Revista Engenharia na Agricultura

VII SIMEAA

Viçosa, MG, DEA/UFV - DOI: 10.13083/reveng.v34iEd.Especial.23439 v.34, Ed. Especial, p.1-7, 2026

SEMANTIC SEGMENTATION OF COFFEE CANOPIES USING UAV MULTISPECTRAL IMAGERY AND U-NET

Gabriel de Morais Campos[1]*, Flora Maria de Melo Villar[1], Domingos Sárvio Magalhães Valente[1], Daniel Marçal de Queiroz[1], Clarisse Mercês Guedes[1], Fabio Daniel Tancredi[1], Michel Castro Moreira[1] & Francisco de Assis de Carvalho Pinto[1]

1 - Federal University of Viçosa, Agricultural Engineering Department, Viçosa, Minas Gerais, Brazil

Keywords:
Machine learning
Remote sensing
Precision agriculture

ABSTRACT

Coffee production in Brazil requires efficient monitoring tools, especially in regions with heterogeneous terrain and planting patterns. Accurate delineation of coffee canopies from UAV imagery is essential for extracting reliable agronomic information and supporting precision agriculture practices. Multispectral UAV images were acquired over a commercial coffee farm in the Matas de Minas region, Brazil. An orthomosaic was generated and manually annotated to create binary ground-truth masks of coffee canopies. A U-Net convolutional neural network was trained using multispectral bands and the Normalized Difference Vegetation Index (NDVI) as input. Model performance was evaluated using Dice coefficient, Intersection over Union (IoU), precision, recall, and accuracy. The proposed model achieved a Dice coefficient of 0.89, IoU of 0.83, and overall accuracy above 0.92. The predicted coffee canopy area (1.50 ha) was consistent with the reference area obtained from manual annotation (1.45 ha), indicating reliable canopy delineation under heterogeneous field conditions. The results demonstrate that semantic segmentation using multispectral UAV imagery and a U-Net model is an effective and applicable approach for coffee canopy mapping, with potential use in precision agriculture and farm-scale decision-support systems.

Palavras-chave:
Aprendizagem de máquina
Sensoriamento remoto
Agricultura de precisão

SEGMENTAÇÃO SEMÂNTICA DE COPAS DE CAFEEIROS UTILIZANDO IMAGENS MULTIESPECTRAIS OBTIDAS POR VANT E U-NET

RESUMO

A produção de café no Brasil requer ferramentas eficientes de monitoramento, especialmente em regiões com relevo heterogêneo e diferentes padrões de plantio. A delimitação precisa das copas de cafeeiros a partir de imagens obtidas por veículos aéreos não tripulados (VANTs) é essencial para a extração de informações agronômicas confiáveis e para o suporte às práticas de agricultura de precisão. Imagens multiespectrais obtidas por VANT foram adquiridas em uma fazenda comercial de café na região das Matas de Minas, Brasil. Um ortomosaico foi gerado e anotado manualmente para a criação de máscaras binárias de referência das copas dos cafeeiros. Uma rede neural convolucional U-Net foi treinada utilizando bandas multiespectrais e o Índice de Vegetação por Diferença Normalizada (NDVI) como dados de entrada. O desempenho do modelo foi avaliado por meio do coeficiente Dice, Interseção sobre União (IoU), precisão, recall e acurácia. O modelo proposto alcançou coeficiente Dice de 0,89, IoU de 0,83 e acurácia global superior a 0,92. A área de copa de cafeeiros predita (1,50 ha) foi consistente com a área de referência obtida por anotação manual (1,45 ha), indicando delimitação confiável das copas sob condições heterogêneas de campo. Os resultados demonstram que a segmentação semântica utilizando imagens multiespectrais obtidas por VANT e o modelo U-Net constitui uma abordagem eficaz e aplicável para o mapeamento de copas de cafeeiros, com potencial de uso na agricultura de precisão e em sistemas de apoio à tomada de decisão em escala agrícola.

Artigo premiado com menção honrosa no VII SIMEAA

Email (corresponding author)*: gabriel.d.campos@ufv.br

INTRODUCTION

Coffee production is one of the most important agricultural activities in Brazil, both in economic and social term (CONAB, 2025). The country has historically occupied a leading position in the global coffee market, and maintaining high productivity and quality levels is essential to ensure competitiveness. In this scenario, the adoption of technologies capable of providing accurate, timely, and spatially detailed information has become increasingly relevant for decision-making in coffee production systems.

