BRIEF COMMUNICATION

LLS-SevEst - Late leaf spot severity estimator. A machine learning approach to assessing *Nothopassalora personata* in peanut

August, 2025

Herrador, T.A.[1]; Migotti Scaglia, J.[1]; Paredes J.A.[2,3]; Cazón, L.I.[2,3]

ABSTRACT

Late leaf spot (LLS), caused by *Nothopassalora personata*, is the most damaging foliar disease in peanut production worldwide. Accurate disease severity assessment is crucial for evaluating and implementing effective management strategies. This study aimed to develop and validate an automated image analysis model, *LLS-SevEst*, for quantifying LLS severity in peanut leaves. A dataset of 190 scanned leaf images was analyzed using three approaches: a fixed threshold-based segmentation, morphological preprocessing and K-means clustering. Exploratory analyses revealed distinct brightness patterns between healthy and diseased tissues, guiding the development of classification functions. The threshold-based model yielded high false positive rates due to its inability to account for natural leaf variation, while the morphological preprocessing method improved segmentation marginally but still required manual adjustments. The K-means clustering approach provided relatively better segmentation performance under the specific conditions tested and showed high potential for automated and reproducible disease severity estimation. This work should be considered a proof-of-concept, and further research is required to develop a robust and generalizable tool for LLS severity estimation.

Keywords: peanut diseases, late leaf spot, image analysis, machine learning, disease quantification.

RESUMEN

La viruela del maní, causada por Nothopassalora personata, es la enfermedad foliar más importante de este cultivo a nivel mundial. La evaluación precisa de la severidad de la enfermedad en la planta es fundamental para la implementación de estrategias de manejo efectivas. Por lo tanto, el objetivo de este estudio fue desarrollar y validar un modelo automatizado de análisis de imágenes, denominado LLS-SevEst, para cuantificar la severidad de la viruela el maní en hojas. Para esto se analizó un conjunto de 190 imágenes escaneadas de hojas de maní utilizando tres enfoques: segmentación basada en umbrales fijos, preprocesamiento morfológico y agrupación de clústeres por K-means. Los análisis exploratorios revelaron patrones de brillo distintos entre los tejidos sanos y enfermos, lo que permitió orientar el desarrollo de funciones de clasificación. El modelo basado en umbrales presentó altas tasas de falsos positivos debido a su incapacidad para considerar la variación natural en la tonalidad de las hojas, mientras que el preprocesamiento morfológico mejoró la segmentación, aunque evidenciando la necesidad de ajustes manuales. El enfoque basado en agrupamientos por K-means ofreció un mejor desempeño relativo para las condiciones evaluadas, mostrando un alto potencial para una estimación automatizada y reproducible de la severidad de la enfermedad. Debido a la naturaleza de nuestros resultados, este trabajo debe considerarse una prueba de concepto, que requiere investigaciones adicionales para constituir una herramienta robusta para la estimación de la severidad de LLS.

Palabras clave: *enfermedades del maní, viruela del maní, análisis de imágenes, aprendizaje automático, cuantificación de enfermedades.*

---

[1] Instituto Universitario de Ciencias Biomédicas de Córdoba. Avenida Naciones Unidas 420 (5012), Córdoba, Argentina.

[2] Instituto Nacional de Tecnología Agropecuaria (INTA), Instituto de Patología Vegetal (IPAVE), Centro de Investigaciones Agropecuarias (CIAP). Avenida 11 de Septiembre 4755 (X5020ICA), Córdoba, Argentina. Correo electrónico: cazon.ignacio@inta.gob.ar

[3] Unidad de Fitopatología y Modelización Agrícola, (UFyMA-CONICET). Avenida 11 de Septiembre 4755 (X5020ICA), Córdoba, Argentina.

