Data and Metadata. 2025; 4:194
doi: 10.56294/dm2025194

ORIGINAL

Automatic weed quantification in potato crops based on a modified convolutional neural network using drone images

Cuantificación automática de malas hierbas en cultivos de patata basada en una red neuronal convolucional modificada utilizando imágenes de drones

Kevin Vinueza[1], Lucía Sandoval-Pillajo[1,2], Adriana Giret-Boggino[2], Diego Trejo-España[1], Marco Pusdá-Chulde[1], Iván García-Santillán[1]

[1]Universidad Técnica del Norte. Ibarra, Ecuador.
[2]Universidad Politécnica de Valencia. Valencia, España.

Cite as: Vinueza K, Sandoval-Pillajo L, Giret-Boggino A, Trejo-España D, Pusdá-Chulde M, García-Santillán I. Automatic weed quantification in potato crops based on a modified convolutional neural network using drone images. Data and Metadata. 2025; 4:194. https://doi.org/10.56294/dm2025194

Submitted: 12-05-2024 **Revised:** 19-09-2024 **Accepted:** 20-02-2025 **Published:** 21-02-2025

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding Author: Iván García-Santillán

ABSTRACT

Identifying and quantifying weeds is a crucial aspect of agriculture for efficiently controlling them. Weeds compete with the crop for nutrients, minerals, physical space, sunlight, and water, causing problems in crops ranging from low production to economic losses and environmental deterioration of the land. Weed quantification is generally a manual process requiring significant time and precision. Convolutional Neural Networks (CNN) are very common in weed quantification. Thus, the purpose of this research is the adaptation of the ResNeXt50 CNN architecture for semantic segmentation tasks, focused on the automatic quantification of weeds (Broadleaf dock, Dandelion, Kikuyo grass, and other unidentified classes) in potato fields using RGB images acquired by the DJI Mavic 2 Pro drone. The analytical model was trained following the Knowledge Discovery in Databases (KDD) methodology using Python and the TensorFlow-Keras frameworks. The results indicate that the modified ResNeXt50 model presented a mean IoU of 0,7350, a performance comparable to the values reported by other authors considering fewer weed classes. The Student´s t-test and Pearson correlation coefficient were applied to contrast the weed coverage from the model predictions and the ground truth, indicating no statistically significant differences between both measurements in most weed classes.

Keywords: Weed Quantification; Deep Learning; UAV Images; Semantic Segmentation; CNN; Resnext50.

RESUMEN

Identificar y cuantificar las malas hierbas es un aspecto crucial de la agricultura para controlarlas eficazmente. Las malas hierbas compiten con el cultivo por los nutrientes, los minerales, el espacio físico, la luz solar y el agua, causando problemas en los cultivos que van desde la baja producción hasta pérdidas económicas y el deterioro medioambiental de la tierra. La cuantificación de las malas hierbas suele ser un proceso manual que requiere mucho tiempo y precisión. Las redes neuronales convolucionales (CNN) son muy comunes en la cuantificación de malas hierbas. Así, el propósito de esta investigación es la adaptación de la arquitectura CNN ResNeXt50 para tareas de segmentación semántica, enfocada a la cuantificación automática de malas hierbas (Broadleaf dock, Dandelion, Kikuyo grass, y otras clases no identificadas) en campos de patatas utilizando imágenes RGB adquiridas por el dron DJI Mavic 2 Pro. El modelo analítico se entrenó siguiendo la metodología Knowledge Discovery in Databases (KDD) utilizando Python y los frameworks TensorFlow-Keras. Los resultados indican que el modelo ResNeXt50 modificado presentó un IoU medio de 0,7350, un rendimiento comparable a los valores reportados por otros autores considerando menos clases de maleza. Se aplicaron la prueba t de Student y el coeficiente de correlación de Pearson para contrastar la cobertura de maleza a partir de las predicciones del modelo y la verdad sobre el terreno, indicando que no había diferencias estadísticamente significativas entre ambas mediciones en la mayoría de las clases de maleza.

Palabras clave: Cuantificación de Maleza; Deep Learning; Imágenes UAV; Segmentación Semántica; CNN; Resnext50.

INTRODUCTION

Problem statement

According to the United Nations postulate, the demand for food is expected to grow to 70 % by 2050. This means that the agricultural sector will have to increase its production to satisfy the food needs, and this increase must be sustainable and without significantly harming the environment.(1)

Weeds are undesirable plants because they harm crops by taking over nutrients and space. This results in nutritional deficiency problems that generate economic losses for farmers.(2) One method used to combat weeds is chemical control. However, farmers commonly apply these chemicals evenly throughout the field,(3) affecting the crops.(4)

Precision agriculture (PA) uses technological tools to improve methods of combating weeds.(5) One approach used for weed identification is Deep Learning, whose algorithms allow the automatic extraction of features from large amounts of data.(6) Convolutional Neural Networks (CNNs) have become very popular in modern years(7,8,9,10,11) since current hardware allows these algorithms to be used efficiently 6,12. CNNs have proven to offer good results concerning weed identification.(13,14,15) This identification task is challenging since the plants (crops and weeds) share similar physical characteristics, the plants overlap, and they are influenced by certain environmental conditions.(16)

Weed quantification is an activity that allows for optimal decision-making to control these plants. This indicator helps to determine their coverage and, based on that, estimate the optimal amount of herbicide to use without significantly harming the crop.(17) Manual weed quantification takes too much time. Thus, many farmers choose to carry out this identification and quantification in a very subjective and incorrect way to reduce time. However, this approach leads to unreliable data, making weed control decision-making inefficient.(18,19) Identifying weeds using CNNs makes this process more efficient, as the results are objective.(17) The goal is to differentiate weeds from crops and other terrain features, and weed cover is estimated by measuring it in a specific area.

The objective of this research is to identify and quantify four categories of weeds: Broadleaf dock (Rumex obtusifolius), Dandelion (Taraxacum officinale), Kikuyo grass (Pennisetum clandestinum), and other unidentified classes in potato crop fields using drone images. For this end, the ResNeXt50 CNN(20) is adapted for semantic segmentation. Potato crops was selected because it is an icon in the northern area of Ecuador, where this research is carried out. The categories of weeds treated are the most prevalent ones observed in the region. Six versions of this algorithm are tested, with modifications to the architecture to make it more specialized and robust for weed segmentation. This adaptation and specialization of the network represents an important improvement and contribution to the field of study.

The methodology used for image data analysis is Knowledge Discovery in Databases (KDD),(21) which allows the corresponding tasks to be carried out in a structured and orderly manner, from acquiring images on the ground to validating and using the algorithm. The algorithm was developed using Python and the TensorFlow-Keras framework.

A dataset containing images of potato crops in Carchi-Ecuador province was collected in 2023. It includes 1500 images of potato crop plants and various types of weeds. The dataset (images and masks) is publicly available on the GitHub platform for research and

comparison purposes at https://bit.ly/4gi7I3O. This symbolizes another important contribution due to the difficulty and time involved in collecting and annotating various weeds in images with their semantic segmentation masks.

Related works

Below, we summarize some works closely related to this study that served as a theoretical basis and for comparison purposes:

Cai et al.(22) proposed an improved PSPNet network architecture for crop weed identification. The images were collected in a pineapple field in China using the DJI Mavic 2 UAV, with a flight height and speed of 5 m and 2 m/s, respectively. 34 images of 5472×3648 with a 0,1 cm/pixel resolution were captured. The training environment was the TensorFlow 2 framework. The best results achieved a Mean IoU of 79,73 %.