In regions such as the Matas de Minas, coffee cultivation is characterized by mountainous terrain, irregular plot geometry, heterogeneous planting densities, and a wide diversity of cultivars and plant ages. These characteristics impose significant limitations on traditional field-based monitoring, which is often time-consuming, labor-intensive, and costly. As a result, producers and researchers have sought alternative methods to monitor crop development and support management practices more efficiently (BARBOSA *et al.*, 2021; DE CASTRO *et al.*, 2023; MARIN *et al.*, 2021; ORLANDO *et al.*, 2024).

Remote sensing using unmanned aerial vehicles (UAVs) has emerged as a promising tool in this context. UAVs equipped with multispectral sensors enable the acquisition of high-resolution imagery with flexible temporal frequency, allowing detailed assessment of crop conditions. Previous studies have demonstrated the potential of UAV imagery for evaluating coffee plant vigor and nutritional status (ALVES *et al.*, 2022), pest and disease occurrence (DE CASTRO *et al.*, 2023; MARIN *et al.*, 2021; ORLANDO *et al.*, 2024), and yield-related variables yield (BARBOSA *et al.*, 2021). However, the extraction of reliable quantitative information from these images depends on accurate identification and delineation of coffee plant canopies.

Among the variables derived from UAV imagery, coffee canopy area stands out due to its relationship with vegetative growth and its potential association with yield, plant health, and management practices (JARAMILLO-BOTERO *et al.*, 2010). Nevertheless, automatic canopy delineation in coffee plantations remains challenging. Coffee plants often exhibit irregular canopy shapes and overlapping crowns, especially in older or densely planted fields. Additionally, the spectral similarity between coffee plants and other vegetation types, such as weeds, pasture, or adjacent forest fragments, complicates the segmentation process.

To address these challenges, machine learning techniques, particularly deep learning approaches, have been increasingly applied to agricultural image analysis. Convolutional neural networks (CNNs) have demonstrated strong performance in extracting complex spatial and spectral patterns from high-resolution imagery (BHADRA *et al.*, 2024; QI *et al.*, 2024). Within this context, semantic segmentation techniques aim to classify each pixel of an image into predefined classes, enabling precise mapping of crop-covered areas.

The U-Net architecture has gained prominence in semantic segmentation tasks due to its encoder–decoder structure with skip connections, which allows simultaneous learning of contextual information and preservation of fine spatial details. Originally developed for biomedical image segmentation (RONNEBERGER; FISCHER; BROX, 2015), U-Net has been successfully adapted for various remote sensing applications, including crop mapping and vegetation analysis (KATTENBORN; EICHEL; FASSNACHT, 2019; ZHU *et al.*, 2025). Its relatively simple architecture and high accuracy make it attractive for practical applications in precision agriculture.

Despite these advances, studies focusing on the applied use of semantic segmentation for coffee canopy mapping under heterogeneous field conditions are still limited. Many existing works rely on simplified scenarios, such as young plantations or uniform planting patterns, which do not fully represent commercial coffee fields. Therefore, there is a need for applied studies that evaluate robust segmentation approaches under realistic production conditions. In this context, the objective of this study was to present and evaluate an applied methodology for semantic segmentation and area estimation of coffee canopies using multispectral UAV imagery and a U-Net convolutional neural network.

MATERIALS AND METHODS

Study area and image acquisition

The study area was the Boa Safra Farm, located in Paula Cândido, in the Matas de Minas region of the state of Minas Gerais, Brazil (20°48'21.84"S, 42°58'57.21"W). Multispectral images were acquired using a DJI Matrice 350 RTK UAV (DJI, China) equipped with a MicaSense RedEdge-P camera (MicaSense, USA) capturing blue, green, red, red-edge, and near-infrared bands. Flights were conducted at 100 m altitude with 80% forward and side overlap, under near-noon illumination conditions to reduce shadow effects. Radiometric calibration was performed using a MicaSense CRP2 reflectance panel.

