

Avaliação de Estratégias de Treinamento para Segmentação Semântica na Detecção de Áreas de Estresse em Lavouras de Feijão

Lívia N. Andrade¹, Marley L. Machado¹, Vanessa C. O. Souza²,
Margarete M. L. Volpato¹, Aurinelza B. T. Condé¹, Clara L. Moreno²

¹Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG)
Belo Horizonte – MG – Brasil

²Universidade Federal de Itajubá (UNIFEI) - IMC/IESTI
Caixa Postal 50 – 37500-903 – Itajuba – MG – Brasil

livia.naiara.andrade@gmail.com, {marley,margarete,aurinelza}@epamig.br,
{vanessasouza,d2022003835}@unifei.edu.br

Abstract. *This work investigates the use of Deep Learning for semantic segmentation of stress-affected areas in bean crops from RGB images acquired by an unmanned aerial vehicle (UAV). A U-Net network with a ResNet34 backbone was trained using images subdivided into tiles of different dimensions and with different loss functions. The experiments evaluated the impact of tile size and training strategies on the model's performance. The results indicate that the configuration with 128×128 -pixel tiles and the Tversky + Cross Entropy loss function achieved the best balance in identifying stress areas in the generalization stage.*

Resumo. *Este trabalho investiga o uso de Deep Learning para segmentação semântica de áreas sob estresse em lavouras de feijão a partir de imagens RGB obtidas por veículo aéreo não tripulado (VANT). Foi treinada uma rede U-Net com backbone ResNet34 utilizando imagens subdivididas em tiles de diferentes dimensões e diferentes funções de perda. Os experimentos avaliaram o impacto do tamanho dos tiles e das estratégias de treinamento no desempenho do modelo. Os resultados indicam que a configuração com tiles de 128×128 pixels e função de perda Tversky + Cross Entropy apresentou melhor equilíbrio na identificação das áreas de estresse na etapa de generalização.*

1. Introdução

A detecção precoce de estresse vegetativo é um desafio para a agricultura moderna, especialmente diante das mudanças climáticas e da necessidade de ampliar a produtividade de forma sustentável. Nesse contexto, a agricultura digital tem se consolidado como uma abordagem tecnológica relevante, integrando sensoriamento remoto e plataformas de aquisição de dados, como Veículos Aéreos Não Tripulados (VANTs), para ampliar a resolução espacial e temporal das informações sobre as lavouras e apoiar práticas de manejo localizado e uso mais eficiente de insumos.

Ao contrário dos satélites, os VANTs podem operar sob demanda, abaixo da cobertura de nuvens, e fornecer imagens com resoluções espaciais de poucos centímetros,

permitindo a detecção de estresse em nível de planta individual [Abbas et al. 2023]. Ademais, a integração de inteligência artificial a esses dados permite automatizar a identificação de estresses hídricos, nutricionais e fitossanitários [Shahi et al. 2023]. Nesse contexto, técnicas de *deep learning* têm sido empregadas na análise de imagens agrícolas para identificação de estresse.

No entanto, conforme discutido por [Gao et al. 2020], a detecção de estresse vegetal a partir de imagens apresenta desafios importantes. Condições de campo, como variações de iluminação e movimento das plantas causado pelo vento, podem afetar a qualidade das imagens. Além disso, as plantas apresentam mudanças fisiológicas ao longo do ciclo de crescimento e, em muitos casos, os sintomas iniciais de estresse não são visualmente evidentes.

Este trabalho integra um projeto voltado ao manejo sustentável de lavouras em Minas Gerais, com foco nas culturas de arroz, feijão, milho e soja, utilizando imagens obtidas por VANTs e técnicas de inteligência artificial para mapear áreas sob estresse vegetativo. Em etapa anterior [Moreno et al. 2025], foram aplicados métodos de *Machine Learning* para a classificação desses padrões na cultura do feijão. Dando continuidade a essa linha de pesquisa, o presente estudo avança para o uso de *Deep Learning*, propondo a aplicação da arquitetura *U-Net* para a segmentação semântica de estresse vegetativo em imagens de feijão obtidas por VANT, visando explorar a capacidade das redes convolucionais em extrair características espaciais complexas.