Nong et al.[23] proposed a method for semantic segmentation of weeds, called SemiWeedNet, to perform semi-supervised training using the DeepLabv3++ architecture. The dataset used was WeedMap, acquired in sugar crops from UAVs. 289 RGB images were chosen, where the present classes were crop, weed, and background. The training environment selected was the Pytorch framework. As a result, the model reached a mean IoU of 0,701 for ResNet50 as a backbone and 0,700 for ResNet101.

Shahi et al.[24] compared CNN architectures with different backbones for binary and multiclass semantic weed segmentation using the public CoFly-WeedDB dataset, which provides 201 RGB images (1280×720). The images were obtained using a DJI Phantom Pro 4 drone over a cotton field in Greece. A total of 3 weed types were detected with a class imbalance. The training environment was Google Colab. The results show that for binary segmentation, the best model was SegNet with DenseNet121 as the backbone with a Mean IoU of 67,56 %, while for multiclass segmentation, it was U-Net with EfficientNetB0 as the backbone with a Mean IoU of 56,21 %.

Gao et al.[5] developed a method for weed segmentation and mapping in maize fields. The UAV-based image data was collected in Belgium. The study classified three classes (weed, maize, soil) regardless of different weed species. A 12 coaxial rotors UAV, equipped with a lightweight visual camera, at 20 m above ground altitude, 6000×4000 was used. The images consisted of over 10,000 images. The deep learning architecture used in the study is similar to U-Net and SegNet, with a higher number of network parameters. The study utilized Python and the TensorFlow framework, achieving a mean IoU of 0,767 in the field test dataset and 0,617 in the UAV orthomosaic imagery.

The literature shows that the number of weed classes identified is limited, mainly addressing a binary semantic segmentation task (weed and background) or a multiclass of 3 categories (crop, weed, and background).

The remainder of the manuscript is as follows: Section 2 presents the methodology, dataset, and software used in developing this study for detecting and quantifying weeds. Section 3 shows the main results and statistical tests obtained in this proposal. Section 4 indicates the discussion with other related research, and finally, the conclusions and future work are presented in Section 5.

METHOD

KDD Methodology

The Knowledge Discovery in Databases (KDD) methodology[21] was used for image data management and analysis, which consists of five phases: Data collection; Data selection, preprocessing, and transformation; Data mining; Evaluation and Interpretation. These are detailed below:

Data collection

RGB mages (5472×3648) were collected from 7 different potato crop fields in Carchi- Ecuador, to incorporate variability in climatic conditions, land, and plants. 2447 images were captured from May to December 2023. The DJI Mavic 2 Pro UAV was used with the Android DroneDeploy application v5.7 to configure all the flight parameters: flight height 9 m, flight speed 1 m/s, and cm/px ratio 0,25. Figure 1 shows examples of images collected in potato fields containing different types of weeds.

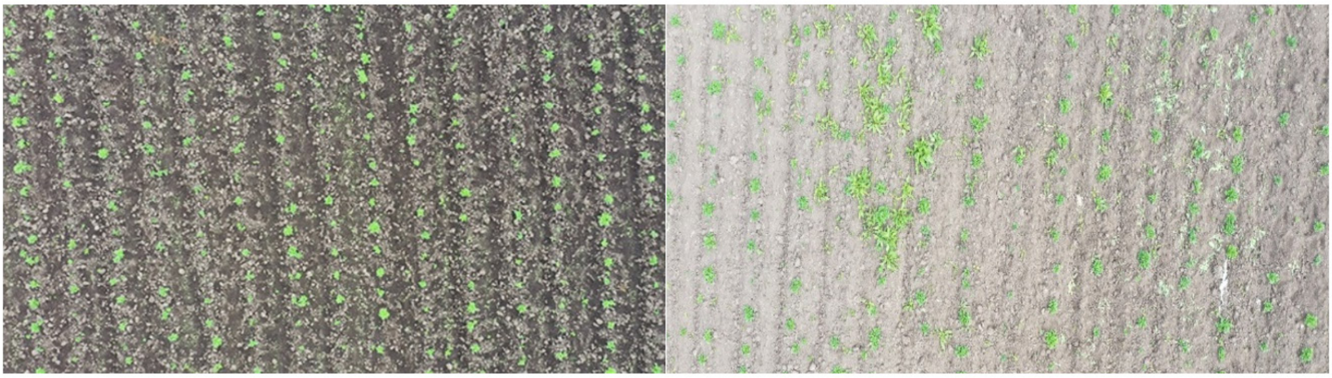


Figure 1. Examples of RGB images (5472×3648) taken at 9 m height by the DJI Mavic 2 Pro UAV in potato fields in Carchi-Ecuador

Data selection, preprocessing, and transformation

Images not suitable for the research were discarded according to the following criteria: (i) images that do not contain weeds, (ii) images with an exaggerated density of weeds, (iii) adjacent images, and (iv) blurred images. After image selection, 108 RGB images were maintained. 250×250 sub-image extraction from the images is used for manual image annotation.[25] This choice is because such resolution almost completely covers most large plants in the original images. Adobe Photoshop 2020 was used to delimit and extract the sub-images, as indicated in figure 2.

Figure 2. Example of delimiting and extracting a 250×250 sub-image

During the extraction of the sub-images, each was categorized, considering the most significant type of plant present in each sub-image, which may contain various kinds of plants.

Weeds with the greatest coverage in the seven crop fields visited were chosen, mainly Broadleaf dock, Dandelion, and Kikuyu. In addition, a category called "other weeds" was used for plants that were difficult to identify visually (unknown or small plants) and those with a low presence in the fields visited.

A total of 1500 correctly classified 250×250 sub-images were obtained. This task required approximately 25 hours of work. Each extracted sub-image was then resized to a resolution of 128×128 for image annotation for semantic segmentation and to be compatible with the input of the RestNeXt50 CNN architecture detailed below. The online platform used for image annotation was Roboflow,[26] which offers functionality to determine the number of individuals (plants) in the entire dataset, as detailed in table 1.

Table 1. Number of individuals for each type of plant and their percentage in the dataset

Type of plant	Number of plants	%
Broadleaf dock	978	14,43
Dandelion	321	4,73
Kikuyo	2670	39,41
Other weeds	997	14,71
Potato (crop)	1808	26,69
Total	6774	100

The class with the fewest individuals is Dandelion (4,73 %), which causes a significant class imbalance in the dataset. Manual annotation on the 1500 128×128 sub-images (6774 plants) required 150 work hours and took approximately 6 minutes for each sub-image.

The annotated dataset was split into training, validation, and testing sets in the ratio of 80 % (1200 sub-images), 10 % (150 sub-images), and 10 % (150 sub-images), respectively. The 80/10/10 split is a good balance in general, where one wants to maximize data usage for model training without sacrificing adequate validation (hyperparameter tuning) or testing of the model's generalization ability on unseen data.

Data augmentation was then performed to extend the dataset size and improve the model training. The operations applied to the sub-images were a 25 % increase and decrease in brightness, Gaussian blur with a 5×5 kernel, and rotations at angles of 90, 180, and 270°. Seven types of data augmentation were applied, obtaining 8400 images for the model training set.

Data mining

This phase consists of 2 main parts: the choice of the Deep Learning algorithm and its training

The Deep Learning Algorithm (ResNeXt50)

The ResNeXt[20] is an improved variant of the residual network ResNet,[27] that uses residual blocks to train very deep networks more effectively. In this study, the ResNeXt50 neural network architecture with 50 layers was specifically used, with certain adaptations in the architecture to perform semantic segmentation. ResNeXt50 maintains the balance between computational efficiency and performance due to the addition of cardinality. Cardinality is the number of parallel transformations (branches) performed on the data within a residual block. For example, if a block has a cardinality of 32, it means that 32 branches are executing parallel transformations. Cardinality is a key concept introduced in the ResNeXt architecture to improve the network's ability to learn rich and diverse features without significantly increasing computational complexity. In this study, the cardinality was C=2, since there would only be two branches within the residual block, seeking a balance between computational requirements and accuracy.