Image processing and orthomosaic generation

The multispectral orthomosaic was generated using Pix4D Mapper software (version 4.10.0; Pix4D SA, Switzerland), maintaining the original spatial resolution. A binary ground-truth mask representing coffee canopy and background was manually created in QGIS and aligned with the orthomosaic.

The image and mask were divided into 256 × 256 pixel patches. Patches containing less than 5% coffee pixels were excluded to reduce class imbalance. The dataset was split into training, validation, and test subsets. Spectral bands were standardized using RobustScaler. The NDVI map was calculated and included as an additional input channel, resulting in a six-channel input for the neural network. The NDVI was computed as the ratio between the sum and the difference of the NIR and Red bands (ROUSE *et al.*, 1974).

U-Net Model and training

A U-Net convolutional neural network was implemented for semantic segmentation. The U-Net model architecture consists in encoder–decoder levels with skip connections. The U-Net architecture was chosen because it was originally proposed for image processing and segmentation tasks (RONNEBERGER; FISCHER; BROX, 2015).

Training was performed using a composite loss function combining Dice Loss and Focal Loss to address class imbalance between canopy and background pixels. Model hyperparameters were optimized prior to final training, and early stopping was applied to prevent overfitting.

Model evaluation

Model performance was evaluated on the independent test set. The performance of the models was evaluated using metrics widely applied in semantic segmentation tasks: Dice coefficient (Equation 1), Intersection over Union (IoU) (Equation 2), precision (Equation 3), recall (Equation 4), and accuracy (Equation 5). These metrics were calculated on the test set after the final model training.

$$Dice = \frac{2\,.TP}{2\,.TP + FP + FN} \tag{1}$$

where Dice is the Dice coefficient, TP (True Positives) is the number of pixels correctly classified as coffee (Class 1); FP (False Positives) is the number of background pixels incorrectly classified as coffee; and FN (False Negatives) is the number of coffee pixels incorrectly classified as background.

$$IoU = \frac{TP}{TP + FP + FN} \tag{2}$$

where, IoU is Intersection over union.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TN (True Negatives) is the number of pixels correctly classified as non-coffee (Class 0).

Predicted and ground-truth areas were calculated in square meters (m^2) for each test patch. The ground-truth coffee area of each patch was obtained from the binary ground truth masks by multiplying the total number of class 1 pixels in the patch by the square of the spatial resolution. Similarly, predicted coffee areas were calculated.

RESULTS AND DISCUSSION

The NDVI map shown in **Figure 1** reveals a spatial variability in vegetation vigor across the study area. Higher NDVI values (up to approximately 0.97) are associated with dense and healthy vegetation, predominantly observed in well-developed crop sectors, indicating elevated photosynthetic activity. Intermediate NDVI values dominate the cultivated plots and reflect differences in crop development, planting geometry, and local management conditions. In contrast, low and negative NDVI values (down to approximately -0.42) correspond to bare soil, access roads, water or sparsely vegetated areas, as evidenced by localized low-value pixels.

Table 1 presents the hyperparameter optimization results for the model. The optimized configuration indicates that a network depth of four levels and a reduced number of initial filters (16) were sufficient to achieve the best performance. A dropout rate of 0.3 was adopted, providing adequate regularization, while a learning rate of 4.67×10^{-4} ensured stable training. The batch size was set to 8, favoring smaller batches during optimization. Additionally, the focal loss parameters ($\alpha = 0.6$ and $\gamma = 3.0$) highlight the importance of handling class imbalance and hard-to-classify samples in the segmentation task.