Diante do exposto, este trabalho investiga o impacto de diferentes estratégias de treinamento no desempenho de modelos de segmentação aplicados a imagens RGB obtidas por VANT. A contribuição deste estudo é de natureza experimental, consistindo em uma análise sistemática do efeito das dimensões das subimagens (*tiles*), obtidas a partir da divisão do mosaico original em recortes menores para treinamento, das funções de perda (*loss functions*) e do número de épocas de treinamento sobre as métricas de segmentação e a capacidade de generalização do modelo.

O restante deste trabalho está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico; a Seção 3 descreve a metodologia; a Seção 4 discute os resultados; e a Seção 5 apresenta as considerações finais.

2. Contexto Tecnológico e Trabalhos relacionados

O sensoriamento remoto por VANTs tem ampliado as possibilidades de monitoramento agrícola, permitindo observar com alto nível de detalhe o desenvolvimento das plantas. A análise dessas imagens possibilita identificar teor de clorofila, área foliar e vigor vegetativo. Já [Aires 2025] destaca que dados obtidos a partir de imagens de drones permitem avaliar características relacionadas ao crescimento, à atividade fotossintética e ao vigor da vegetação, evidenciando o potencial dessas técnicas para o monitoramento agrícola. No entanto, na agricultura digital, é comum encontrar trabalhos que identificam estresse ou doenças nas plantas a partir de fotografias de suas folhas [Silva Neto 2023], enquanto a identificação de estresse vegetativo a partir do dossel da cultura ainda é relativamente pouco explorada na literatura.

A aplicação de técnicas de *Machine Learning* tem se consolidado como uma abordagem promissora na agricultura de precisão. Estudos de revisão indicam que a

combinação de imagens RGB ou multiespectrais com modelos baseados em aprendizado profundo permite extrair representações detalhadas da cobertura vegetal, mesmo em cenários caracterizados por elevada heterogeneidade espacial e variações fenológicas [Bouguettaya et al. 2022].

No trabalho de [Moreno et al. 2025], técnicas de *Machine Learning* foram aplicadas à classificação de padrões de estresse em imagens multiespectrais da cultura do feijão, demonstrando o potencial dessas abordagens para a identificação de variações fisiológicas nas lavouras. Os autores reportaram acurácia de 83% e média geométrica de 77,63% para o melhor modelo obtido na fase reprodutiva da cultura. Contudo, por se tratar de uma abordagem baseada na classificação espectral individual de pixels, a representação resultante carece de contexto espacial, o que pode gerar fragmentação na identificação das áreas afetadas e dificultar a delimitação precisa das manchas de estresse.

A segmentação semântica tem recebido destaque no processamento de imagens agrícolas por permitir a classificação pixel a pixel, possibilitando a delimitação de regiões correspondentes a diferentes classes de interesse, como solo, vegetação saudável e áreas de estresse. Nesse contexto, arquiteturas do tipo *encoder-decoder*, em especial a *U-Net*, tornaram-se referência devido à sua capacidade de preservar informações espaciais, aspecto essencial em cenários agrícolas [Ronneberger et al. 2015]. Em aplicações práticas, a fragmentação das imagens em *tiles* é frequentemente adotada por limitações computacionais. Contudo, essa estratégia pode impactar a modelagem do contexto espacial, pois a divisão arbitrária da imagem pode fragmentar objetos de interesse ou isolar características contextuais presentes apenas em vizinhanças maiores, dificultando a identificação de classes raras, como áreas sob estresse [Chiu et al. 2020].

Frente a essas limitações, o equilíbrio entre granularidade local e contexto global permanece como um dos principais desafios na segmentação semântica de imagens capturadas por VANTs, exigindo cuidados tanto no *pipeline* de processamento quanto nas estratégias de avaliação dos modelos [Ma et al. 2019].

3. Desenvolvimento metodológico

A metodologia utilizada foi adaptada de [Müller and Kramer 2021], originalmente proposto para segmentação de imagens médicas. No presente estudo, o *pipeline* foi ajustado para o processamento de imagens RGB obtidas por VANT, incluindo etapas de preparação dos dados, treinamento e avaliação do modelo de segmentação.

3.1. Pré-processamento e Preparação de Dados

A aquisição dos dados ocorreu em um talhão cultivado com feijão, localizado no município de Arcos/MG, com 17.549 m². Nesse talhão foram delimitadas 32 áreas submetidas a estresse induzido, por meio da aplicação controlada de defensivos agrícolas. Cada área possui aproximadamente 3 m², totalizando 96 m² de estresse na área experimental.

