

A era digital e suas implicações sociais: Desafios e contribuições

CLASSIFICAÇÃO E SEGMENTAÇÃO DE PLANTAS DANINHAS EM PASTAGEM CULTIVADA UTILIZANDO O ALGORITMO k-NEAREST NEIGHBORS

Amanda Teixeira Moreira Capacia, Thiago Alexandre da Silva Oliveira, João Mendes Cicarini Hott, Maísa Martins de Oliveira, Isaías Brinati Valentim, Samuel de Assis Silva

Universidade Federal do Espírito Santo, Alto Universitário, Cx.P 16, Guararema – 29.500-000 – Alegre-ES, Brasil, amandacapacia@gmail.com, thiagoalexandre984@gmail.com, joaoagro18@hotmail.com, maisamartins20@gmail.com, isaiasbrinativalentim@gmail.com, samuel.silva@ufes.br.

Resumo

As plantas daninhas estão entre os principais fatores bióticos geradores de perdas em cultivos agrícolas. Com isso, um bom monitoramento de plantas daninhas torna-se necessário para minimizar os prejuízos por elas causados. Durante os últimos anos, a agricultura de precisão tem papel de destaque no monitoramento dessas plantas, com tecnologias como o sensoriamento remoto em conjunto com o aprendizado de máquinas. Neste estudo, foram utilizadas técnicas de sensoriamento remoto e classificação supervisionada para a identificação e segmentação de plantas daninhas em pastagens cultivadas, com uso do algoritmo *k-Nearest Neighbors* (k-NN) a partir de imagens obtidas por sensores RGB e multiespectral. O algoritmo apresentou bons resultados e foi capaz de identificar e classificar as plantas daninhas, distinguindo-as da pastagem cultivada e do solo exposto.

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

Área do Conhecimento: Engenharia Agrônoma.

Introdução

Dentre as principais ameaças bióticas que causam prejuízos aos cultivos agrícolas, estão as plantas daninhas (Esposito *et al.*, 2018), que rigorosamente competem por recursos como água, luz e nutrientes (Korav *et al.*, 2018). Com o crescente aumento da população mundial, o monitoramento para o controle das plantas daninhas será um grande desafio (Westwood *et al.*, 2018).

O monitoramento de plantas daninhas de forma convencional apresenta alguns pontos negativos como alto custo e trabalho intensivo, dificuldade de identificação, alta frequência de monitoramento, limitações sazonais e tempo de resposta (Rosle *et al.*, 2021; Roslim *et al.*, 2021).

O sensoriamento remoto é uma das tecnologias mais aplicadas na agricultura de precisão e é altamente eficaz no monitoramento e manejo de plantas daninhas (Hunter *et al.*, 2020; John *et al.*, 2018). O planejamento mais preciso aumenta a eficácia de métodos de controle químico reduzindo a disseminação de herbicidas e consequentemente gerando menos impacto ao meio ambiente (López-Granados *et al.*, 2016; Ray *et al.*, 2013), além de ser uma prática segura, economicamente viável e de elevada resolução espacial (Lan *et al.*, 2021). A alta resolução espacial contribui para a identificação de plantas daninhas de forma mais precisa e coesa nas pastagens, visto que atributos como distribuição de touceiras e a arquitetura das plantas pode dificultar o processo quando este é realizado de forma convencional.

A obtenção de maior precisão na classificação de alvos, especialmente no estudo de cobertura do solo por meio de sensoriamento remoto, como no caso das plantas daninhas, pode ser alcançada ao combinar imagens aéreas com modelos de inteligência computacional. Essa associação permite aprimorar significativamente a capacidade de identificação e classificação dos elementos em análise. Nos últimos anos, o aprendizado de máquinas a partir de algoritmos foi capaz de oferecer maior desempenho de áreas de produção a partir da gestão operacional agrícola no que diz respeito de identificar, medir e analisar grandes quantidades de dados e/ou dados complexos (Roslim *et al.*, 2021; Lan *et al.*, 2021). Dentre os algoritmos mais utilizados estão o *Random Forest* (De Casto *et al.*, 2018),

A era digital e suas implicações sociais: Desafios e contribuições

Support Vector Machine (Zhao et al., 2022) e *k-Nearest Neighbors* (kNN). O kNN, modelo introduzido por Cover e Hart em 1967, determina de maneira eficiente o rótulo de classe das amostras utilizando o parâmetro “k” do tamanho da vizinhança e a maioria simples de votos entre os “k” vizinhos mais próximos (Gou et al., 2019). Diversos autores vêm trabalhando com a utilização deste algoritmo na identificação de plantas daninhas (Alulema et al., 2023; Imanni et al., 2023; Islam et al., 2021).