For ResNeXt to be used for semantic segmentation, the ResNet architecture was modified,[28] removing the fully connected layers and replacing them with convolutional layers and upsampling layers to obtain the semantic segmentation mask, as shown in figure 9. The modified ResNeXt architecture generally consists of two parts: an encoder and a decoder. The encoder performs feature extraction from the image, while the decoder is responsible for reconstruction and semantic segmentation.

ResNeXt50 training

Six versions of the ResNeXt50 architecture were proposed, where modifications were made to the architecture and the hyperparameters used by different authors.[29,30,31,32,33,34] One-hot encoding was used for ResNeXt50 predictions (6 classes). The training environment chosen was Kaggle, which used GPUs.

Baseline model (version 1)

The baseline model is the original ResNeXt50 architecture[28] for semantic segmentation. The training hyperparameters chosen were batch size=8, number of epochs=100, learning rate=0,0001, and optimizer=Adam.

Version 2 of the model

The architecture of the baseline model has 8×8, 16×16, and 32×32 upsampling layers in the decoder part, which is responsible for the semantic segmentation mask. This causes the feature maps to increase their dimension too quickly. Therefore, the first proposal for version 2 was to replace these upsampling layers with 2×2 dimensions so that the feature maps increase their dimension gradually and not abruptly. For example, the single 8×8 upsampling layer was replaced by 3 2×2 layers, the 16×16 layer by 4 2×2 layers, and the 32×32 layer by 5 2×2 layers. This makes the silhouettes of the plants in the semantic segmentation masks more similar and fit the ground truth, resulting in a more specialized type of architecture for semantic weed segmentation using UAV images. This idea is because it is better to apply masks of small sizes in a cascade fashion.[36] These changes allow the network to better focus on different object scales and fine details, delineating weed edges more clearly (including small plants) and improving the accuracy of semantic segmentation.

The second major modification was the replacement of the ReLU activation function with the Mish function[37] which offers a negative range of values, a greater capacity to model complex features (in small weeds), and an ability to improve model generalization.

The third proposal in this version was to implement a function that decreases the value of the learning rate as the model training epochs progress so that it converges as far as possible to the global minimum in a more controlled manner than at the beginning of training.[38]

To this end, where the cosine annealing function is used to vary the learning rate. Thus, the learning rate decreases as a cosine function over the epochs, allowing smooth fluctuations and continuous decay.[39]

Figure 9 visualizes the resulting diagram of the architecture of version 2 of the model. The training hyperparameters chosen here were batch size=8, number of epochs=100, learning rate=0,0001 with decay every ten epochs (cycle) using the cosine annealing function, and optimizer=Adam.

Version 3 of the model

The modifications already proposed and the foundations of the ResNeXt architecture are considered, where it indicates that cardinality helps to improve the prediction results since the network should be wider than deep.[20] Therefore, the same identity blocks used in the encoder part of the network were implemented in the decoder part, which was in charge of generating the semantic segmentation mask. The resulting architecture of the model can be seen in figure 9. The training hyperparameters chosen here were as before: batch size=8, number of epochs=100, learning rate=0,0001 with decay every ten epochs (cycle) using the cosine annealing function, and optimizer=Adam.

Versions 4, 5, and 6 of the models

Because the architecture proposed in version 3 of the model was sufficiently robust in preliminary results, it was decided to use it in versions 4, 5, and 6 but testing with different combinations of the training hyperparameters indicated in table 2.

https://doi.org/10.56294/dm2025194

Table 2. Hyperparameter values used in versions 4, 5, and 6 of the models

Hyper parameters	Version 4	Version 5	Version 6
Batch size	16	32	2
Number of epochs	100	200	100
Learning rate	0,00001	0,0001 with decay using the cosine annealing function.	0,001 with decay using the cosine annealing function.
Optimizer	Adam	Adam	Adam

Evaluation and Interpretation

For the performance evaluation of the six versions of ResNeXt50, three common metrics in semantic segmentation were used: dice loss, mean dice coefficient, and mean intersection over union (IoU).

Dice loss is a function used, especially when imbalanced classes exist. It measures the match between the ground truth masks and the one predicted by the model at the shape and overlap level. Its range is defined between [0,1], with 0 indicating a perfect match between the segmentation masks.(40)

Mean Dice Coefficient is a metric used to evaluate the accuracy of models that segment or classify pixels into different classes. Its value range is between [0,1], with 1 indicating a perfect match between segmentation masks.(35)

Mean IoU is another widely used metric to evaluate the accuracy of semantic segmentation models. It provides a clear and balanced view of the model's performance across all classes and is robust in class imbalance situations. Its value range is between [0,1], where values close to 1 indicate that the model has excellent segmentation performance and high accuracy in class prediction.

Regarding the interpretation of the model, the purpose is to give a value meaning to the predictions made by the model, providing information that is understandable and useful for decision-making. Equation 1 is useful for calculating weed coverage:(17)

$$\text{Weed coverage} = (\text{Weed area})/(\text{Total area}) \times 100\ \%$$

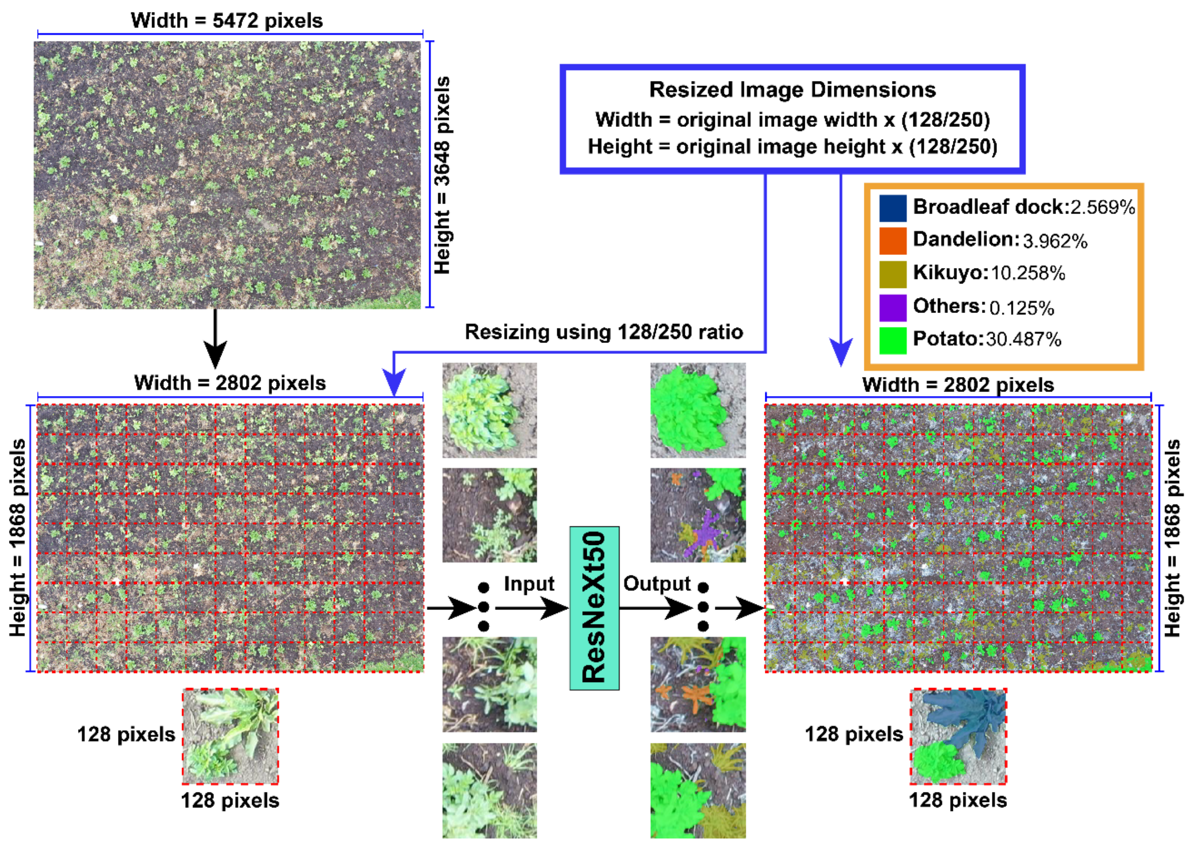


Figure 3. The entire process for automatic weed quantification from UAV images using the best model ResNeXt50

To obtain the weed area, the number of pixels belonging to each category (Broadleaf dock, Dandelion, Kikuyu, other weeds) provided by the model's semantic segmentation mask must be calculated. The entire process for weed quantification is shown in figure 3.