As imagens foram capturadas com VANT DJI Mavic 3M com RTK, em 30/04/2024, 64 dias após o plantio, quando a cultura estava no estágio R5 (início da fase reprodutiva), fase em que os sintomas de estresse tendem a estar mais evidentes. As imagens foram processadas em um mosaico RGB ortorretificado, com resolução espacial aproximada de 0,018 m (cerca de 1,8 cm por pixel).

Além do mosaico RGB, foi construída manualmente uma máscara de referência com as mesmas dimensões e características espaciais das bandas RGB. Nessa máscara, o valor 0 corresponde ao *background*, o 1 cultura saudável, o 2 estresse natural e o 3 estresse induzido durante o experimento.

Para avaliar a capacidade de generalização dos modelos estudados, definiu-se uma Área de Avaliação Externa (*hold-out* externo), correspondente a uma região do mosaico original não utilizada nas etapas de treinamento e teste. Essa área foi delimitada por um recorte retangular destacado em vermelho na Figura 1, com aproximadamente 7.800 m², abrangendo regiões do talhão e áreas adjacentes sem informação. Considerando apenas pixels com valores digitais válidos, a área efetivamente utilizada foi de cerca de 3.826 m².

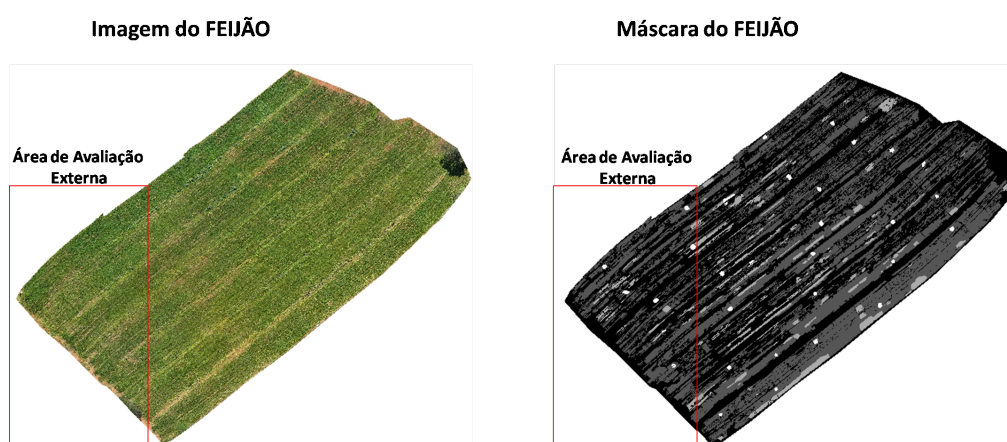


Figura 1. Mosaico RGB da área experimental (à esquerda) e a respectiva máscara de referência (à direita). A área delimitada em vermelho indica a Área de Avaliação Externa utilizada para validação do modelo.

Para o treinamento, o mosaico e a máscara foram divididos em subimagens (*tiles*) de diferentes dimensões (512, 256, 128 e 96 pixels) para avaliar o compromisso entre o número de amostras e a preservação do contexto espacial.

O processo de geração dos *tiles* seguiu o fluxo da Figura 2, incluindo alinhamento, filtragem de amostras sem dados e aplicação de *padding*. Os pares imagem–máscara foram organizados em conjuntos de treinamento (80%) e teste (20%). A divisão foi realizada de forma aleatória, com um script configurado com critério adicional para garantir que todas as classes, inclusive as minoritárias, estivessem representadas de forma equilibrada em ambos os conjuntos, analisando as máscaras e distribuindo as amostras das classes menos frequentes da maneira mais uniforme possível.

