ICPA
17th INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE
ConBAP 2026
CONGRESSO BRASILEIRO DE AGRICULTURA DE PRECISÃO E DIGITAL
13-16 July 2026 | Porto Alegre, Brazil
ISPA
AsBraAP

Detection of Weed-Related Anomalies in Sugarcane Fields Using Sentinel-2 Imagery

Rafaella Pironato Amaro[1], Fernando Amstalden[1], Daniel Garbellini Duft[1], Carlos Daniel Berro Filho[1]

[1]CTC – Centro de Tecnologia Canavieira. Piracicaba, São Paulo- Brasil

A paper from the Proceedings of the

17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture

13-15 July 2026

Porto Alegre, Brazil

Abstract.

Weed infestation is one of the main causes of yield losses in agricultural systems, particularly in large-scale crops such as sugarcane. Conventional weed management, based on uniform herbicide application, often ignores the spatial variability of infestations, resulting in higher production costs and environmental impacts. In this context, remote sensing and machine learning techniques are recently being used as a solution for automation and precision in crop monitoring.

In this study, we identified anomalies in sugarcane fields associated with weed presence by using Sentinel-2 satellite imagery and Support Vector Regression (SVR). The model calibration was performed using high-resolution Unmanned Aerial Vehicle (UAV) imagery acquired over two representative areas. Then, we identified spatially the weed infestation on a grid with the Sentinel-2 spatial resolution. The sensor multispectral bands from Sentinel-2 were used as predictor variables. The model was calibrated using cross-validation and evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Bias Error (MBE) metrics, and subsequently applied to the selected fields at the mill scale.

The model presents satisfactory performance, with MAE of approximately 6% and RMSE of 8%. In addition, a MBE of 0.22 indicates a tendency toward overestimation of weed-infested areas, particularly in regions with mixed sugarcane and weed cover. This behavior might be attributed to the limited representation of pixels with full canopy or full weed cover in the training dataset. The model application enabled the spatial identification of high-risk infestation areas of approximately 4,870 ha, which corresponds to 16% of the total mill area and 36% of the analyzed area. Higher weed infestation levels were predominantly associated with advanced ratoon stages. However, a relevant incidence was also detected in second-ratoon fields, indicating potential long-term impacts on crop productivity.

The proposed approach proved effective for the detection of weed-related anomalies in sugarcane fields, supporting site-specific management and strategic decision-making. Limitations related to species discrimination and canopy conditions indicate the need to expand the calibration dataset and integrate complementary tools, such as biomass monitoring and soil maps, to achieve a more comprehensive assessment of crop status and weed management planning.

Keywords.

Remote sensing, Machine learning, Support vector regression, Decision Support

---

Introduction

In recent decades, weeds have been a major driver of yield losses in agricultural systems worldwide, with reported reductions ranging from 10% to 80% depending on crop, environment, and management practices (Oerke and Dehne, 2004; Gerhards et al., 2017).

These losses arise from both direct and indirect effects. Direct impacts include competition for essential resources such as nutrients, light, water, and space, as well as parasitism. Indirect effects are also relevant, particularly in mechanized systems such as sugarcane production, where weed presence can interfere with crop operations, reduce harvesting efficiency, shorten field longevity, and negatively affect the industrial quality of the raw material (Kuva et al., 2003).

To reduce these effects, herbicides are commonly applied uniformly across fields, often disregarding the spatial variability of weed infestation (Wang, Zhang, and Wei, 2019). Although operationally simple, this approach increases production costs and raises environmental and human health concerns. In this context, site-specific weed management has emerged as a more sustainable alternative, offering economic benefits while reducing environmental load through targeted herbicide application (Walter et al., 2002; Pätzold et al., 2008).

The implementation of site-specific management strategies relies on accurate spatial information to support decision-making. Weed mapping enables the visualization of spatial patterns and supports the identification of infestation hotspots. Traditionally, the generation of such maps require field observations combined with ancillary data, including soil properties, climate conditions, crop management, and yield data. These datasets can be used to identify risk factors, quantify weed occurrence, and evaluate the effectiveness of control measures.

Remote sensing has become a valuable tool for weed detection and mapping, using imagery acquired from unmanned aerial vehicles (UAVs) or high-resolution satellite platforms (Pérez-Ortiz et al., 2016; López-Granados, 2016; Castillejo-González et al., 2014). These technologies enable the monitoring of large agricultural areas and provide information on spatiotemporal variability, supporting continuous crop assessment and the implementation of preventive and targeted management practices (de Jeu et al., 2008; Marques et al., 2009; Randall and James, 2012).

However, accurate weed detection using remote sensing depends on the ability to discriminate weeds from crops and soil based on their spectral signatures (López-Granados, 2011). This remains a challenge in many cases, as spectral similarity between weeds and crops can limit classification performance, particularly when species exhibit comparable reflectance characteristics (Carleer and Wolff, 2004).

In this context, machine learning techniques have significantly advanced the automated identification of weeds. Approaches combining remote sensing data with algorithms such as Support Vector Machines (SVM) have shown promising results across different cropping systems (Alexandridis et al., 2017; de Castro et al., 2018; Wang et al., 2019; Yu et al., 2019). For instance, successful discrimination between weeds and crops has been reported in sugarcane (Souza et al., 2020), rice (Barrero and Perdomo, 2018), and winter wheat systems (Lambert et al., 2018).