As the 2802×1868 images are post-processed, the pixels belonging to each category of plant found are counted, and then weed coverage (equation 1) is applied for each type of plant.

Finally, to contrast the weed coverages calculated from the predictions of the best model (ResNeXt50) and the ground truth, the Student's t-test and Pearson correlation coefficient (r) statistical tests were used to verify whether there are statistically significant differences and a linear association between both measurements,[41] respectively.

RESULTS

ResNeXt50 performance metrics

Figures 4 to 6 show the training curves of the six proposed versions of ResNeXt50 on the validation set (10 %) using the three-performance metrics.

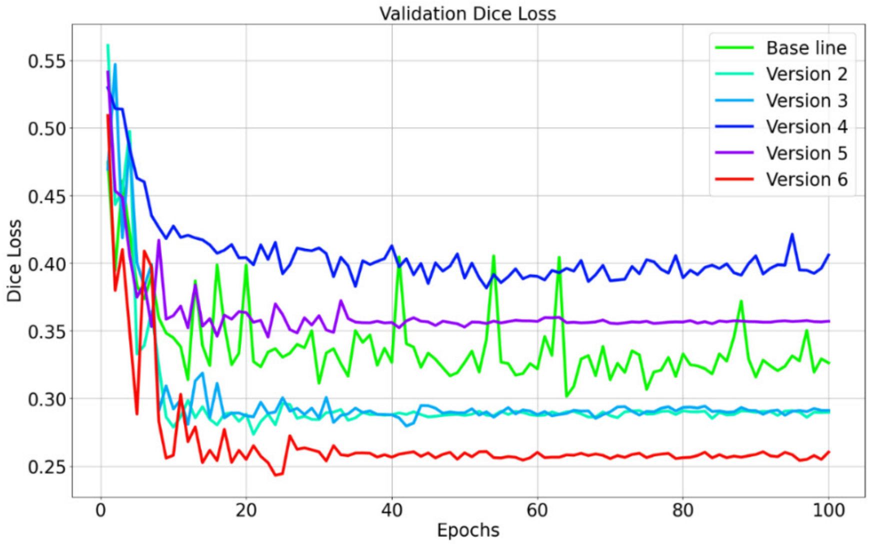


Figure 4. Dice loss metric on the validation set of the six proposed versions of ResNeXt50

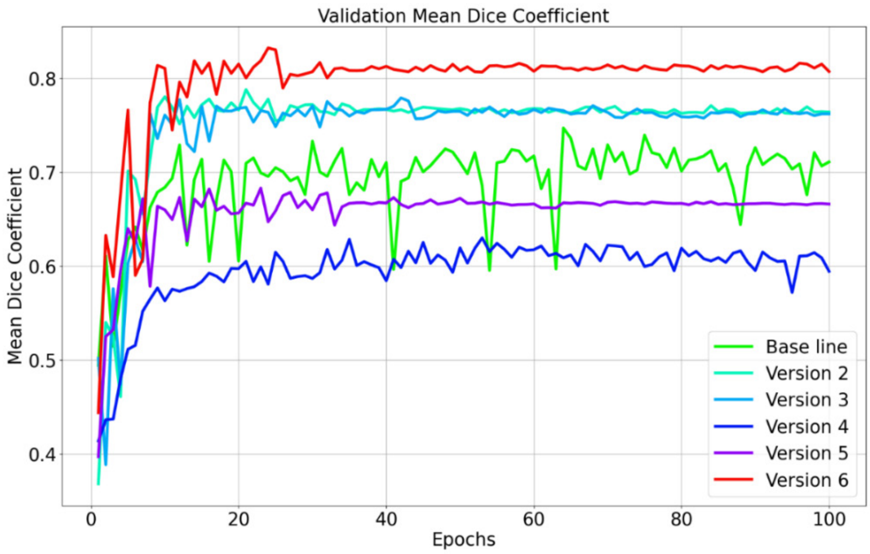


Figure 5. Mean dice coefficient metric on the validation set of the six proposed versions of ResNeXt50

Table 3 summarizes the values obtained in the validation stage of the evaluation metrics of the six versions of ResNeXt50.

The proposed ResNeXt50 version 6 performed best in all evaluation metrics. This indicates that the modified network architecture and the optimal combination of hyperparameters played important roles in the model's performance.

Table 4 details the IoU obtained in each class of the best ResNeXt50 v6 model dataset on the test set (10 %, 150 images), that is, on data not seen during training.

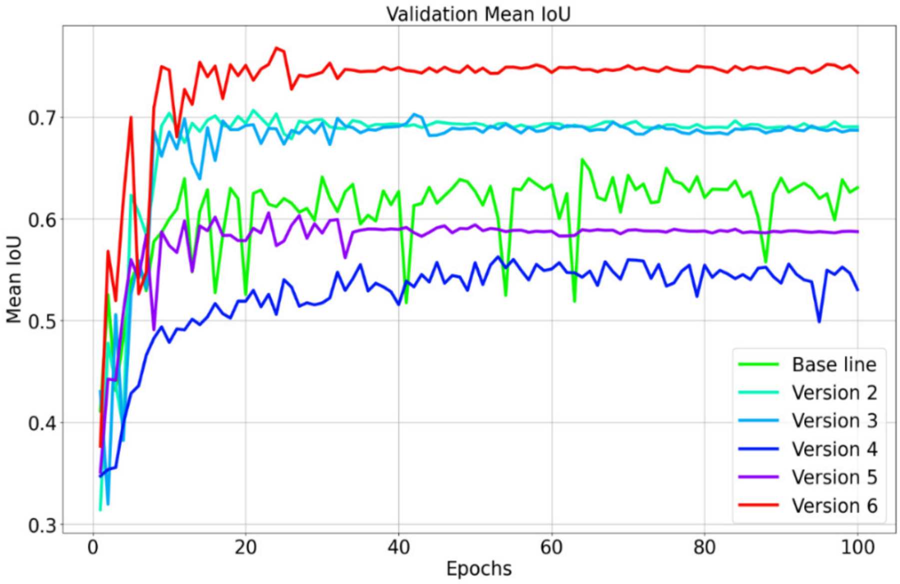


Figure 6. Mean IoU metric on the validation set of the six proposed versions of ResNeXt50

Table 3. Evaluation metrics were reached in the six versions of ResNeXt50 on the validation set (10 % of the dataset, 150 images) for 100 epochs

Version	Dice loss	Mean dice coefficient	Mean IoU
Baseline	0,326	0,710	0,630
Version 2	0,289	0,763	0,690
Version 3	0,291	0,761	0,686
Version 4	0,406	0,593	0,529
Version 5	0,357	0,665	0,586
Version 6	0,260	0,806	0,743

Table 4. IoU metric of each ResNeXt50 v6 model dataset class on the test set (10 %, 150 images)

Version	Background IoU	Broadleaf dock IoU	Dandelion IoU	Kikuyo IoU	Other weeds IoU	Potato IoU	Mean IoU	Dice Loss	Mean Dice Coefficient
Version 6	0,956	0,826	0,583	0,703	0,434	0,908	**0,735**	0,265	0,787

The class with the highest IoU is the background, which agrees with results obtained by other researchers[22,23,24] due to its significant presence in all the images in the dataset. The second class with the highest IoU value is Potato, which is in the same situation. The class with the lowest IoU is "other weeds" because this class encompasses the different types of weeds that could not be visually identified as a specific category. This makes it difficult for the CNN to find and learn unique patterns of the plants in this category, significantly harming the mean IoU of the model. Still, the best version 6 obtained a mean IoU of 0,735, indicating that, on average, the model segments all six classes well. In addition, a Mean Dice Coefficient=0,787 and Dice Loss=0,265 were obtained.