3.2. Arquitetura do Modelo e Treinamento

Os experimentos foram realizados no Google Colab, em ambiente Python 3, com aceleração por GPU NVIDIA L4 (23 GB de memória dedicada, CUDA 13.0). O modelo foi implementado em PyTorch utilizando a biblioteca *Segmentation Models PyTorch* (SMP), com suporte das bibliotecas *Albumentations*, *Rasterio*, *OpenCV* e *TorchMetrics*. Para a segmentação semântica, utilizou-se a arquitetura *U-Net* com *backbone* ResNet-34 e pesos pré-treinados no ImageNet. A escolha da *U-Net* deve-se à sua ampla utilização em tarefas de segmentação semântica em imagens agrícolas, além de apresentar boa capaci-

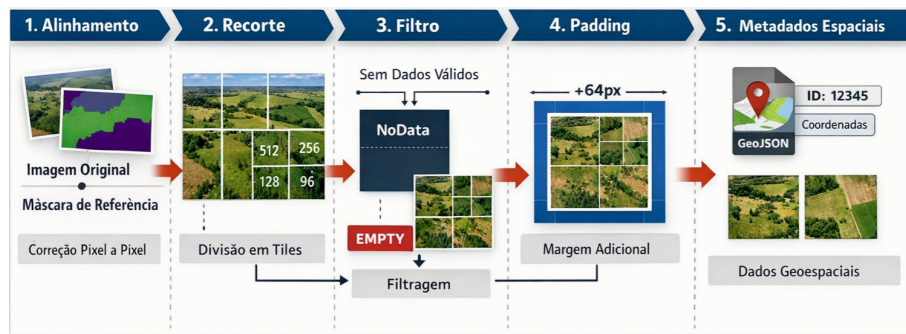


Figura 2. Diagrama do processo de geração dos tiles.

dade de preservação de informações espaciais e uma relação adequada entre desempenho e custo computacional.

O treinamento foi realizado utilizando o otimizador *Adam*, com taxa de aprendizado 1×10^{-4} e tamanho de lote (*batch*) de quatro imagens. Quanto à função de perda, foi utilizada inicialmente a *Cross-Entropy* (CE); contudo, devido ao desbalanceamento entre classes, adotou-se uma abordagem híbrida combinando CE e *Tversky Loss*, com pesos de 0,3 e 0,7, respectivamente, de modo a priorizar a contribuição da *Tversky Loss* e aumentar a sensibilidade do modelo às classes minoritárias.

Durante o carregamento dos dados para o treinamento, foi adotado um mecanismo de amostragem aleatória, no qual 80% das amostras disponíveis eram selecionadas a cada iteração, introduzindo variabilidade entre épocas e atuando como mecanismo auxiliar de regularização.

Dessa forma, foram conduzidos experimentos variando três fatores: tamanho dos *tiles*, função de perda e número de épocas de treinamento. Foram avaliados quatro tamanhos de *tiles* (512, 256, 128 e 96 pixels), duas funções de perda — *Cross Entropy* (CE) e *Tversky + CE* — e três durações de treinamento (30, 50 e 100 épocas), totalizando 24 combinações experimentais.

3.3. Avaliação de Desempenho

Durante o treinamento foram monitoradas as métricas Loss, IoU (*Intersection over Union*) e F1-Score. Na etapa de teste, o desempenho foi avaliado por métricas globais e por classe (IoU, F1-Score, Precisão e *Recall*). O treinamento foi realizado com *tiles* extraídos do mosaico original; contudo, por se tratar de uma arquitetura totalmente convolucional, o modelo pode ser aplicado a imagens de diferentes dimensões. Assim, na validação foi utilizada a área de validação, na qual o modelo foi aplicado diretamente sobre a imagem completa, sendo também analisadas a matriz de confusão e os mapas de erro. A análise dos resultados foi conduzida de forma descritiva e comparativa, sem aplicação de testes estatísticos inferenciais, devido à ausência de replicações independentes dos experimentos. Em função do elevado custo computacional e das limitações de tempo de execução, optou-se por uma única execução por configuração experimental.

4. Resultados e Discussão

A Tabela 1 apresenta os resultados dos experimentos de treinamento, realizados de forma independente com 30, 50 e 100 épocas. De modo geral, a *CrossEntropyLoss* isolada

apresentou melhor desempenho nos *tiles* de 96, 128 e 256 pixels, com menores valores de Loss e maiores índices de IoU e F1-Score. Embora a combinação com *Tversky Loss* tenha sido avaliada para aumentar a sensibilidade às classes minoritárias, seus resultados globais foram inferiores, possivelmente devido ao aumento de falsos positivos, o que impacta métricas como IoU e F1-Score.

Tabela 1. Métricas obtidas na última época do treinamento.