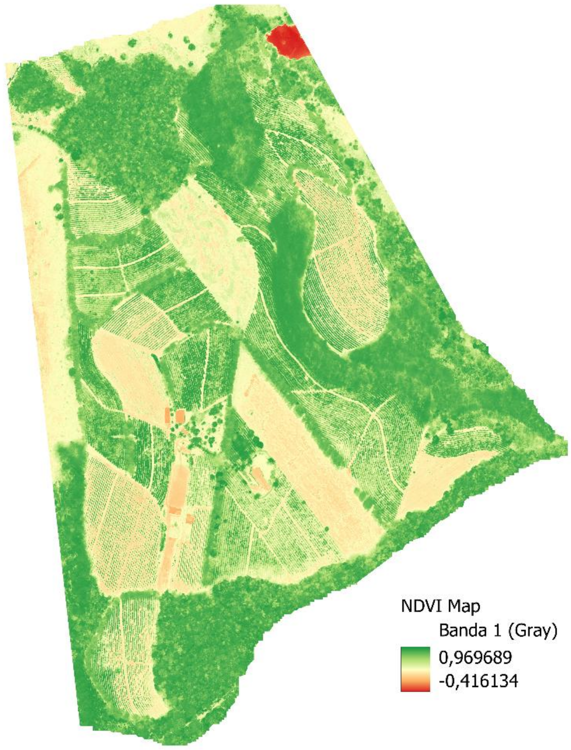

Figure 1. Normalized Difference Vegetation Index (NDVI) map of Boa Safra Farm

Table 1. Search intervals and model's optimized hyperparameters values using Optuna

Hyperparameter	Search space	Type	Optimization results
Deep	3 a 5	Discrete (int)	4
Inicial filters	16, 32, 64	Categorical	16
Dropout rate	0.1 to 0.3 (step 0.1)	Continuous (Float)	0.3
Learning rate	$1x10^{-4}$ to $1x10^{-3}$ (log scale)	Continuous (Float)	$4.67x10^{-4}$
Batch size	8, 16, 32	Categorical	8
Focal α	0.5 to 1.0 (step 0.1)	Continuous (Float)	0.6
Focal γ	1.0 to 3.0 (step 0.5)	Continuous (Float)	3.0

The U-Net model demonstrated strong performance on the test dataset, achieving a Dice coefficient of 0.89 and an Intersection over Union (IoU) of 0.83. Overall accuracy exceeded 0.92, with a precision of 0.88, a recall of 0.92, and a loss value of 0.19. These metrics indicate a high level of agreement between the predicted segmentation masks and the manually delineated ground truth, confirming the model's capability to accurately discriminate coffee canopies from background elements.

Qualitative analysis of the segmentation results revealed that the model was able to delineate coffee canopies under a wide range of field conditions, including variations in canopy density, exposed soil, and the presence of other vegetation types such as weeds and forest fragments. Even in plots with dense and overlapping canopies, the model preserved canopy boundaries with satisfactory spatial coherence, which is essential for reliable canopy area estimation. The ability of CNN-based segmentation models to handle background variability and complex canopy structures in UAV imagery has been documented in other agricultural contexts. Studies have found that deep convolutional architectures (including U-Net) outperform traditional image processing methods when confronted with heterogeneous scenes involving mixed vegetation, soil backgrounds, and occlusions (SILVA *et al.*, 2024).

The inclusion of NDVI as an additional input channel probably played a key role in improving segmentation robustness, particularly in areas with sparse canopy cover or higher soil exposure. NDVI may have enhanced the contrast between vegetated and non-vegetated pixels, facilitating the discrimination of coffee plants from bare soil and low-vigor vegetation. This result reinforces the importance of integrating spectral indices with raw multispectral bands when addressing segmentation tasks in heterogeneous agricultural environments.

The canopy area computed from the reference mask was 1.45 ha, while predicted canopy area was 1.50 ha. From a practical standpoint, the estimated total coffee canopy area derived from the predicted mask was consistent with the reference value obtained from manual annotation. This agreement suggests that the proposed methodology can be reliably applied at the farm scale to support crop mapping and monitoring activities. Accurate canopy area estimation can contribute to the assessment of vegetative development, identification of spatial variability within plots, and support for yield-related analyses.

When compared with results reported in the literature for canopy segmentation in other perennial crops, such as viticulture and citrus (BONO *et al.*, 2024; LI *et al.*, 2024), the performance achieved in this study is comparable or superior, despite the structural complexity of coffee plantations. The robustness of the model under heterogeneous conditions highlights its potential applicability in real production scenarios, where uniform planting patterns are rarely observed.

From an engineering and operational perspective, the proposed approach presents a favorable balance between accuracy and computational demand. The U-Net architecture, combined with a relatively simple training strategy, enables efficient processing while maintaining high segmentation quality. This characteristic is particularly relevant for applied contexts, where computational resources may be constrained.