The performance of the model on the validation (table 3) and test (table 4) sets are comparable, i.e., the difference in each evaluation metric is small (≤ 2 %), indicating that both data sets are representative, and the model generalizes well to unseen data. Figure 7 shows the RGB images, ground truth masks, and predictions of the best version of RestNeXt50 v6 on four example sub-images of 128×128.

The predicted segmentation mask matches the ground truth, especially in the image (c), which is more evident due to several weeds.

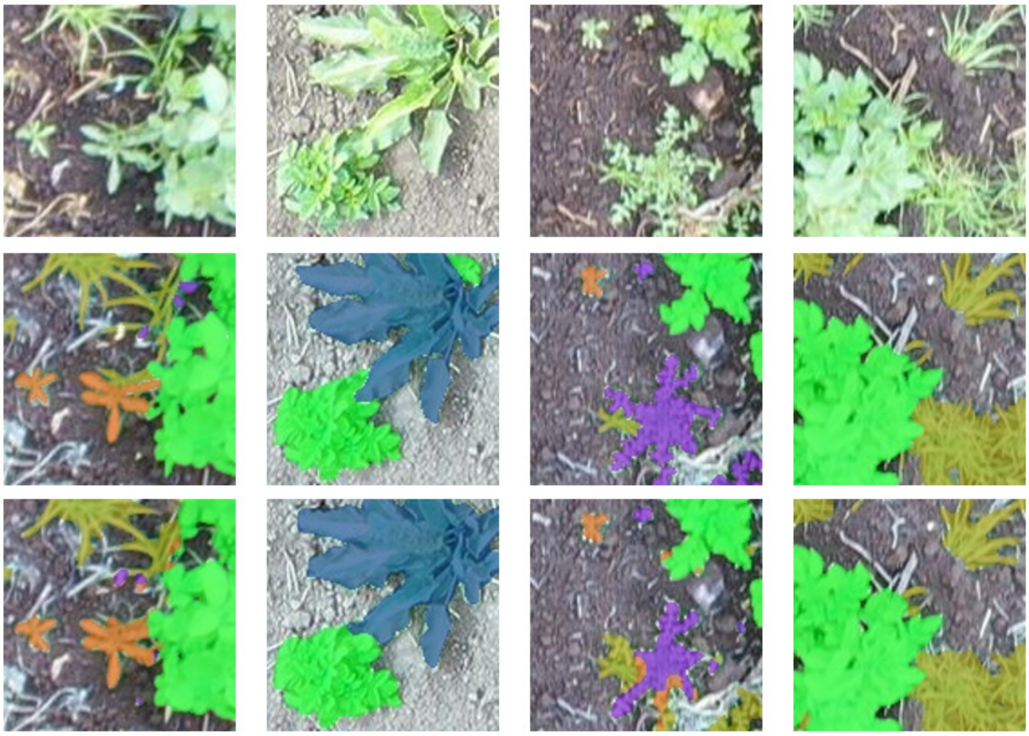


Figure 7. RGB images (first row), Ground truth (second row), and ResNeXt50 v6 model prediction (third row) on four examples of 128×128 sub-images. Kikuyu weed appears in mustard yellow, dandelion in orange, yellow dock in blue, potato crop in green, and other weeds in purple

Statistical validation of model predictions

The Student's t-test is a parametric statistical test to determine whether there is a significant difference in the weed coverage means between the ground truth and the values predicted by the best model (version 6). This allows us to objectively determine whether the predictions of ResNeXt50 v6 are reliable. In this case, they are considered dependent samples in the test since the manual assessments (ground truth) and those predicted by the model are carried out on the same set of images.[42] The null hypothesis H0 and alternative hypothesis H1 are established as follows:

H_0: d=0 (There is no significant difference in the weed coverage means of both groups)

H_1: $d \neq 0$

The decision rule is: If p-value ≤ 0,05, reject H_0.

Statistical evaluation was applied on the test set (10 %) containing 150 images of 128×128 using IBM SPSS v26 software. The paired t-test assumes the differences between paired observations (ground truth and model) follow a normal distribution. The sample is large enough (n ≥ 30) so that, according to the Central Limit Theorem, the differences approximate a normal distribution.[42] This makes the t-test valid and robust, even without explicit verification of the normality of the differences.

Applying the Student´s t-test on the coverage of the Broadleaf dock weed, a p-value = 0,068 was obtained. Therefore, according to the decision rule, H_0 is not rejected, indicating no significant difference between coverage measurements. The Pearson correlation coefficient r = 0,995 (p < 0,05) indicates a very high linear association between both measurements.

Using the t-test on the coverage of the Dandelion, a p-value = 0,726 was reached, indicating no significant difference between both coverage measurements. The coefficient r = 0,892 (p < 0,05) means a high linear association between both measurements.

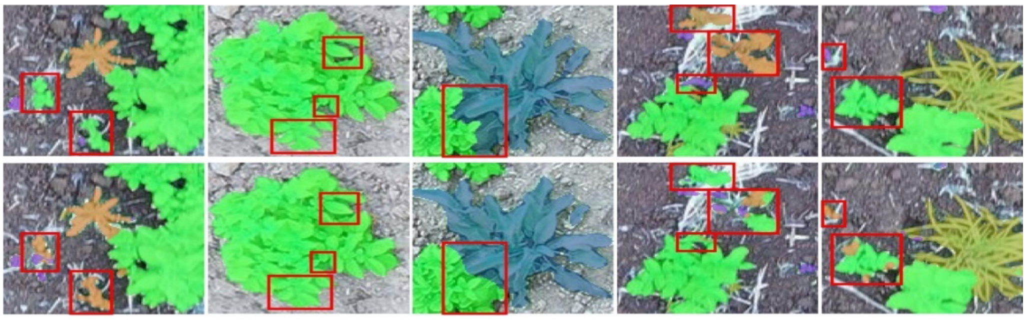
Utilizing the t-test on the coverage of the Kikuyo, a p-value = 0,323 was achieved, indicating no significant difference between both coverage measurements. The coefficient r = 0,994 (p < 0,05) suggests a very high linear association between both measurements.

Continuing with the t-test on "other weeds" coverage, a p-value = 0,039 was obtained, indicating a significant difference between both coverage measurements. The coefficient r = 0,953 (p < 0,05) shows a very high linear association between both measurements.

Lastly, with the t-test on the coverage of potato crops, a p-value = 0,027 was reached, indicating a significant difference between both coverage measurements. The coefficient r = 0,997 (p < 0,05) means a very high linear association between both measurements.