Tile	Nº de Épocas	CrossEntropy (CE)			Tversky + CrossEntropy		
		Loss	IoU	F1Score	Loss	IoU	F1Score
96 x 96 px	30	0.0278	0.8721	0.9136	0.1999	0.7041	0.7680
	50	0.0146	0.9227	0.9476	0.1544	0.7818	0.8336
	100	0.0061	0.9672	0.9793	0.1157	0.8507	0.8798
128 x 128 px	30	0.0385	0.8494	0.8955	0.1895	0.7179	0.7895
	50	0.0160	0.9378	0.9612	0.1492	0.8025	0.8557
	100	0.0073	0.9737	0.9845	0.0964	0.9228	0.9518
256 x 256 px	30	0.0636	0.8714	0.9195	0.1451	0.7996	0.8728
	50	0.0255	0.9528	0.9746	0.0820	0.8894	0.9330
	100	0.0173	0.9611	0.9777	0.0476	0.9374	0.9591
512 x 512 px	30	0.1076	0.8615	0.9205	0.1263	0.8374	0.9092
	50	0.0568	0.9313	0.9639	0.1468	0.8080	0.8812
	100	0.0394	0.9498	0.9741	0.0293	0.9574	0.9781

Apesar de algumas configurações apresentarem métricas elevadas no treinamento - como o modelo com *tiles* de 96 pixels e *CrossEntropy* por 100 épocas - esses resultados devem ser interpretados com cautela, pois podem refletir sobreajuste. Assim, a análise do desempenho no conjunto de teste torna-se necessária, cujos resultados são apresentados na Figura 3.

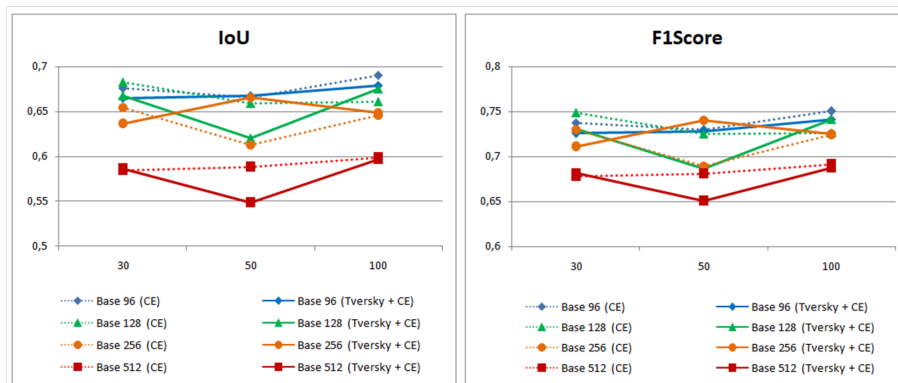


Figura 3. Gráficos com valores de IoU e F1Score calculados sobre a base de teste.

Os gráficos da Figura 3 mostram que, diferentemente do observado no treinamento, os resultados na base de teste não apresentam crescimento consistente com o aumento do número de épocas. Em algumas configurações observa-se estabilidade ou pequenas variações negativas, entre 1% e 7%. De modo geral, os *tiles* de 96 e 128 pixels apresentaram os melhores resultados médios de IoU e F1-Score. A *CrossEntropyLoss* manteve desempenho ligeiramente superior em relação à combinação *Tversky + CrossEntropy*, com diferenças médias inferiores a 4%, comportamento consistente com o observado na etapa de treinamento. O *tile* de 256 pixels apresentou maior oscilação, enquanto o *tile* de 512 pixels apresentou os menores resultados, possivelmente devido à perda de detalhamento espacial em recortes maiores.

A Figura 4 apresenta as métricas *IoU*, F1-Score, Precisão e *Recall* para a classe 2 (feijão com estresse induzido) na base de teste, classe de maior interesse agrônomico e com menor representatividade no conjunto de dados. De modo geral, os valores são inferiores às métricas globais, refletindo o desbalanceamento entre classes e a maior complexidade da segmentação das áreas de estresse. O *tile* de 256 pixels apresentou evolução progressiva com o aumento das épocas, especialmente com *CrossEntropyLoss*, alcançando *IoU* de 0,3978 e F1-Score de 0,5614 em 100 épocas, sugerindo que um contexto espacial intermediário favorece sua identificação. Já os *tiles* de 96 e 128 pixels apresentaram oscilações entre as funções de perda, sem tendência consistente de melhoria, enquanto o *tile* de 512 pixels mostrou desempenho mais moderado, com incremento principalmente no *Recall* na combinação *Tversky Loss* + *CrossEntropyLoss*.