Embora tenham sido conduzidas várias pesquisas nos últimos anos sobre esse tema, poucos estudos têm se dedicado à combinação do uso de *machine learning* e sensoriamento remoto aéreo para propor a classificação e segmentação de plantas daninhas em ambientes produtivos e complexos. O objetivo deste trabalho foi utilizar técnicas de sensoriamento remoto e *machine learning*, com ênfase no algoritmo *k-Nearest Neighbors*, para automatizar a identificação e classificação da incidência de plantas daninhas em áreas cultivadas com pastagem.

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 as 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 subcentimétrico. 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 k-Nearest Neighbor* (k-NN) 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

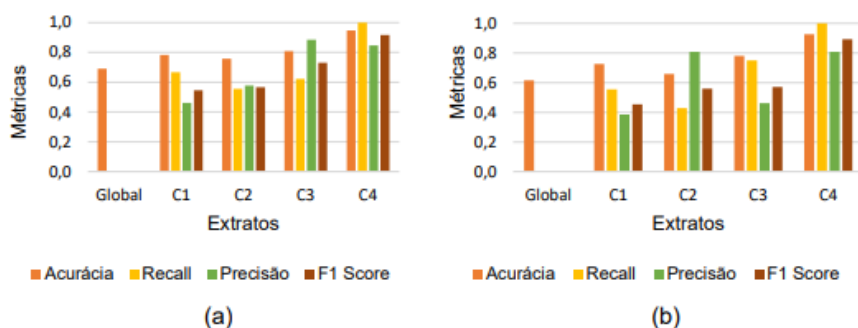
A era digital e suas implicações sociais: Desafios e contribuições

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 acurácia mede a fração de predições corretas, independentemente da classe, englobando todas as *features* que foram classificadas corretamente; a precisão verifica a proporção de acertos dentro daquilo que o classificador previu como verdadeiro; o *recall* se concentra na proporção de acertos dentro do que realmente é verdadeiro e o F1-score, atribui uma igual importância a precisão e ao recall, dado pela média ponderada entre as duas.

Resultados

As métricas utilizadas para avaliar o desempenho do classificador k-NN estão representadas na Figura 1.

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



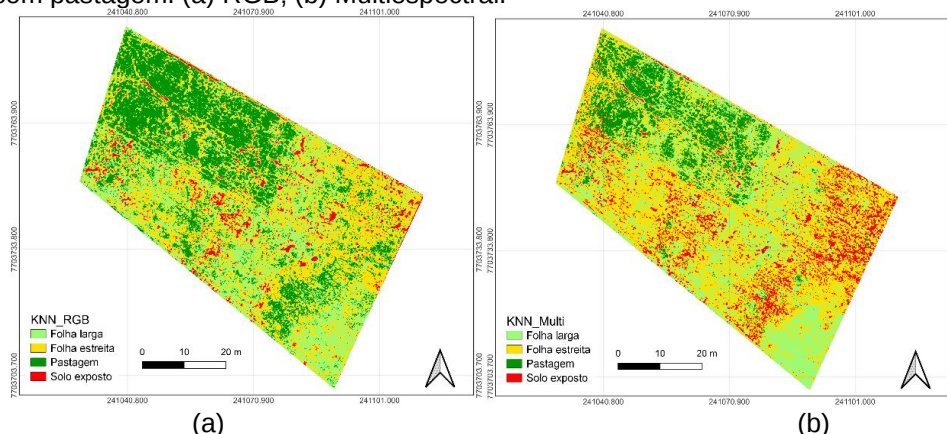
Fonte: os autores.

As figuras apresentam os valores das métricas calculadas a partir da matriz de confusão. Para cada classe, foram calculadas as métricas, gerando os gráficos (a e b) que estão representados de forma conjunta na figura. As métricas calculadas foram acurácia, *recall*, precisão e F1-score para que se obtenha diferentes pontos de vista do desempenho do classificador; e estas são calculadas a partir do número de acertos e/ou erros, dentro do que cada métrica se propõe a avaliar. Os valores resultantes variam entre 0 e 1, onde 1 corresponde a 100% de acerto. A acurácia, além de calculada para cada classe, também foi calculada de forma global, para proporcionar uma análise mais completa sobre os resultados obtidos.

Os mapas classificados por k-NN a partir das classes avaliadas estão representados na Figura 2.

A era digital e suas implicações sociais: Desafios e contribuições

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



Fonte: os autores.

Os mapas acima são representados por quatro classes distintas, sendo essas diferenciadas por cores. As plantas daninhas de folha larga (C1), são representadas pela cor verde claro; as de folha estreita (C2), amarelo; a pastagem cultivada (C3), verde escuro e o solo exposto (C4), vermelho. O mapa (a) foi gerado a partir de uma imagem RGB e o mapa (b), a partir de uma imagem multiespectral.