Among machine learning methods, SVM has been widely applied in agricultural studies due to its robustness in handling high-dimensional data and its effectiveness in classification tasks. Applications include yield prediction based on environmental variables (Brdar et al., 2011) and plant disease detection (Rumpf et al., 2010). Specifically, for weed detection, Louargant et al. (2018) achieved detection accuracies of up to 89% by integrating spatial and spectral features, while Castillejo-González et al. (2014) reported accuracies above 91% for wild oat patches in wheat fields.

Overall, the integration of remote sensing and machine learning represents a promising approach for agricultural monitoring, particularly in sugarcane systems, enabling automated and continuous assessment of crop conditions.

Therefore, the objective of this study was to identify weed-related anomalies in sugarcane fields using Sentinel-2 satellite imagery and a Support Vector Machine (SVM) classification approach.

Material and Method

Study Area

The study area is in the southeastern region of the state of São Paulo, Brazil (Figure 1), encompassing the municipalities of Americana, Artur Nogueira, Campinas, Conchal, Cosmópolis, Engenheiro Coelho, Holambra, Itapira, Jaguariúna, Limeira, Mogi Guaçu, Mogi Mirim, Paulínia, and Santo Antônio de Posse. The total area covers approximately 29,918.26 ha.

The terrain exhibits variability ranging from flat to strongly undulating relief (Rossi, 2017). According to slope classes defined by Santos et al. (2013), the area can be categorized as follows: flat (0–3%), gently undulating (3–8%), undulating (8–20%), strongly undulating (20–45%), mountainous (45–75%), and steep (>75%). The predominant class in the study area is gently undulating terrain, representing approximately 50% of the fields.

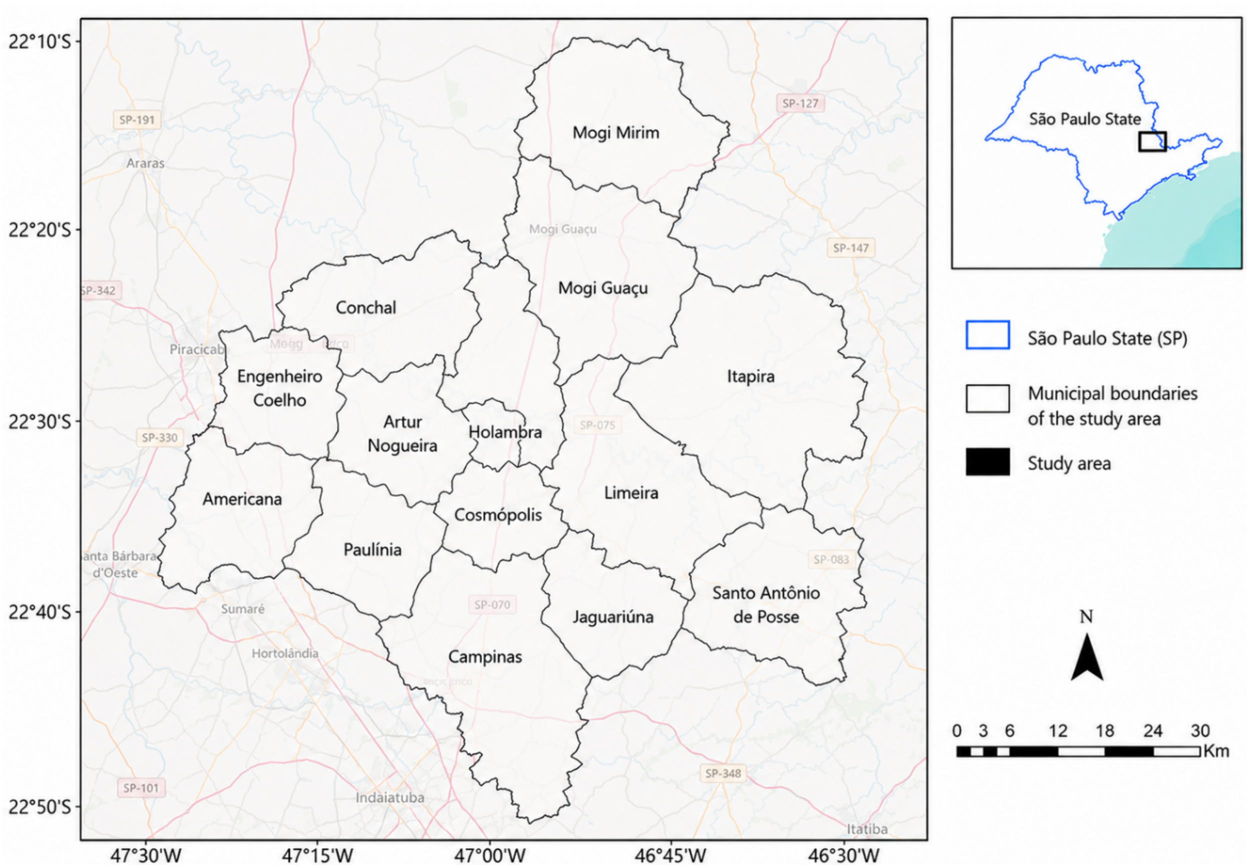

Figure 1. Study area.