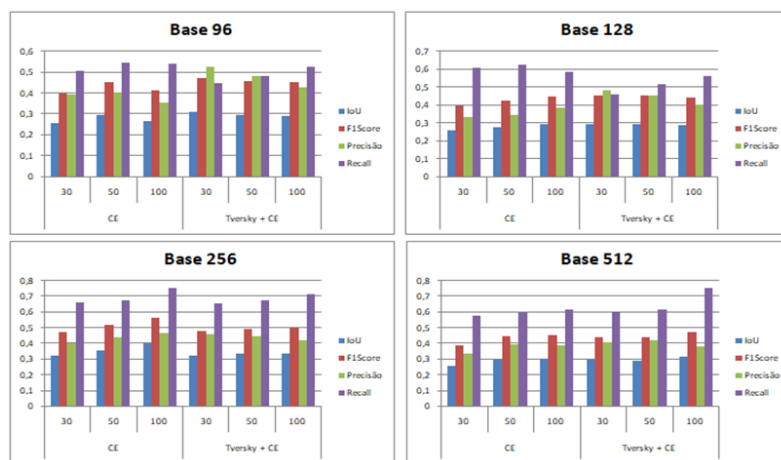


Figura 4. Gráficos com valores de IoU, F1 Score, Precisão e *Recall* da classe 2 (áreas de feijão com estresse), calculados sobre a base de teste.

A análise conjunta de Precisão e *Recall* ajuda a interpretar de maneira mais profunda os resultados da Figura 4. Em diversas configurações, o *Recall* supera a Precisão, indicando maior capacidade de detecção das áreas de estresse, com maior ocorrência de falsos positivos. Por outro lado, configurações com maior Precisão e menor *Recall* indicam segmentações mais conservadoras.

Após a avaliação interna, procedeu-se à análise da capacidade de generalização do modelo. Foram analisadas métricas por classe, com ênfase na classe 2, seguida de análise visual comparativa entre a imagem RGB, a máscara de referência, a segmentação predita e o mapa de diferenças.

Embora a Base 256 tenha apresentado melhor desempenho nas métricas globais, a análise específica da classe 2 (Figura 5) revela um cenário distinto. As maiores taxas de *IoU* para essa classe foram observadas nas configurações Base 96 (*Tversky* + CE, 30 épocas) e Base 128 (*Tversky* + CE, 50 épocas), ambas próximas de 0,30. Considerando *IoU*, F1-score, Precisão e *Recall*, a configuração Base 128 com *Tversky* + CE e 50 épocas apresentou comportamento mais equilibrado, com maior *Recall* (0,5902), embora com redução na Precisão (0,3838). A matriz de confusão (Tabela 2) confirma esse comportamento. Para a classe 2, aproximadamente 59% dos pixels foram corretamente identificados, enquanto parte foi confundida com as classes 0 e 1, indicando similaridade visual entre essas categorias. Também foram observados falsos positivos provenientes das

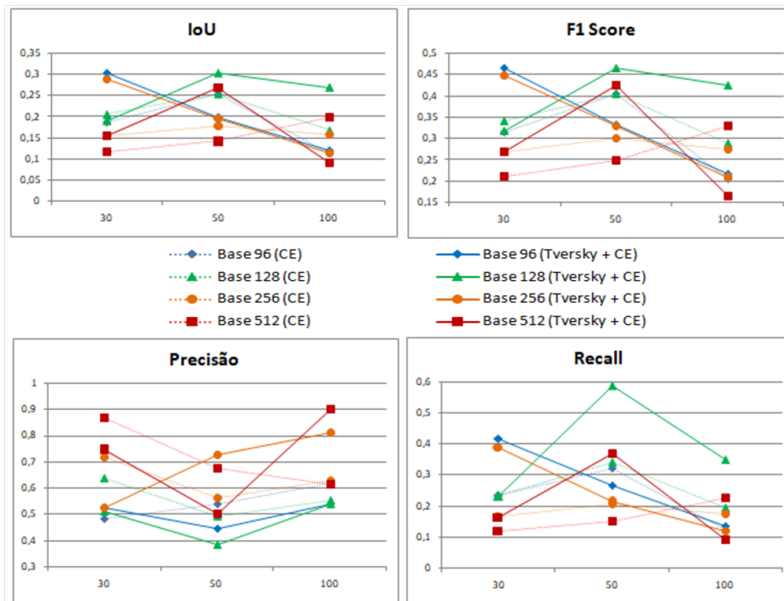


Figura 5. Métricas da classe de estresse na imagem de validação em função do número de épocas e das configurações de base/*tile*.

classes majoritárias, o que explica a redução na Precisão. Ainda assim, essa configuração apresentou maior proporção de acertos para a classe de interesse.