It was concluded that the ResNeXt50 v6 model has a high concordance concerning the predictions of the three specific weed types (Broadleaf dock, Dandelion, and Kikuyo). However, this level of agreement is reduced for detecting weeds belonging to the "other classes" category and potato crops. This may be due to the significant imbalance of classes in the dataset since they are the minority and majority categories. Furthermore, this could be due to several factors related to unusual characteristics of the plants (size, shape, density), variability in light conditions, capture angles, presence of shadows, overlapping plants, abrupt changes in the texture, presence of species not adequately represented in the training data, noise in the images (mislabeled, blurry examples) and foreign objects (stones, sticks, branches).

Focusing on the potato class, a high IoU of 0,908 is maintained (table 4), but contradictorily, there are significant differences between the actual and predicted measurements according to the Student's t-test. This means that, although the model predicts the potato well, on average, there are certain atypical images where the coverage prediction is substantially different from the actual coverage (ground truth). This was evidenced by the presence of moderate and extreme outliers identified in a box plot regarding the significant difference in coverage between the Ground truth and the model v6. Figure 8 shows some example images with significant coverage differences in the test set due to some unusual factors of the plants and the dataset mentioned above.



(a) (b) (c) (d) (e)

Figure 8. Images (128×128) of the ground truth (first row) and predicted (second row) with the largest differences in plant coverage on the test set. Kikuyu weed appears in mustard yellow, Dandelion in orange, Broadleaf dock in blue, Potato crop in green, and other weeds in purple

The model v6 was trained on the Kaggle platform for 100 epochs in 12,9 hours (considered medium), with a relatively fast inference time (128×128 image) of 94 milliseconds (10,6 fps) and a large size of 668 MB.

DISCUSSION

The ResNeXt50 architecture, was used for the semantic segmentation task. Six versions are proposed. The first 3 modify the architecture so that the upsampling masks better fit the silhouettes and fine details of the plants. This adaptation and specialization of the network represents an important improvement and contribution to the field of study. The other three remaining versions experimented with different hyperparameters.

According to table 3, model version 6 is the best model, with a mean IoU of 0,743. Version 4 is the worst model, with a mean IoU of 0,529, whereas the baseline version obtains a modest mean IoU of 0,630. This shows that the modifications proposed in our work to the architecture in the upsampling layers are successful, adjusting better to the silhouettes of the plants, considering the small size ones that turn out to be a very challenging task using UAV images.[43]

Table 5 compares our proposal against different CNN models used in semantic weed segmentation using UAV images on the test set.

Table 5. Comparison of CNN models used in semantic weed segmentation using UAVs

Authors	Year	Model architecture	Number of classes	Mean IoU
Cai et al.[22]	2023	PSPNet with ECA module	2 classes: weed and background	0,7973
Nong et al.[23]	2022	DeepLabv3++ with ResNet50 backbone	3 classes: crop, weed, and background	0,7010
Shahi et al.[24]	2023	SegNet with DenseNet121 backbone U-net with EfficientNetB0 backbone	2 classes: weed and background 4 classes: Johnson grass, field bindweed, purslane, and background	0,6756 0,5621
Gao et al.[5]	2024	Similar to U-Net and SegNet	3 classes: weed, maize, and soil	0,767
This proposal	2024	ResNeXt50 v6 adapted for semantic segmentation	6 classes: Broadleaf dock, Dandelion, Kikuyo, other weeds, Potato, and background	0,7350

https://doi.org/10.56294/dm2025194

The PSPNet model is the best, with a mean IoU of 0,7973, followed by the U-net architecture with mean IoU=0,767 and our ResNeXt v6 model with a mean IoU of 0,7350. All three models have the mean IoU metrics in the same order of magnitude, indicating no drastic difference. Furthermore, our model is superior to the other models, DeepLabv3++, SegNet, and U-net, with 0,7010, 0,6756, and 0,5621, respectively.

Regarding the number of classes, the value in Cai et al.[22] is the best mean IoU focused on identifying only two categories, background and weeds, while our work analyzes six classes: background, potato crop, and four types of weeds. Shahi et al.[24] propose two approaches: the first one, which makes a binary classification with background and weeds, and the other multiclass approach that analyzes four classes, the background and three types of weeds, obtaining a mean IoU of 0,6756 and 0,5621, respectively. In Gao et al. 5 analyze only three classes (weed, maize, and soil), unlike ours, which identifies six classes. This shows that the greater the number of categories in the data set, the lower the value in the mean IoU metric. However, our study provides a semantic segmentation model with a mean IoU comparable to other algorithms but trained on a dataset with a larger number of classes, representing an important contribution to the field of study.

Regarding this study's limitations: The dataset's number of classes is restricted to six categories. The collected and annotated dataset has a significant imbalance of classes, mainly concerning the Dandelion and "other weeds" categories. Another restriction is the exclusive use of the ResNeXt50 architecture focused on semantic segmentation, leaving aside other current deep architectures or a combination of them.

Furthermore, most embedded devices consider our model large (668 MB). Finally, our work uses a low-cost UAV to capture RGB images, mainly because they are popular and affordable, offering a cost-efficient solution.[17]

CONCLUSIONS

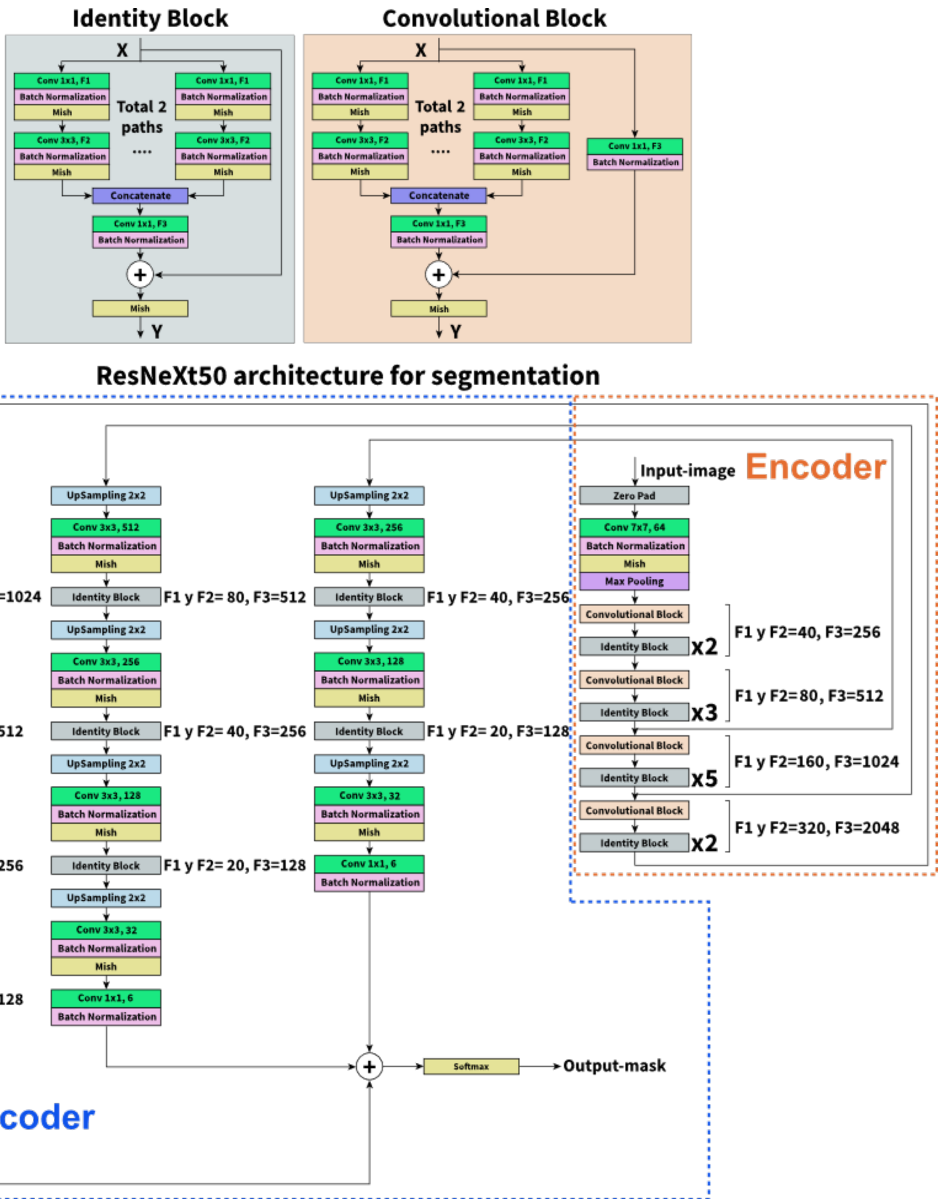


Figure 9. The modified architecture of the ResNeXt50 model for semantic segmentation (our proposal)