Soil and climate

According to the Köppen climate classification, the study area is characterized by Cwa and Cfa climate types (Figure 2). The Cfa climate corresponds to a humid subtropical climate with hot summers, where mean temperatures exceed 22°C during the warmest months and precipitation in the driest month remains above 30 mm. In contrast, the Cwa climate is defined as a humid subtropical climate with dry winters, characterized by mean temperatures below 18°C in the coldest month and hot summers with temperatures above 22°C. This climate type predominates across sizable portions of the state of São Paulo, particularly in the central, eastern, and western regions (Ventura, 1964).

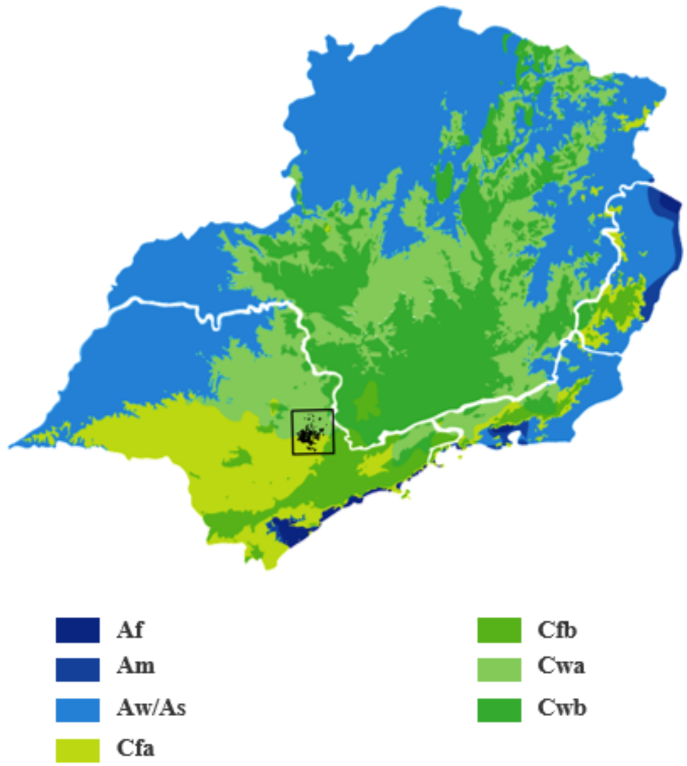

Figure 2. Climate classification. São Paulo and Minas Gerais. Source: Adapted from Alvares et al. (2014).

The dominant soil classes within the study area are Rhodic Ferralsols (9,948.81 ha), Xanthic Ferralsols (9,865.06 ha), and Acrisols (7,648.43 ha), according to Rossi (2017). Ferralsols are characterized by a well-developed latosolic B horizon occurring immediately below the A horizon, typically within 200 cm of the soil surface (or within 300 cm when the A horizon exceeds 150 cm in thickness). In contrast, Acrisols exhibit a textural B horizon with low-activity clay (Santos et al., 1999).

Soil depth in the study area ranges from shallow to very deep, with deep soils predominating and representing approximately 64% of the total area (Rossi, 2017). The prevailing soil texture is predominantly clayey to very clayey.

Dataset: acquisition and processing

The calibration dataset was collected using the senseFly eBee X Duet T UAV (Figure 3), which carried a senseFly S.O.D.A camera sensor. This setup captured images with a spatial resolution of 5 cm. The flight was conducted at an altitude of 120 m, and the images were acquired on July 21, 2020.

Figure 3. Unmanned aerial vehicle (UAV), SenseFly.

The UAV overflight covered two farms, Farm A (Figure 4a) and Farm B (Figure 4b), totaling 240 hectares.

(a) (b)