Nevertheless, some limitations should be considered. Model performance may vary across different phenological stages or management practices not represented in the dataset. Additionally, the manual creation of ground-truth masks is inherently subject to operator interpretation, which may introduce uncertainty into the training process. Expanding the dataset to include multi-temporal acquisitions and refining annotation protocols are recommended to further improve model generalization.

Overall, the results demonstrate that semantic segmentation of coffee canopies using multispectral UAV imagery and a U-Net neural network is a viable and effective approach for applied agricultural monitoring. The methodology shows strong potential for integration into decision-support systems aimed at improving the efficiency and sustainability of coffee production.

CONCLUSIONS

- This study demonstrated an applied methodology for semantic segmentation and area estimation of coffee canopies using multispectral UAV imagery and a U-Net neural network. The proposed model achieved high segmentation metrics under heterogeneous field conditions, highlighting its potential for practical use in precision coffee agriculture.

The model, which incorporates NDVI as an additional input band and was configured with five encoder–decoder blocks (network depth), 16 initial filters, a dropout rate of 0.2, a learning rate of 6.23 x 10^{-4}, a batch size of 8, $\alpha = 0.9$, and $\gamma = 1.0$, achieved a Dice coefficient of 0.89. The canopy area calculated by the model was 1.50 ha, while the area in the ground truth was 1.45 ha.

AUTHORSHIP CONTRIBUTION STATEMENT

CAMPOS, G. M.: Conceptualization, Data curation, Formal Analysis, Methodology, Writing – original draft; **VILLAR, F. M. M.:** Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing; **VALENTE, D. S. M.:** Conceptualization, Methodology, Validation, Writing – review & editing; **QUEIROZ, D. M.:** Software, Supervision, Writing – review & editing; **GUEDES, C. M.:** Data curation, Visualization, Writing – review & editing; **TANCREDI, F. D.:** Funding acquisition, Resources, Software, Supervision, Writing – review & editing; **MOREIRA, M. C.:** Writing – review & editing; **PINTO, F. A. C.:** Writing – review & editing.

DECLARATION OF INTEREST

The authors declare that they have no financial or personal interests that could influence the work reported in this article.

REFERENCES

ALVES, M. de C.; SANCHES, L.; POZZA, E. A.; POZZA, A. A. A.; SILVA, F. M. da; The role of machine learning on Arabica coffee crop yield based on remote sensing and mineral nutrition monitoring. **Biosystems Engineering**, v. 221, p. 81-104, 2022.

BARBOSA, B. D. S.; FERRAZ, G. A. S.; COSTA, L.; AMPATZIDIS, Y.; VIJAYAKUMAR, V.; SANTOS, L. M. UAV-based coffee yield prediction utilizing feature selection and deep learning. **Smart Agricultural Technology**, v. 1, p. 100010, 1 dez. 2021.

BHADRA, S.; SAGAN, V.; SKOBALSKI, J.; GRIGNOLA, F.; SARKAR, S.; VILBIG, J. End-to-end 3D CNN for plot-scale soybean yield prediction using multitemporal UAV-based RGB images. **Precision Agriculture**, v. 25, n. 2, p. 834–864, 2024. Available at: https://doi.org/10.1007/s11119-023-10096-8. Access at: 30 jan. 2026.

BONO, A.; MARANI, R.; GUARAGNELLA, C.; D'ORAZIO, T. Biomass characterization with semantic segmentation models and point cloud analysis for precision viticulture. **Computers and Electronics in Agriculture**, v. 218, p. 108712, 2024. Available at: https://doi.org/10.1016/j.compag.2024.108712. Access at: 24 jun. 2025.

CONAB. **Acompanhamento da safra brasileira de café**. Brasília: CONAB, 2025. Available at: https://www.gov.br/conab/pt-br/atuacao/informacoes-agropecuarias/safras/safra-de-cafe/4o-levantamento-de-cafe-safra-2025/boletim-cafe-dezembro-2025. Access at: 19 jan. 2026.

DE CASTRO, G. D. M.; VILELA, E. F.; DE FARIA, A. L. R.; SILVA, R. A.; FERREIRA, W. P. M. New vegetation index for monitoring coffee rust using sentinel-2 multispectral imagery. **Coffee Science** - ISSN 1984-3909, v. 18, p. e182170–e182170, 2023. Available at: https://doi.org/10.25186/.V18I.2170. Access at: 18 abr. 2024.