The study proposes an adapted version of the ResNeXt50 architecture for the semantic segmentation of several types of weeds from drone images. Six categories are successfully processed: background, Potato, Broadleaf dock, Dandelion, Kikuyu, and other weeds. The adaptation consists mainly of modifying parameters of the upsampling layers in the decoder of the architecture used to better adjust the resulting segmentation to the silhouettes of the plants under study. Model training uses a custom dataset from several lands in Carchi-Ecuador. The model obtains comparable results (mean IoU=0,735) to those existing in the literature (table 5), reaching a relatively fast inference time (128×128 image) of 94 milliseconds (10,6 fps).

Future work would suggest increasing the variability of images and weed classes in the data set, prioritizing class balancing. It would also be advised to experiment with some combination of modern deep networks based on CNN or Transformers. Lighter models optimized for embedded devices applying techniques such as quantization or pruning are also recommended. Weed mapping is necessary for creating georeferenced maps that indicate the spatial distribution and density of weeds within a field following this research line.

BIBLIOGRAPHIC REFERENCES

1. United Nations. THE 17 GOALS | Sustainable Development 2018. https://sdgs.un.org/goals (accedido 21 de septiembre de 2024).

2. Babaei-Ghaghelestany A, Alebrahim MT, Farzaneh S, Mehrabi M. The anticancer and antibacterial properties of aqueous and methanol extracts of weeds. J Agric Food Res 2022;10:100433. https://doi.org/10.1016/j.jafr.2022.100433.

3. Paušič A, Tojnko S, Lešnik M. Permanent, undisturbed, in-row living mulch: A realistic option to replace glyphosate-dominated chemical weed control in intensive pear orchards. Agric Ecosyst Environ 2021;318:107502. https://doi.org/10.1016/j.agee.2021.107502.

4. Zimdahl RL. Weed Reproduction and Dispersal. Fundam. Weed Sci., Elsevier; 2018, p. 83-121. https://doi.org/10.1016/B978-0-12-811143-7.00005-6.

5. Gao J, Liao W, Nuyttens D, Lootens P, Xue W, Alexandersson E, et al. Cross-domain transfer learning for weed segmentation and mapping in precision farming using ground and UAV images. Expert Syst Appl 2024;246:122980. https://doi.org/10.1016/j.eswa.2023.122980.

6. Coulibaly S, Kamsu-Foguem B, Kamissoko D, Traore D. Deep Convolution Neural Network sharing for the multi-label images classification. Mach Learn Appl 2022;10:100422. https://doi.org/10.1016/j.mlwa.2022.100422.

7. Chacua B, García I, Rosero P, Suárez L, Ramírez I, Simbaña Z, et al. People Identification through Facial Recognition using Deep Learning. 2019 IEEE Lat. Am. Conf. Comput. Intell. -CCI, 2019, p. 1-6. https://doi.org/10.1109/LA-CCI47412.2019.9037043.

8. Montenegro S, Pusdá-Chulde M, Caranqui-Sánchez V, Herrera-Tapia J, Ortega-Bustamante C, García-Santillán I. Android Mobile Application for Cattle Body Condition Score Using Convolutional Neural Networks. En: Narváez FR, Urgilés F, Bastos-Filho TF, Salgado-Guerrero JP, editores. Smart Technol. Syst. Appl., vol. 1705, Cham: Springer Nature Switzerland; 2023, p. 91-105. https://doi.org/10.1007/978-3-031-32213-6_7.

9. Cevallos M, Sandoval-Pillajo L, Caranqui-Sánchez V, Ortega-Bustamante C, Pusdá-Chulde M, García-Santillán I. Morphological Defects Classification in Coffee Beans Based on Convolutional Neural Networks. En: Valencia-García R, Borodulina T, Del Cioppo-Morstadt J, Moran-Castro CE, Vera-Lucio N, editores. Technol. Innov., vol. 2276, Cham: Springer Nature Switzerland; 2025, p. 3-15. https://doi.org/10.1007/978-3-031-75702-0_1.

10. Ulloa F, Sandoval-Pillajo L, Landeta-López P, Granda-Peñafiel N, Pusdá-Chulde M, García-Santillán I. Identification of Diabetic Retinopathy from Retinography Images Using a Convolutional Neural Network. En: Valencia-García R, Borodulina T, Del Cioppo-Morstadt J, Moran-Castro CE, Vera-Lucio N, editores. Technol. Innov., vol. 2276, Cham: Springer Nature Switzerland; 2025, p. 121-36. https://doi.org/10.1007/978-3-031-75702-0_10.

11. Salazar-Fierro F, Cumbal C, Trejo-España D, León-Fernández C, Pusdá-Chulde M, García-Santillán I. Detection of Scoliosis in X-Ray Images Using a Convolutional Neural Network. En: Valencia-García R, Borodulina T, Del Cioppo-Morstadt J, Moran-Castro CE, Vera-Lucio N, editores. Technol. Innov., vol. 2276, Cham: Springer Nature Switzerland; 2025, p. 167-83. https://doi.org/10.1007/978-3-031-75702-0_13.

12. Valizadeh M, Wolff SJ. Convolutional Neural Network applications in additive manufacturing: A review. Adv Ind Manuf Eng 2022;4:100072. https://doi.org/10.1016/j.aime.2022.100072.

13. Espejo-Garcia B, Mylonas N, Athanasakos L, Fountas S, Vasilakoglou I. Towards weeds identification assistance through transfer learning. Comput Electron Agric 2020;171:105306. https://doi.org/10.1016/j.compag.2020.105306.

14. McCool C, Perez T, Upcroft B. Mixtures of Lightweight Deep Convolutional Neural Networks: Applied to Agricultural Robotics. IEEE Robot Autom Lett 2017;2:1344-51. https://doi.org/10.1109/LRA.2017.2667039.

15. Pusdá-Chulde MR, Salazar-Fierro FA, Sandoval-Pillajo L, Herrera-Granda EP, García-Santillán ID, De Giusti A. Image Analysis Based on Heterogeneous Architectures for Precision Agriculture: A Systematic Literature Review. En: Nummenmaa J, Pérez-González F, Domenech-Lega B, Vaunat J, Oscar Fernández-Peña F, editores. Adv. Appl. Comput. Sci. Electron. Ind. Eng., vol. 1078, Cham: Springer International Publishing; 2020, p. 51-70. https://doi.org/10.1007/978-3-030-33614-1_4.

16. García-Santillán ID, Pajares G. On-line crop/weed discrimination through the Mahalanobis distance from images in maize fields. Biosyst Eng 2018;166:28-43. https://doi.org/10.1016/j.biosystemseng.2017.11.003.