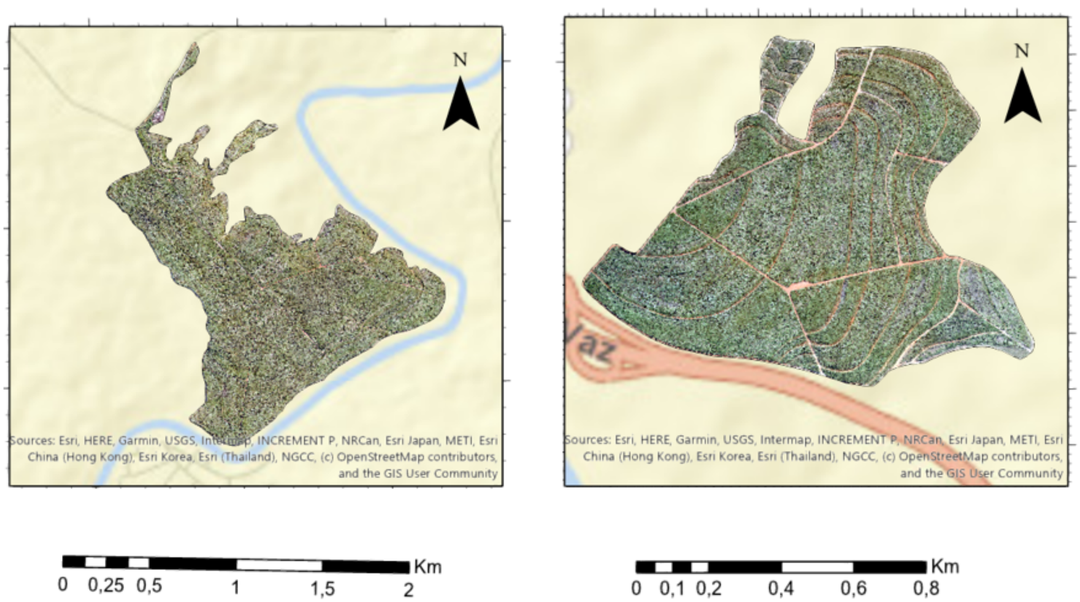

Figure 4. Drone aerial photography of farms. Farm A (a) and Farm B (b)

Farm A has soils characterized by Association of Acrisols (typical Red-Yellow Argisol), with moderate to prominent A horizon and sandy to medium texture, and Cambisols (Haplic Cambisol), with moderate A horizon and medium texture; both soils are dystrophic, occurring under undulating relief (Rossi, 2017). At the time of the analysis, the farm was a ratoon area in its second harvest.

Farm B is a farm with Association of Acrisols (abruptic or typical Red-Yellow Argisol), with moderate to prominent A horizon and sandy to medium texture, and Ferralsols (typical Red-Yellow Latosol), with moderate A horizon and medium texture, alic and dystrophic; occurring under strongly undulating to undulating relief. Soils (Rossi, 2017), and at the time of the analysis, the farm was a sugarcane plant.

Following UAV image acquisition, the area affected by weeds occurrence was identified across the study farms using a supervised machine learning approach based on the Support Vector Machine (SVM), implemented in ArcGIS Pro 2.4.3. This procedure enabled the generation of a wall-to-wall weed distribution map for the entire study area.

To ensure compatibility between UAV-derived data and satellite imagery, a regular grid of 10 m × 10 m was generated over the study area. This spatial aggregation allowed the estimation of affected areas within each grid cell, corresponding to the spatial resolution of Sentinel-2 pixels. The resulting dataset was used as the reference (target) layer for model development.

Multispectral imagery from the Sentinel-2 MSI sensor, acquired on July 17, 2020, corresponding to tiles 23KKR and 22KHV, was used as input data. These images have a revisit time of 5 days and spatial resolution depending on the spectral bands used (Table 1).

A mosaicking procedure was applied to merge the satellite scenes prior to their use as model input. All image processing steps were performed using ArcGIS Pro 2.4.3.

Table 1. Spatial resolution by spectral bands for the Sentinel-2 satellite.

Spatial resolution (m)	Band number	Band name	Spectral range (nm)
10	B02	Blue	458 – 523
10	B03	Green	543 – 578
10	B04	Red	650 – 680
10	B08	Near Infrared (NIR)	785 – 900
20	B05	Red Edge 1	698 – 713
20	B06	Red Edge 2	733 – 748
20	B07	Red Edge 3	773 – 793
20	B8A	Red Edge 4	855 – 875
20	B11	SWIR 1	1565 – 1655
20	B12	SWIR 2	2100 – 2280

Model Development and Evaluation

An anomaly detection model was developed using the Support Vector Regression (SVR), a regression-based extension of the Support Vector Machine (SVM) algorithm. SVR is widely applied to continuous prediction problems and was implemented using the e1071 library in R (Meyer et al., 2015).

A regression framework was adopted to enable continuous estimation of weed-infested areas, allowing a more detailed characterization of the relationship between the response and predictor variables (Breiman et al., 1984).

For model development, the dataset was randomly split into training and testing subsets. The training dataset, comprising two-thirds of the observations from farms, was used for model calibration, i.e., parameter estimation to best represent the underlying process.

Hyperparameter tuning was performed using a random search strategy, as proposed by Bergstra and Bengio (2012), sampling from uniform distributions within predefined ranges (Table 2). The optimal parameter combination was selected based on the lowest mean absolute error (MAE) obtained through 5-fold cross-validation, in which the training data were partitioned into five subsets and iteratively used for model fitting and validation.

Table 2. Parameters and lower and upper bounds used in the random search for the technique employed.

Parameters	Lower bound	Upper bound
Cost (C)	0,2	0,9
Gama (g)	0,5	1,0
Epsilon (eps)	0,05	0,2

The testing dataset, consisting of the remaining one-third of the observations, was used for model evaluation. Model performance was assessed using the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), and the Mean Bias Error (MBE). MBE indicates whether the model tends to overestimate (MBE > 0) or underestimate (MBE < 0) the observed values (Willmott and Matsuura, 2005).

Following calibration and validation, the model was applied to additional fields within the study area. The input data consisted of spectral bands from Sentinel-2 imagery and a reference weed infestation map derived from farm records, which was used for model training.

Model generalization was evaluated by applying the trained model to independent fields. Only fields with crop development stages comparable to the training dataset were selected, specifically those with harvest dates between May and October 2019. Out of 271 farms in the study area, 81 met this criterion, representing approximately 45% of the total cultivated area.