A Figura 6 apresenta essa configuração, na qual o modelo demonstra maior sensibilidade na identificação das áreas de estresse, com delimitação espacial coerente em relação à máscara de referência, embora com pequenas superestimações em regiões adjacentes.

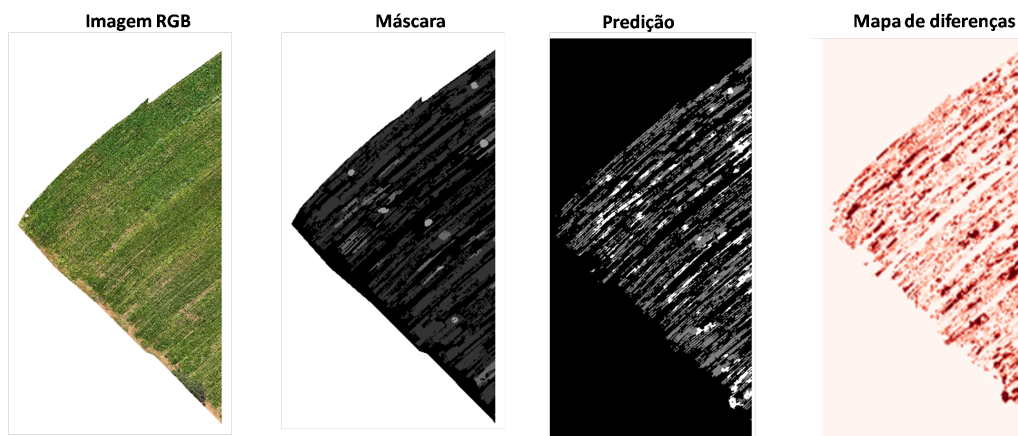


Figura 6. Segmentação obtida com a configuração Base 128×128 px (*Tversky + CE*, 50 épocas): imagem RGB, máscara de referência, predição do modelo e mapa de diferenças (erros em vermelho).

Considerando que, em aplicações de monitoramento agrícola, a não detecção de áreas sob estresse pode implicar perdas produtivas mais significativas do que a sinalização excessiva de regiões saudáveis, a priorização do *Recall* torna-se estrategicamente mais adequada. Nesse contexto, a configuração Base 128 × 128 pixels, combinada à função de

Tabela 2. Matriz de confusão (%) para a área de validação externa, utilizando *tiles* 128×128 px (*Tversky* + CE, 50 épocas).

	Classe predita 0	Classe predita 1	Classe predita 2
Classe real 0	97	2.6	0.4
Classe real 1	34.4	56.6	8.9
Classe real 2	15.9	25.1	59

perda *Tversky* + *CrossEntropy* e 50 épocas de treinamento, mostrou-se a alternativa mais apropriada para a identificação das áreas de estresse na etapa de generalização.

Por dar continuidade a investigações conduzidas na mesma área experimental, os resultados deste estudo podem ser interpretados à luz das evidências apresentadas por [Moreno et al. 2025], que também relatam dificuldades na identificação de estresse em lavouras de feijão devido à heterogeneidade espacial da área e à presença de solo exposto e falhas na cultura, que podem gerar confusão para os modelos ao distinguir vegetação saudável de regiões com menor densidade ou áreas estressadas. Esses fatores indicam que parte dos desafios observados não se restringe à modelagem, mas também às características intrínsecas das imagens agrícolas. Na validação externa, a configuração com *tiles* de 128 × 128 pixels, função de perda *Tversky* + *CrossEntropy* e 50 épocas apresentou melhor desempenho para a classe de estresse, com IoU próxima de 0,30 e Recall de 0,59, indicando maior sensibilidade na detecção dessas áreas quando comparada à abordagem de classificação espectral pixel a pixel.

