82. A precisão, o recall e o F1 score da imagem multiespectral foram superiores. Para C3, a acurácia e a precisão de ambas as imagens foram superiores à 0,86. O recall foi maior na imagem multiespectral. Por outro lado, o F1 score foi superior na imagem RGB, com valor de 0,80 enquanto na imagem multiespectral foi 0,75. Analisando C4, podemos observar que a acurácia e a precisão da imagem multiespectral foram maiores quando comparadas à imagem RGB. O recall na imagem RGB atingiu valor 1 e na imagem multiespectral foi superior à 0,91. O F1 score para ambas as imagens foi superior à 0,89. Os valores das métricas obtidos nesta classe indicam uma melhor classificação quando se compara aos obtidos nas classes anteriores. A acurácia global para a imagem RGB foi de 0,74 e para a imagem multiespectral foi de 0,75.

De maneira geral, os melhores resultados para C1 e C2 foram obtidos utilizando a imagem multiespectral, apesar de a acurácia ter sido maior para a imagem RGB em C1. Em C3, a imagem RGB se sobressaiu, pois apesar do valor do recall ter sido maior para a imagem multiespectral, os valores da precisão e do F1 score foram maiores para a imagem RGB. Em C4, a melhor classificação foi na imagem multiespectral, apesar da ausência de falsos negativos na imagem RGB, o valor da precisão e do F1 foram maiores para a imagem multiespectral. As imagens multiespectrais, por carregar informações de reflectância além do espectro eletromagnético do visível, é capaz de descrever melhor as diferenças entre objetos em solo (MEDAUAR et al., 2020).

Ao analisar os mapas (Figura 2), é possível observar melhor desempenho do classificador para a imagem multiespectral (b). Além de ter sido classificada mais pastagem quando comparada à imagem RGB, apresentou melhor identificação para plantas daninhas de folha larga. Ao identificar maior quantidade de plantas daninhas de folha estreita, é possível relacionar o fato da semelhança das mesmas com a pastagem e, portanto, um melhor desempenho do classificador, isto porque, de acordo com Li et al. (2021), as espécies de folhas largas são ligeiramente mais simples de identificar do que as espécies de folhas estreitas devido ao fato de coleta de dados espectrais mais uniformes, fato este que não se observa na imagem RGB, que classificou mais plantas daninhas de folha larga e que, por estas plantas apresentarem uma disposição e arquitetura de folhas mais distintas quando relacionadas com a pastagem, faz com que a classificação seja mais errônea.

Em linhas gerais os resultados apontam para a possibilidade de uso do RF para a classificação de plantas daninhas em áreas cultivadas com pastagem. Flecher (2016) aplicando RF para a classificação de plantas daninhas em áreas de cultivo de soja, encontrou resultados semelhantes ao desse trabalho. Segundo esse autor, o desempenho do algoritmo é semelhante utilizando as bandas espectrais (conforme utilizado nessa pesquisa) ou a partir de índices de vegetação.

Conclusão

Foi possível classificar de forma automatizada plantas daninhas em área cultivada com pastagem, bem como separá-las de acordo com sua arquitetura em plantas de folha larga e folha estreita. O algoritmo RF reportou bons resultados no processo de classificação, sendo capaz de segmentar as áreas infestadas com plantas daninhas, separando-as da pastagem cultivada.

Referências

ALI, K.; SHAIKH, Z. A.; KHAN, A. A.; LAGHARI, A. A. Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer. **Neuroscience Informatics**, v. 2, n. 4, p. 100034, 2022.

ASSAD, M. H.; BAIS, A. Weed detection in canola Fields using maximum likelihood classification and deep convolutional neural network. *Information Processing in Agriculture*, v. 7, n. 4, p. 535-545, 2020.

A era digital e suas implicações sociais: Desafios e contribuições

BREIMAN, L. Random forests. **Machine learning**, v. 45, p. 5-32, 2001.