The overall workflow adopted for large-scale model application is illustrated in Figure 5.

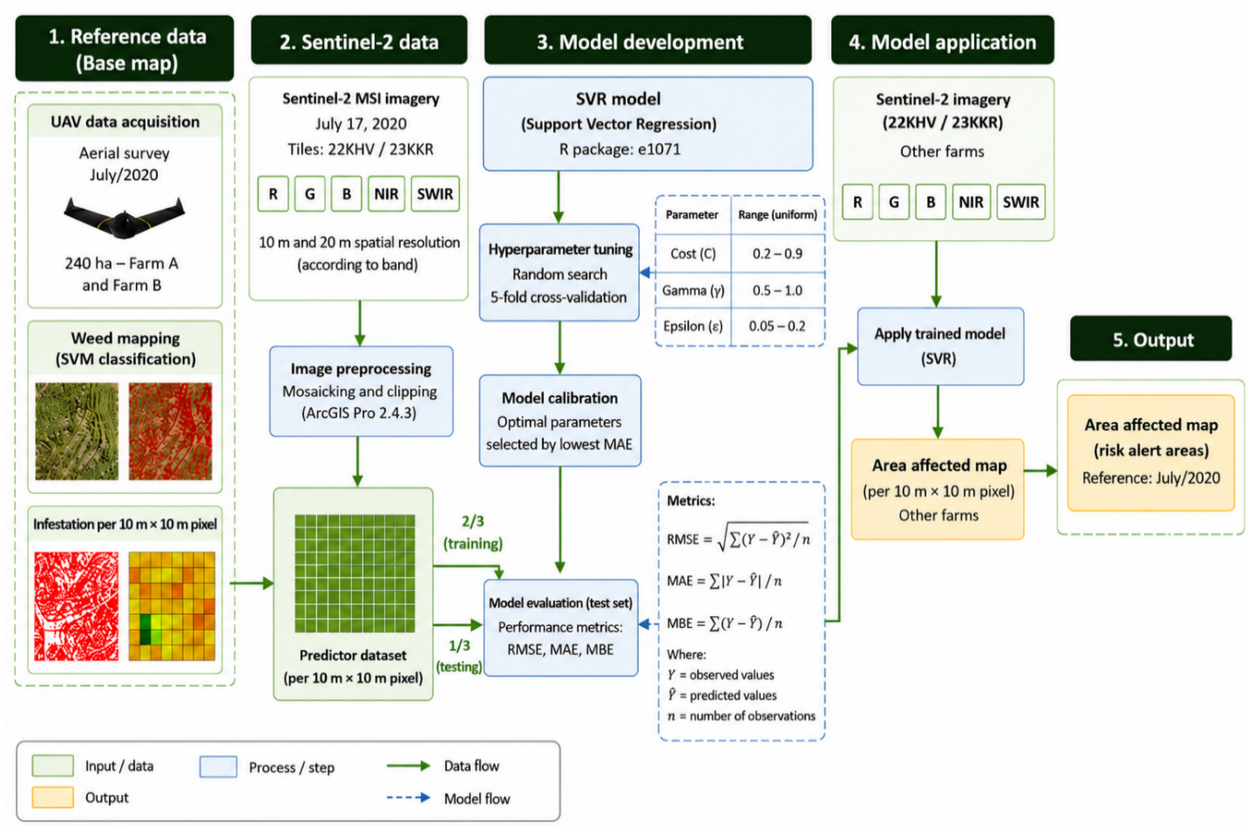

Figure 5. General methodology for applying the model throughout the mill.

Results and discussion

The supervised classification based on the Support Vector Machine (SVM) enabled the spatial identification of weed occurrence across the study area (Figure 6).

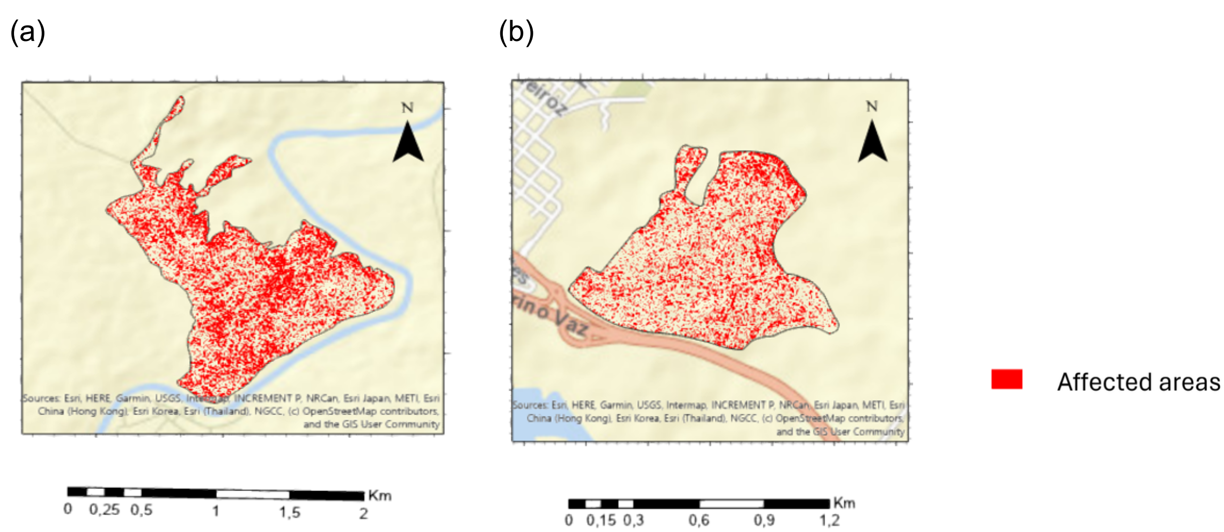

Figure 6. Affected areas by weeds. Farm A (a) and Farm B (b).