https://doi.org/10.58149/2xz3-6879

INTRODUCTION

Peanut is a key global crop. In Argentina, production is regionally important, with over 70% concentrated in Córdoba. During the 2023/24 season, 300.000 hectares were planted, generating over $1 billion in exports. However, peanut production faces significant phytosanitary challenges, with late leaf spot (LLS), caused by *Nothopassalora personata* (Berk. & M.A. Curtis), being the most damaging disease worldwide (Giordano *et al*., 2021). Under favorable conditions (95% relative humidity, ~18°C), and without effective field management, LLS can produce losses exceeding 50% (Shokes and Culbreath, 1997). The symptoms consist of dark leaf spots surrounded by yellow halos, which, in severe cases, coalesce and drastically reduce the photosynthetic area (Marinelli and March 2005; Oddino *et al*., 2018).

Control relies primarily on fungicides (e.g., strobilurins, triazoles, carboxamides, and chlorothalonil) (Giordano *et al*., 2021; Monguillot *et al*., 2023). Evaluating the efficacy of these tools requires accurate disease severity assessment, which remains predominantly visual despite the availability of advanced techniques. While visual assessment is cost-effective, it is inherently subjective and influenced by the evaluator's expertise and pathosystem characteristics (Bock *et al*., 2020; Del Ponte *et al*., 2021).

Various tools have been developed to improve visual assessments, including online training systems and a standard area diagram (Cazón *et al*., 2025; Del Ponte, 2023). Smart agriculture technologies, particularly multispectral and hyperspectral imaging, offer promising alternatives for disease severity quantification assessment (Chen *et al*., 2019; Omran, 2016). However, adoption remains limited due to operational complexity and costs. In contrast, RGB imaging with deep learning has shown high accuracy in peanut foliar disease detection, yet no validated tool currently exists for LLS severity quantification (Xu *et al*., 2023).

In this context, this study aimed to develop a proof-of-concept model using Python software, based on image segmentation methods. With this, we seek to lay the groundwork for the future development of automated models for quantifying the severity of LLS.

MATERIAL AND METHODS

- **Image acquisition.** A total of 190 peanut leaves with varying disease severity were collected in April 2023 from plants grown under controlled conditions at IPAVE-CIAP, Córdoba (Latitude: −31.46895, Longitude: −64.14730). The abaxial leaf surface was scanned using a CanoScan LIDE 300 flatbed scanner at 300 dpi. A group of 50 representative images were selected for model development. Additionally, the *Pliman* package (Olivoto, 2022) in *R* (R Core Team, 2022) was used for comparison, applying segmentation techniques to distinguish healthy from diseased tissue.
- **Image processing.** Image processing was conducted using *Python 3.x* (Python Software Foundation, 2023) in *Jupyter Notebooks* (Kluyver *et al*., 2016) on *Google Colab*. Exploratory analyses identified pixel luminosity differences between healthy and diseased areas. To enhance contrast and visualize brightness distribution, various filters and histogram plots were applied using pandas (McKinney, 2023), NumPy (Numpy, 2023), OpenCV (Itseez, 2023), and Matplotlib (Hunter, 2023). Based on these patterns, three methods were used to estimate the percentage of leaf area affected by *N. personata*:

1. Threshold-Based Model: Pixels with intensity <80 (from histogram analysis) were classified as lesions; others as healthy. Severity was calculated as the proportion of lesion pixels relative to the total leaf area.
2. Morphological Preprocessing Model: Erosion followed by dilation (3×3 elliptical element) improved segmentation, reducing misclassification. Users could manually adjust thresholds based on histograms for better accuracy.
3. K-Means Clustering Model: Using scikit-learn (Pedregosa *et al*., 2011), images were transformed into RGB matrices, smoothed and converted into datasets of pixel positions and color values. Color-difference features were added and normalized (StandardScaler). K-means clustering (3 clusters, 10 iterations) classified pixels; severity was calculated as lesion pixels over the total leaf area.

RESULTS AND DISCUSSION

Exploratory Analysis

The grayscale conversion effectively distinguished healthy and diseased areas based on pixel luminosity (fig. 1A). The histogram analysis revealed a distinct intensity peak corresponding to the background (255 intensity units). In contrast, leaf areas (healthy + diseased) displayed a Gaussian distribution (fig. 1B).