FERREIRA, E. A.; PAIVA, M. C. G.; PEREIRA, G. A. M.; OLIVEIRA, M. C.; SILVA, E. B. Fitossociologia de plantas daninhas na cultura do milho submetida à aplicação de doses de nitrogênio. *Revista de Agricultura Neotropical*, v. 6, n. 2, p. 100-107, 2019.

FLETCHER, R. S. Using vegetation indices as input into Random forest for soybean and weed classification. **American Journal of Plant Sciences**, v. 7, n. 15, p. 2186- 2198, 2016.

GHANIZADEH, H.; HARRINGTON, K. C. Weed management in New Zeland pastures. *Agronomy*, v. 9, n. 8, p. 448, 2019.

HASAN, A. S. M. M.; SOHEL, F.; DIEPEVEEN, D.; LAGA, H.; JONES, M. G. K. A survey of deep learning techniques for weed detection from images. *Computers and Electronics in Agriculture*, v. 184, p. 106067, 2021.

JHA, P.; KUMAR, V.; GODARA, R. K.; CHAUHAN, B. S. Weed managment using crop competition in the United States: A review. *Crop Protection*, v. 95, p. 31-37, 2017.

LAN, Y.; HUANG, K.; YANG, C.; LEI, L.; YE, J.; ZHANG, J.; DENG, J. Real-time identification of rice weeds by UAV low-altitude remote sensing based on improved semantic segmentation model. *Remote Sensing*, v. 13, n. 21, p. 4370, 2021.

LI, W.; LIU, H.; WANG, Y.; LI, Z.; JIA, Y.; GUI, G. Deep learning-based classification methods for remote sensing images in urban built-up areas. **IEEE Access**, v. 7, p. 36274-36284, 2019.

LI, Y.; AL-SARAYREH, M.; IRIE, K.; HACKELL, D.; BOURDOT, G.; REIS, M. M.; GHAMKHAR, K. Identification of weeds based on hyperspectral Imaging and machine learning, **Frontiers in Plant Science**, v. 11, p. 611622, 2021.

LORENZI, H.; NICOLAI, M.; BIANCHI, P. J.; GUIMARÃES, S. C. **Manual de Identificação e Controle de Plantas Daninhas**. 2014.

LUQUE, A.; CARRASCO, A.; MARTÍN, A.; LAS HERAS, A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. **Pattern Recognition**, v. 91, p. 216-231, 2019.

MEDAUAR, C.C.; SILVA, S.A.; CARVALHO, L.C.C.; TIBÚRCIO, R.A.S.; LIMA, J.S.S.; OLIVEIRA, P.S. Eradication of eucalyptus sprouts after chemical weeding over time in State of Bahia, Brazil. **Anais da Academia Brasileira de Ciências**, v. 92, p. e20190601, 2020.

REEDHA, R.; DERICQUEBOURG, E.; CANALS, R.; HAFIANE, A. Transformer Neural Network for Weed and Crop Classification of High Resolution UAV Images. **Remote Sensing**, v. 14, n. 3, p. 592, 2022.

ROSLIM, M. H. M.; JURAIMI, A. S.; CHE'YA, N. N.; SULAIMAN, N.; MANAF, M. N. H. A.; RAMLI, Z.; MOTMAINNA, M. Using remote sensing and na unmanned aerial system for weed management in agricultural crops: A review. **Agronomy**, v. 11, n. 9, p. 1809, 2021.

SHAWKY, O. A.; HAGAG, A.; EL-DASHAN, E. S. A.; ISMAIL, M. A. Remote sensing image scene classification using CNN-MLP with data augmentation. **Optik**, v. 221, p. 165356, 2020.

SOUZA, M. F.; SILVA, T. S.; DOS SANTOS, J. B.; CARNEIRO, G. D. O. P., REGINALDO, L. T. R. T.; BANDEIRA, J. N.; SILVA, D. V. Soil water availability alter the weed Community and its interference on onion crops. **Scientia Horticulturae**, v. 272, p. 109573, 2020.

YOUNG, S.; PIERCE, J. **Automation: The Future of Weed Control in Cropping Systems**. Springer Science & Business Media, 2013.

ZIMDAHL, R. L. **Fundaments of weed Science**. Academic press. 2018.