Although Farm B exhibited lower levels of weed infestation, this does not necessarily indicate better crop conditions. Field observations documented dead sugarcane plants in multiple locations across Farm B. This issue has the potential to adversely impact overall crop performance.

Following weed detection, a 10 m × 10 m grid was applied to estimate affected areas by weed at the pixel level for both Farm A (Figure 7a) and Farm B (Figure 7b), generating the base map used for model development.

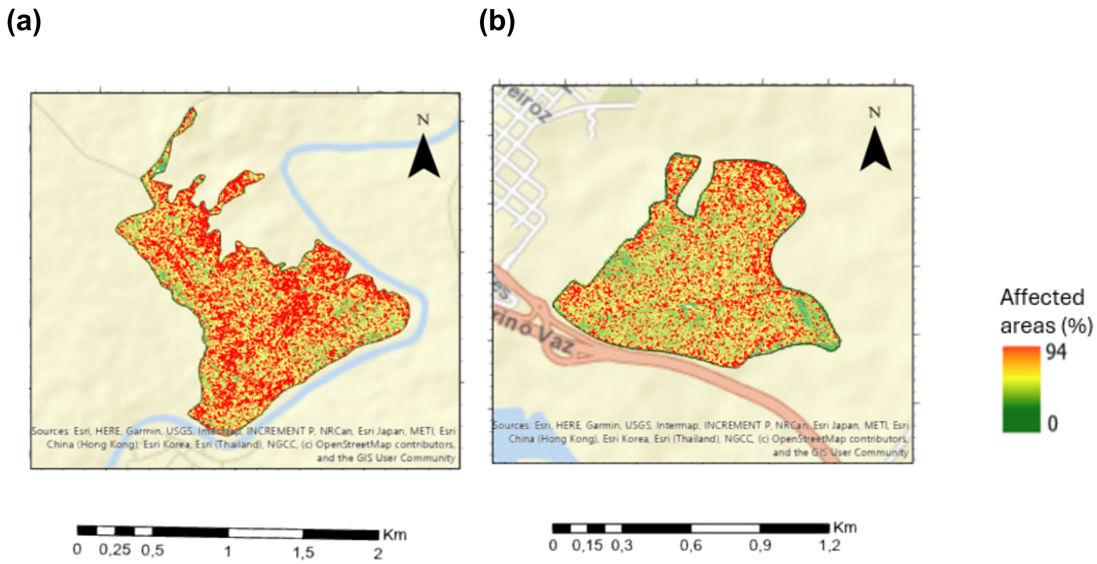

Figure 7. Base map—Affected areas per pixel. Farm A (a) and Farm B (b).

Using the base map derived from Farm A and B, together with spectral information from Sentinel-2 imagery, the model was calibrated and evaluated, resulting in a Mean Absolute Error (MAE) of approximately 6% and a Root Mean Square Error (RMSE) of 8%.

Overall, the model showed a tendency to overestimate areas affected by weed infestation, as indicated by a positive Mean Bias Error (MBE = 0.22). This behavior was more evident in areas with low weed occurrence, whereas for infestation levels close to 75%, the model tended to underestimate.

This pattern is associated with the limited representation of extreme conditions in the training dataset, particularly pixels with either full sugarcane cover or full weed cover. The scarcity of such samples reduces the model's ability to accurately capture these scenarios. To address this limitation, additional UAV imagery from areas exhibiting these conditions should be incorporated into the calibration dataset.

After calibration and validation, the model was applied to the remaining selected fields, enabling the spatial extrapolation of weed infestation. The total estimated infested area was 4,870 ha, corresponding to approximately 16% of the total area and 36% of the selected fields.

The largest absolute infested areas were associated with cutting stages 5, 2, and 4, which also represent the most extensive cultivated areas (Figure 8).