Discussão

Analisando os valores obtidos nos cálculos das métricas (Figura 1) pelo algoritmo individualmente por cada classe, em C1, os melhores valores são oriundos da imagem RGB. Em C2, somente a precisão obteve resultado superior e significativo na imagem multiespectral. O *recall* e o F1-score da imagem RGB foram superiores a 0,55 e da imagem multiespectral, obtiveram valores de 0,42 e 0,56 respectivamente. Em C3, a imagem RGB apresentou valores de acurácia e de precisão superiores a 0,8. Somente o *recall* na imagem multiespectral foi maior.

Em C4, a acurácia foi significativa para as duas imagens sendo superior a 0,9 e a precisão superior a 0,8; o *recall* obteve valor 1 para as duas e o F1 score foi superior na imagem RGB, com valor acima de 0,9 enquanto na imagem multiespectral o valor foi de 0,89. Estes valores indicam uma melhor interpretação da realidade por parte do classificador. A acurácia global da imagem RGB foi de 0,69 e da imagem multiespectral foi de 0,62.

De forma geral, os melhores resultados para todas as classes provêm da classificação utilizando a imagem RGB. Em C1, os valores foram superiores para todas as métricas calculadas; em C2, apesar da precisão ter sido superior para a imagem multiespectral, o valor do *recall* foi mais baixo e consequentemente, o F1-score também; para C3, a imagem RGB se sobressai, devido à diferença discrepante da precisão quando comparada à imagem multiespectral, que, apesar de ter obtido uma menor quantidade de falsos positivos, ainda assim, apresenta uma precisão inferior. Os resultados para C4 foram bons para ambas as imagens, onde ambas apresentaram *recall* equivalente a 1, ou seja, a ausência de falsos negativos; entretanto, a precisão foi ligeiramente maior para a imagem RGB.

Ao analisar os mapas (Figura 2), é possível distinguir que a área norte foi o local que obteve uma maior classificação de pastagem para as duas imagens; a isto, a explicação é dada pelo maior vigor vegetativo da pastagem e uma maior densidade, permitindo uma maior diferença espectral. Quanto à maior precisão para plantas daninhas de folha estreita na imagem RGB, a explicação pode se dar pelo fato de que, com menores informações de bandas, o confundimento por parte do algoritmo pode ter sido menor; isto porque a maioria das plantas daninhas amostradas possuíam uma arquitetura e distribuição semelhantes à da pastagem além de que, muitas delas se encontravam no centro das touceiras da própria pastagem, e seguindo pela premissa de que a imagem multiespectral consegue capturar uma maior quantidade de informações nos pixels (Medauar, *et al.*, 2020), a possibilidade de confusão é maior.

A era digital e suas implicações sociais: Desafios e contribuições

Em contrapartida, a imagem multiespectral diferenciou razoavelmente melhor o solo exposto, pois a utilização da informação NIR separa a vegetação e o solo exposto de forma mais eficiente (Lottes *et al.*, 2017). Porém, houve superestimação por parte do classificador. A superestimação na imagem multiespectral das plantas daninhas tanto para C1 quanto para C2, pode se dar ao fato de que, por possuírem forma irregular, dificultaram um bom delineamento manual, conforme relatado também por de Castro *et al.* (2018). Além disso, a superestimação pode ser explicada pelo vigor vegetativo das mesmas, que, em alguns casos, era superior ao da pastagem.

Em algumas áreas de solo exposto, o classificador confundiu com plantas daninhas de folha estreita nas duas imagens, isto porque a maioria das amostras coletadas apresentavam um baixo vigor vegetativo, tinham aspecto rasteiro e em algumas amostras, estavam em meio ao solo. A reflectância espectral interfere em plantas com estresses fisiológicos (Carter; Knapp, 2001). Portanto, a imagem RGB demonstrou melhores resultados.

Conclusão

Foi possível identificar e classificar de forma automatizada a incidência de plantas daninhas em área cultivada com pastagem, bem como separá-las de acordo com sua arquitetura e distingui-las da pastagem e solo exposto. O classificador apresentou bons resultados neste processo, com destaque significativo para a imagem RGB.