JARAMILLO-BOTERO, C.; SANTOS, R. H. S.; MARTINEZ, H. E. P.; CECON, P. R.; FARDIN, M. P. Production and vegetative growth of coffee trees under fertilization and shade levels. **Scientia Agricola**, v. 67, n. 6, p. 639–645, 2010. Available at: https://doi.org/10.1590/s0103-90162010000600004. Access at: 30 jan. 2026.

KATTENBORN, T.; EICHEL, J.; FASSNACHT, F. E. Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. **Scientific Reports 2019 9:1**, v. 9, n. 1, p. 17656-, 2019. Available at: https://doi.org/10.1038/s41598-019-53797-9. Access at: 30 jan. 2026.

LI, Z.; DENG, X.; LAN, Y.; LIU, C.; QING, J. Fruit tree canopy segmentation from UAV orthophoto maps based on a lightweight improved U-Net. **Computers and Electronics in Agriculture**, v. 217, p. 108538, 2024. Available at: https://doi.org/10.1016/j.compag.2023.108538. Access at: 24 jun. 2025.

MARIN, D. B.; FERRAZ, G. A. e. S.; SANTANA, L. S.; BARBOSA, B. D. S.; BARATA, R. A. P.; OSCO, L. P.; RAMOS, A. P. M.; GUIMARÃES, P. H. S. Detecting coffee leaf rust with UAV-based vegetation indices and decision tree machine learning models. **Computers and Electronics in Agriculture**, v. 190, p. 106476, 2021. Available at: https://doi.org/10.1016/J.COMPAG.2021.106476

ORLANDO, V. S. W.; DE LOURDES BUENO TRINDADE GALO, M.; MARTINS, G. D.; LINGUA, A. M.; ANDALÓ, V. UAV imaging for spectral characterization of Coffee Leaf Miner (Leucoptera coffeella) infestation in the Cerrado Mineiro region. In: 2024, **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences**. Copernicus GmbH, 2024. p. 285–291.Available at: https://doi.org/10.5194/isprs-annals-X-3-2024-285-2024

QI, L.; ZUO, D.; WANG, Y.; TAO, Y.; TANG, R.; SHI, J.; GONG, J.; LI, B. Convolutional Neural Network-Based Method for Agriculture Plot Segmentation in Remote Sensing Images. **Remote Sensing** 2024, Vol. 16, Page 346, v. 16, n. 2, p. 346, 2024. Available at: https://doi.org/10.3390/RS16020346. Access at: 30 jan. 2026.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: 2015, **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. : Springer, Cham, 2015. p. 234–241.Available at: https://doi.org/10.1007/978-3-319-24574-4_28. Access at: 11 abr. 2025.

ROUSE, R. W. H.; HAAS, J. A. W.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the Great Plains with ERTS (Earth Resources Technology Satellites). **Goddard Space Flight Center 3d ERTS-1**, v. 1, n. 1, p. 309–317, 1974. Available at: https://doi.org/10.1590/0034-737X201663010008

SILVA, J. A. O. S.; SIQUEIRA, V. S. de; MESQUITA, M.; VALE, L. S. R.; MARQUES, T. do N. B.; SILVA, J. L. B. da; SILVA, M. V. da; LACERDA, L. N.; OLIVEIRA-JÚNIOR, J. F. de; LIMA, J. L. M. P. de; OLIVEIRA, H. F. E. de. Deep Learning for Weed Detection and Segmentation in Agricultural Crops Using Images Captured by an Unmanned Aerial Vehicle. **Remote Sensing**, v. 16, n. 23, p. 4394, 2024. Available at: https://doi.org/10.3390/rs16234394. Access at: 30 jan. 2026.

ZHU, Z.; CHEN, Y.; LU, C.; YANG, M.; XIA, Y.; HUANG, D.; LV, J. Research on Crop Classification Using U-Net Integrated with Multimodal Remote Sensing Temporal Features. **Sensors**, v. 25, n. 16, p. 5005, 2025. Available at: https://doi.org/10.3390/s25165005. Access at: 30 jan. 2026.