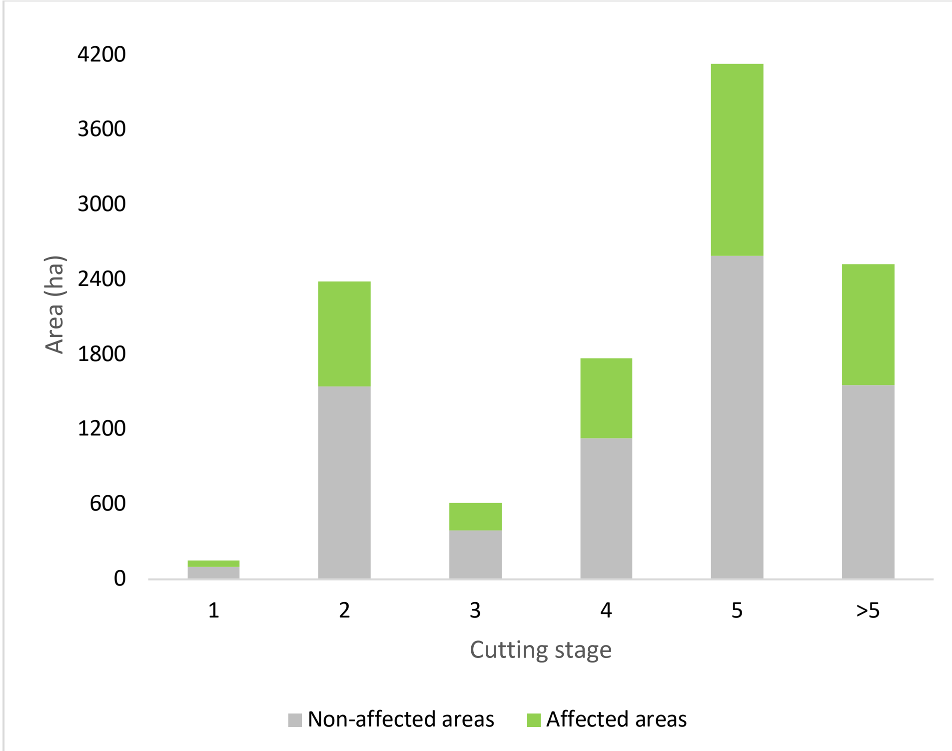

Figure 8. Affected areas by weeds according to cutting stage.

When considering the proportion of infestation relative to the total area within each cutting stage, fields with more than five harvest cycles exhibited the highest infestation levels. This trend is anticipated because sugarcane fields that have been cultivated for a longer time generally get fewer management resources.

Among fields with fewer than five harvest cycles, the second cutting stage showed the highest weed incidence. This finding is particularly relevant, as early infestations in young fields may compromise crop performance over subsequent harvest cycles, considering the average sugarcane lifespan of approximately six years with up to five harvests (CONAB, 2017).

Given the model's tendency to both overestimate and underestimate infestation under different conditions, the results were interpreted as weed infestation alert areas, defined as regions with infestation levels greater than 45%. It is important to note that the presence of weeds does not fully explain crop health status. Areas with low or no detected weed infestation may still present reduced performance due to other factors, such as diseases, planting failures, or plant mortality.

Therefore, the proposed approach should be integrated with complementary monitoring tools, particularly biomass estimation methods, to provide a more comprehensive assessment of crop status. Additionally, the use of soil maps can support more effective weed management planning, as soil properties, such as clay, silt, and sand content, aggregate stability, porosity, and water availability are known to influence weed composition and spatial distribution (Petzold et al., 2020).

Conclusions

This study demonstrates that the proposed model is capable of automatically detecting anomalies associated with weed infestation in sugarcane fields, supporting site-specific management and strategic decision-making by sugarcane growers. The proposed approach proved effective for the detection of weed-related anomalies, reinforcing their applicability as a tool for operational use.

However, due to the timing of aerial image acquisition and the crop development stage of the fields used for model calibration, the analysis was limited to broadleaf weeds under closed-canopy conditions, without species-level discrimination. These limitations highlight the need to expand the calibration dataset and improve model generalization, particularly to enable the detection of grass weeds using drone imagery acquired prior to canopy closure.

Furthermore, achieving a more comprehensive assessment of crop status will require the integration of complementary tools, such as biomass monitoring systems and soil maps, thereby enhancing weed management planning and supporting more precise and efficient agricultural practices.

The SVM model has good evaluation metrics, but other models that are less complex and faster can generate accurate information more efficiently.

References

Alexandridis, T. K., Tamouridou, A. A., Pantazi, X. A., Lagopodi, A. L., Kashefi, J., Ovakoglou, G., Polychonos, V., & Moshou, D. (2017). Novelty detection classifiers in weed mapping: *Silybum marianum* detection on UAV multispectral images. *Sensors*, 17, 2007.

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Moraes, G., Leonardo, J., & Sparovek, G. (2014). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, 22(6), 711–728.

Barrero, O., & Perdomo, S. A. (2018). RGB and multispectral UAV image fusion for Gramineae weed detection in rice fields. *Precision Agriculture*, 19, 809–822.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 281–305.

Brdar, S., Ćulibrk, D., Marinković, B., Crnobarac, J., & Crnojević, V. (2011). Support vector machines with features contribution analysis for agricultural yield prediction. In *Proceedings of the Second International Workshop on Sensing Technologies in Agriculture, Forestry and Environment (EcoSense 2011)* (pp. 43–47).

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Boca Raton, USA: Chapman & Hall/CRC.

Carleer, A., & Wolff, E. (2004). Exploitation of very high resolution satellite data for tree species identification. *Photogrammetric Engineering & Remote Sensing*, 70, 135–140.