17. Zou K, Chen X, Zhang F, Zhou H, Zhang C. A Field Weed Density Evaluation Method Based on UAV Imaging and Modified U-Net. Remote Sens 2021;13:310. https://doi.org/10.3390/rs13020310.

18. Osorio Delgado AK. Método para la estimación de maleza en cultivos de lechuga utilizando aprendizaje profundo e imágenes multiespectrales. Trabajo de grado - Maestría. Universidad Nacional de Colombia, 2021.

19. Puerto Lara AE. Clasificación y cuantificación de maleza en cultivos de hortalizas por medio de procesamiento de imágenes digitales multiespectrales. Universidad Nacional de Colombia, 2018.

20. Xie S, Girshick R, Dollar P, Tu Z, He K. Aggregated Residual Transformations for Deep Neural Networks. 2017 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, Honolulu, HI: IEEE; 2017, p. 5987-95. https://doi.org/10.1109/CVPR.2017.634.

21. Fayyad U, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knowledge Discovery in Databases. AI Mag 1996;17:37-37. https://doi.org/10.1609/aimag.v17i3.1230.

22. Cai Y, Zeng F, Xiao J, Ai W, Kang G, Lin Y, et al. Attention-aided semantic segmentation network for weed identification in pineapple field. Comput Electron Agric 2023;210:107881. https://doi.org/10.1016/j.compag.2023.107881.

23. Nong C, Fan X, Wang J. Semi-supervised Learning for Weed and Crop Segmentation Using UAV Imagery. Front Plant Sci 2022;13:927368. https://doi.org/10.3389/fpls.2022.927368.

24. Shahi TB, Dahal S, Sitaula C, Neupane A, Guo W. Deep Learning-Based Weed Detection Using UAV Images: A Comparative Study. Drones 2023;7:624. https://doi.org/10.3390/drones7100624.

25. Sarvini T, Sneha T, Sukanya Gowthami G, Sushmitha S, Kumaraswamy R. Performance Comparison of Weed Detection Algorithms. 2019 Int. Conf. Commun. Signal Process. ICCSP, Chennai, India: IEEE; 2019, p. 0843-7. https://doi.org/10.1109/ICCSP.2019.8698094.

26. Roboflow Docs. Create a Project | Roboflow Docs 2024. https://docs.roboflow.com/datasets/create-a-project (accedido 28 de enero de 2024).

27. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, Las Vegas, NV, USA: IEEE; 2016, p. 770-8. https://doi.org/10.1109/CVPR.2016.90.

28. Mou L, Zhu XX. Vehicle Instance Segmentation From Aerial Image and Video Using a Multitask Learning Residual Fully Convolutional Network. IEEE Trans Geosci Remote Sens 2018;56:6699-711. https://doi.org/10.1109/TGRS.2018.2841808.

29. Li X, Duan F, Hu M, Hua J, Du X. Weed Density Detection Method Based on a High Weed Pressure Dataset and Improved PSP Net. IEEE Access 2023;11:98244-55. https://doi.org/10.1109/ACCESS.2023.3312191.

30. Hu X-Z, Jeon W-S, Rhee S-Y. Sugar Beets and Weed Detection using Semantic Segmentation. 2022 Int. Conf. Fuzzy Theory Its Appl. IFUZZY, Kaohsiung, Taiwan: IEEE; 2022, p. 1-4. https://doi.org/10.1109/iFUZZY55320.2022.9985222.

31. Lottes P, Behley J, Milioto A, Stachniss C. Fully Convolutional Networks With Sequential Information for Robust Crop and Weed Detection in Precision Farming. IEEE Robot Autom Lett 2018;3:2870-7. https://doi.org/10.1109/LRA.2018.2846289.

32. Weyler J, Läbe T, Magistri F, Behley J, Stachniss C. Towards Domain Generalization in Crop and Weed Segmentation for Precision Farming Robots. IEEE Robot Autom Lett 2023;8:3310-7. https://doi.org/10.1109/LRA.2023.3262417.

33. Sa I, Chen Z, Popovic M, Khanna R, Liebisch F, Nieto J, et al. weedNet: Dense Semantic Weed Classification Using Multispectral Images and MAV for Smart Farming. IEEE Robot Autom Lett 2018;3:588-95. https://doi.org/10.1109/LRA.2017.2774979.

34. Gonçalves ÉC, Almeida GPD, Silva ELD, Schein TT, Evald PJDDO, Drews-Jr PLJ. Line Detection and Segmentation of Annual Crops Using Hybrid Method. 2023 Lat. Am. Robot. Symp. LARS 2023 Braz. Symp. Robot. SBR 2023 Workshop Robot. Educ. WRE, Salvador, Brazil: IEEE; 2023, p. 472-7. https://doi.org/10.1109/LARS/SBR/WRE59448.2023.10332920.

35. Ullah HS, Asad MH, Bais A. End to End Segmentation of Canola Field Images Using Dilated U-Net. IEEE Access 2021;9:59741-53. https://doi.org/10.1109/ACCESS.2021.3073715.

36. Sossa J, Rodríguez R. Procesamiento y Análisis Digital de Imágenes. 1.a ed. 430: RA-MA; 2011.

37. Misra D. Mish: A Self Regularized Non-Monotonic Activation Function 2019. https://doi.org/10.48550/ARXIV.1908.08681.

38. Smith LN. Cyclical Learning Rates for Training Neural Networks 2015. https://doi.org/10.48550/ARXIV.1506.01186.

39. Loshchilov I, Hutter F. SGDR: Stochastic Gradient Descent with Warm Restarts 2017.

40. Celikkan E, Saberioon M, Herold M, Klein N. Semantic Segmentation of Crops and Weeds with Probabilistic Modeling and Uncertainty Quantification. 2023 IEEECVF Int. Conf. Comput. Vis. Workshop ICCVW, 2023, p. 582-92. https://doi.org/10.1109/ICCVW60793.2023.00065.

41. Juma A, Rodríguez J, Caraguay J, Naranjo M, Quiña-Mera A, García-Santillán I. Integration and Evaluation of Social Networks in Virtual Learning Environments: A Case Study. En: Botto-Tobar M, Pizarro G, Zúñiga-Prieto M, D'Armas M, Zúñiga Sánchez M, editores. Technol. Trends, vol. 895, Cham: Springer International Publishing; 2019, p. 245-58. https://doi.org/10.1007/978-3-030-05532-5_18.

42. Lind D, Marchal W, Wathen S. Basic Statistics in Business and Economics. 10.a ed. MCGraw Hill; 2022.

43. Sapkota R, Stenger J, Ostlie M, Flores P. Towards reducing chemical usage for weed control in agriculture using UAS imagery analysis and computer vision techniques. Sci Rep 2023;13:6548. https://doi.org/10.1038/s41598-023-33042-0.

FINANCING

No financing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Data curation: Kevin Vinueza, Lucía Sandoval-Pillajo, Adriana Giret-Boggino, Diego Trejo-España, Marco Pusdá-Chulde, Iván García-Santillán.

Methodology: Kevin Vinueza, Lucía Sandoval-Pillajo, Adriana Giret-Boggino, Diego Trejo-España, Marco Pusdá-Chulde, Iván García-Santillán.

Software: Kevin Vinueza, Lucía Sandoval-Pillajo, Adriana Giret-Boggino, Diego Trejo-España, Marco Pusdá-Chulde, Iván García-Santillán.

Drafting - original draft: Kevin Vinueza, Lucía Sandoval-Pillajo, Adriana Giret-Boggino, Diego Trejo-España, Marco Pusdá-Chulde, Iván García-Santillán.

Writing - proofreading and editing: Kevin Vinueza, Lucía Sandoval-Pillajo, Adriana Giret-Boggino, Diego Trejo-España, Marco Pusdá-Chulde, Iván García-Santillán.