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.
- ALULEMA, J. C. M.; LOZANO, Á. G. G.; RAMOS, J. M. O.; GOMEZ, J. A. N. MACHINE-LEARNING ALGORITHMS FOR WEED MANAGEMENT USING KNN. **Ann. For. Res.**, v. 66, n. 1, p. 2330-2340, 2023.
- CARTER, G.; KNAPP, A. K. Leaf optical Properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. **American journal of botany**, v. 88, n. 4, p. 677-684, 2001.
- DE CASTRO, A. I.; TORRES-SÁNCHEZ, J.; PEÑA, J. M.; JIMÉNEZ-BRENES, F. M.; CSILLIK, O.; LÓPEZ-GRANADOS, F. An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. **Remote Sensing**, v. 10, n. 2, p. 285, 2018.
- DE CASTRO, A.; TORRES-SÁNCHEZ, J.; PEÑA, J. M.; JIMÉNEZ-BRENES, F. M.; CSILLIK, O.; LÓPEZ-GRANADOS, F. An automatic Random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. **Remote Sensing**, v. 10, n. 2, p. 285, 2018.
- EL IMANNI, H. S.; EL HARTI, A.; BACHAOUI, E. M.; MOUNCIF, H.; EDDASSOUQUI, F.; HASNAI, M. A.; ZINELABIDINE, M. I. Multispectral UAV data for detection of weeds in a citrus farm using machine learning and Google Earth Engine: Case study of Morocco. **Remote Sensing Applications: Society and Environment**, v. 30, p. 100941, 2023.
- ESPOSITO, M.; CRIMALDI, M.; CIRILLO, V.; SARGHINI, F.; MAGGIO, A. Drone and sensor technology for sustainable weed management: A review. **Chemical and Biological Technologies in Agriculture**, v. 8, n. 1, p. 1-11, 2021.
- GOU, J.; MA, H.; OU, W.; ZENG, S.; RAO, Y.; YANG, H. A generalized mean distance-based k-nearest neighbor classifier. **Expert Systems with Applications**, v. 155, p. 356-372, 2019.
- HUNTER III, J. E.; GANNON, T. W.; RICHARDSON, R. J.; YELVERTON, F. H.; LEON, R. G. Integration of remote-weed mapping and an autonomous spraying unmanned aerial vehicle for site-specific weed management. **Pest Management Science**, v. 76, n. 4, p. 1386-1392, 2020.
- ISLAM, N.; RASHID, M. M.; WIBOWO, S.; XU, C. Y.; MORSHED, A.; WASIMI, S. A.; ...; RAHMAN, S. M. Early weed detection using image processing and machine learning techniques in an Australian chilli farm. **Agriculture**, v. 11, n. 5, p. 387, 2021.

A era digital e suas implicações sociais: Desafios e contribuições

JOHN, K. N.; VALENTIN, V.; ABDULLAH, B.; BAYAT, M.; KARGAR, M. H.; ZARGAR, M. Weed mapping technologies in discerning and managing weed infestation levels of farming systems. **Research on Crops**, v. 21, n. 1, p. 93-98, 2020.

KORAV, S.; RAM, V.; RAY, L. I.; KRISHNAPPA, R.; SINGH, N. J.; PREMARADHYA, N. Weed pressure on growth and yield of groundnut (*Arachis hypogaea* L.) in Meghalaya, India. **Int. J. Curr. Microbiol. App. Sci**, v. 7, n. 3, p. 2852-2858, 2018.

LAMB, D. W.; BROWN, R. B. Remote-sensing and mapping of weed in crops. **Journal of Agricultural Engineering Research**, v. 78, n. 2, p. 117-125, 2001.

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, 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.

LÓPEZ-GRANADOS, F.; TORRES-SÁNCHEZ, J.; SERRANO-PÉREZ, A.; DE CASTRO, A. I.; MESAS-CARRASCOSA, F. J.; & PEÑA, J. M. Early season weed mapping in sunflower using UAV technology: variability of herbicide treatment maps against weed thresholds. **Precision agriculture**, v. 17, p. 183-199, 2016.

LOTTE, P.; KHANNA, R.; PFEIFER, J.; SIEGWART, R.; STACHNISS, C. UAV-based crop and weed classification for smart farming. In: 2017 **IEEE international conference on robotics and automation (ICRA)**. IEEE, 2017. p. 3024-3031.

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.

RAY, D. K.; MUELLER, N. D.; WEST, P. C.; FOLEY, J. A. Yield trends are insufficient to double global crop production by 2050. **PloS one**, v. 8, n. 6, p. e66428, 2013.

ROSLE, R.; CHE'YA, N. N.; ANG, Y.; RAHMAT, F.; WAYAYOK, A.; BERAHIM, Z.; ...; OMAR, M. H. Weed detection in rice fields using remote sensing technique: A review. **Applied sciences**, v. 11, n. 22, p. 10701, 2021.

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 an unmanned aerial system for weed management in agricultural crops: A review. **Agronomy**, v. 11, n. 9, p. 1809, 2021.

WESTWOOD, J. H.; CHARUDATTAN, R.; DUKE, S. O.; FENNIMORE, S. A.; MARRONE, P.; SLAUGHTER, D. C.; ...; ZOLLINGER, R. Weed management in 2050: Perspectives on the future of weed science. **Weed science**, v. 66, n. 3, p. 275-285, 2018.

ZHAO, X.; WANG, X.; LI, C.; FU, H.; YANG, S.; ZHAI, C. Cabbage and weed identification based on machine learning and target spraying system design. **Frontiers in Plant Science**, v. 13, p. 924973, 2022.