5. Considerações Finais

Os resultados mostraram que o tamanho dos *tiles* influencia diretamente o desempenho do modelo. De modo geral, *tiles* de 96 e 128 pixels apresentaram melhores métricas globais, enquanto *tiles* maiores, como 512 pixels, apresentaram redução de desempenho, possivelmente associada à perda de detalhamento espacial. A análise das funções de perda indicou que a *CrossEntropy* apresentou desempenho global superior em parte das configurações. Contudo, na etapa de generalização, a combinação *Tversky* + *CrossEntropy* mostrou maior sensibilidade na identificação da classe de estresse, favorecendo o *Recall* dessa classe minoritária. Nesse contexto, a configuração com *tiles* de 128 × 128 pixels, função de perda *Tversky* + *CrossEntropy* e 50 épocas apresentou o melhor equilíbrio entre sensibilidade e controle de superestimacões na validação externa.

De forma geral, os resultados indicam que o desempenho de modelos de segmentação em imagens agrícolas depende não apenas da arquitetura utilizada, mas também das escolhas do *pipeline* de treinamento e da escala espacial dos dados. Assim, a principal contribuição deste estudo é experimental, ao apresentar uma análise comparativa de diferentes configurações de treinamento para a identificação de estresse vegetativo em imagens de VANT. Estratégias otimizadas de treinamento contribuem para a viabilidade da agricultura digital, permitindo que modelos de *deep learning* gerem mapas de estresse mais precisos para o manejo localizado.

Como trabalhos futuros, pretende-se investigar a generalização temporal dos modelos, bem como avaliar novas estratégias para aprimorar a detecção de classes minoritárias e ampliar a robustez da avaliação experimental.

Notas sobre a elaboração do artigo: Ferramentas de inteligência artificial ChatGPT e Gemini foram usadas para revisão textual e geração de imagens. Todas as decisões analíticas e interpretativas permaneceram sob responsabilidade dos autores.

Agradecimentos

Os autores agradecem à FAPEMIG e à EPAMIG pelo apoio aos projetos de pesquisa, bem como à UNIFEI pelo apoio institucional às pesquisas vinculadas a este trabalho.

Referências

- Abbas, I., Liu, J., Faheem, M., Noor, R. S., Shaikh, S. A., Solangi, K. A., and Raza, S. M. (2023). Drones in plant disease assessment, efficient monitoring, and detection: A way forward to smart agriculture. *Agronomy*, 13(6):1524.
- Aires, A. N. R. (2025). Uso de imagens coletadas por aeronaves remotamente pilotadas para fenotipagem digital da cultura de soja. Trabalho de conclusão de curso, Universidade Federal de Uberlândia, Uberlândia, MG, Brasil.
- Bouguettaya, A., Zarzour, H., Kechida, A., and Taberkit, A. M. (2022). Deep learning techniques to classify agricultural crops through uav imagery: a review. *Neural Computing and Applications*, 34(12):9511–9536.
- Chiu, M.-T. et al. (2020). Agriculture-vision: A large aerial image database for agricultural pattern analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2828–2838. IEEE.
- Gao, Z., Luo, Z., Zhang, W., Lv, Z., and Xu, Y. (2020). Deep learning application in plant stress imaging: A review. *AgriEngineering*, 2(3):430–446.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152:166–177.
- Moreno, C. L., Souza, V. C. O., Machado, M. L., and Freitas, B. H. G. S. (2025). Aprendizado sensível ao custo aplicado na detecção de estresse agrícola na cultura do feijão. In *Anais do XVI Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA)*, Maceió, AL, Brasil. Sociedade Brasileira de Computação.
- Müller, D. and Kramer, F. (2021). Miscnn: a framework for medical image segmentation with convolutional neural networks and deep learning. *BMC Medical Imaging*, 21(1).
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer.
- Shahi, T. B., Xu, C.-Y., Neupane, A., and Guo, W. (2023). Recent advances in crop disease detection using uav and deep learning techniques. *Remote Sensing*, 15(9):2450.
- Silva Neto, H. d. (2023). Segmentação de pragas e doenças em folhas de café utilizando redes convolucionais. Dissertação (mestrado em tecnologias sustentáveis), Instituto Federal do Espírito Santo, Campus Vitória, Vitória, ES.