Castillejo-González, I. L., Peña-Barragán, J. M., Jurado-Expósito, M., Mesas-Carrascosa, F. J., & López-Granados, F. (2014). Evaluation of pixel- and object-based approaches for mapping wild oat (*Avena sterilis*) weed patches in wheat fields using QuickBird imagery for site-specific management. *European Journal of Agronomy*, 59, 57–66.

CONAB. (2017). *Acompanhamento da safra brasileira de cana-de-açúcar: Safra 2017/18 – Primeiro levantamento*. Brasília, Brazil.

De Castro, A. I., Torres-Sánchez, J., Peña, J. M., Jiménez-Brenes, F. M., Csillik, O., & López-Granados, F. (2018). An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. *Remote Sensing*, 10(2), 285.

De Jeu, R., Wagner, W., Holmes, T. R. H., Dolman, A. J., Van de Giesen, N. C., & Friesen, J. (2008). Global soil moisture patterns observed by spaceborne microwave radiometers and scatterometers. *Surveys in Geophysics*, 29(4–5), 399–420.

Gerhards, R., Bezhin, K., & Santel, H. J. (2017). Sugar beet yield loss predicted by relative weed cover, weed biomass and weed density. *Plant Protection Science*, 53, 118–125.

Kuva, M. A., Gravena, R., Pitelli, R. A., Christoffoleti, P. J., & Alves, P. L. C. A. (2003). Periods of weed interference in sugarcane crop: III – *Brachiaria decumbens*. *Planta Daninha*, 21(1), 37–44.

Lambert, J. P. T., Hicks, H. L., Childs, D. Z., & Freckleton, R. P. (2018). Evaluating the potential of unmanned aerial systems for mapping weeds at field scales. *Weed Research*, 58, 35–45.

López-Granados, F. (2011). Weed detection for site-specific weed management: Mapping and real-time approaches. *Weed Research*, 51, 1–11.

López-Granados, F., Torres-Sánchez, J., Serrano-Pérez, A., De Castro, A. I., Mesas-Carrascosa, F. J., & Peña, J. M. (2016). Early season weed mapping in sunflower using UAV technology. *Precision Agriculture*, 17, 183–199. https://doi.org/10.1007/s11119-015-9415-8

Louargant, M., Jones, G., Faroux, R., Paoli, J. N., Maillot, T., Gée, C., & Villette, S. (2018). Unsupervised classification algorithm for early weed detection. *Remote Sensing*, 10, 761.

Marques, L. S., Mello, F. R., & Chaves, M. (2009). Land use and land cover mapping using remote sensing. In *Proceedings of the National Symposium on Applied Physical Geography* (Vol. 13).

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2015). *e1071: Misc functions of the Department of Statistics*. Vienna, Austria: TU Wien.

Oerke, E. C., & Dehne, H. W. (2004). Safeguarding production—Losses in major crops. *Crop Protection*, 23, 275–285.

Pätzold, S., Mertens, F. M., Bornemann, L., Koleczek, B., Franke, J., & Feilhauer, J. (2008). Soil heterogeneity at the field scale. *Precision Agriculture*, 9, 367–390.

Pätzold, S., Hbirko, C., Dicke, D., Gerhards, R., & others. (2020). Linking weed patterns with soil properties. *Precision Agriculture*, 21, 569–588.

Pérez-Ortiz, M., Peña, J. M., Gutierrez, P. A., Torres-Sánchez, J., Hervás-Martínez, C., & López-Granados, F. (2016). Selecting patterns for weed mapping using UAV imagery. *Expert Systems with Applications*, 47, 85–94.

Randall, N. P., & James, K. L. (2012). The effectiveness of integrated farm management. *Environmental Evidence*, 1, 4.

Rossi, M. (2017). *Soil map of the state of São Paulo*. São Paulo, Brazil: Instituto Florestal.

Rumpf, T., Mahlein, A. K., Steiner, U., Oerke, E. C., Dehne, H. W., & Plümer, L. (2010). Early detection of plant diseases using SVM. *Computers and Electronics in Agriculture*, 74, 91–99.

Santos, H. G., et al. (1999). *Brazilian soil classification system*. Brasília, Brazil: Embrapa.

Santos, H. G., et al. (2013). *Brazilian soil classification system* (3rd ed.). Rio de Janeiro, Brazil: Embrapa.

Souza, M. F., Amaral, L. R., Oliveira, S. R. M., Coutinho, M. A. N., & Netto, C. F. (2020). Spectral differentiation of sugarcane from weeds. *Biosystems Engineering*, 190, 41–46.

Ventura, A. (1964). Technical problems of forestry in São Paulo. *Silvicultura em São Paulo*, 3(3), 61–80.

Walter, A. M., Christensen, S., & Simmelsgaard, S. E. (2002). Spatial correlation between weed species densities and soil properties. *Weed Research*, 42, 26–38.

Wang, A., Zhang, W., & Wei, W. (2019). A review on weed detection using machine vision. *Computers and Electronics in Agriculture*, 158, 226–240.

Willmott, C. J., & Matsuura, K. (2005). Advantages of MAE over RMSE. *Climate Research*, 30(1), 79–82.

Yu, J., Sharpe, S. M., Schumann, A. W., & Boyd, N. S. (2019). Deep learning for weed detection. *European Journal of Agronomy*, 104, 78–84.