Further analysis showed different intensity patterns between healthy and diseased regions (fig. 2). Healthy tissue exhibited a Gaussian distribution (fig. 2B), while lesions displayed bimodal distributions due to overlapping brightness intensities at lesion margins (fig. 2A, C).

Model performance evaluation

For the Threshold-Based Model, a threshold of 80 grayscale intensity units (ranging from 0 for black to 255 for white) was set for the classification of different areas. Pixels below this threshold were classified as lesions, while those above were considered healthy leaf tissue. However, this approach failed to accurately differentiate between healthy and diseased areas, resulting in a high rate of false positives. One example of this misclassification was the identification of shadows cast by the leaf's midrib as diseased areas (fig. 3A). This limitation led to considerable variability when comparing the severities obtained with this model and those calculated using *Pliman* (fig. 3B). Although this approach is conceptually valid (Barbedo, 2016), the results suggest that segmentation based on a fixed luminosity threshold is insufficient for accurately distinguishing between healthy and lesioned areas, particularly in leaves with natural variations in brightness and color in this pathosystem. The function was later modified to allow for manual threshold adjustment, enabling better adaptation to the specific characteristics of each leaf image.

Regarding the Morphological Preprocessing Model, the dilation followed by erosion function helped to "smooth" the images, reducing some false positives (Gonzalez and Woods, 2018). However, darker-toned areas remained undistinguished (fig. 3C), still causing significant dispersion in the severity estimates when compared to *Pliman* (fig. 3D). These results suggest that a fixed threshold cannot be universally applied, and that manual adjustments are required for each specific case, which reduces the practicality and automation of the method. It is important to note that achieving accurate segmentation remains a significant challenge in image-based automatic plant disease identification (Barbedo, 2016).

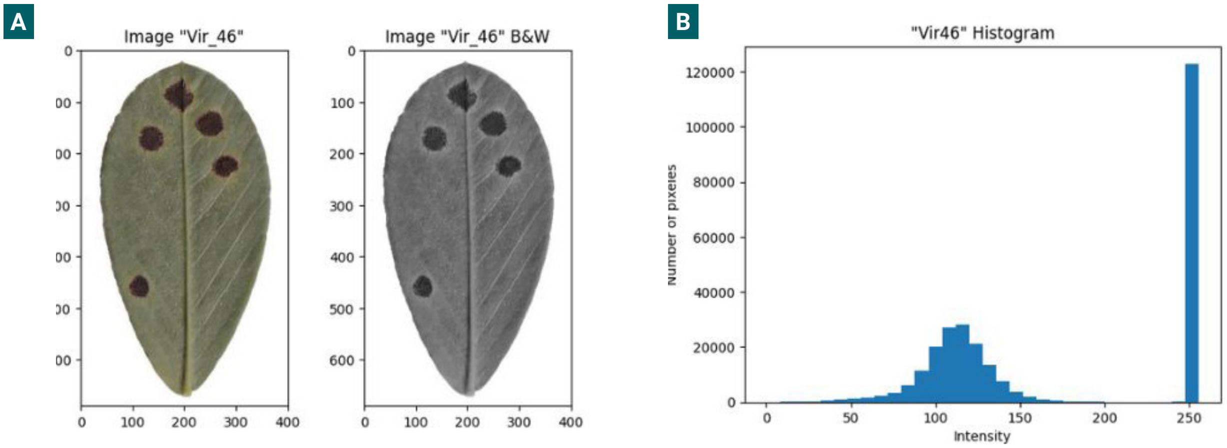

Figure 1. A. RGB image of a peanut leaflet affected by *N. personata* (left), and its grayscale conversion (right) using the color_rgb2gray from the OpenCV library (cv2). **B.** Pixel intensity histogram corresponding to the grayscale image in A. Peack near to 255 correspond to the background. Gaussian distribution between 60–150 correspond with healthy and diseased tissues.

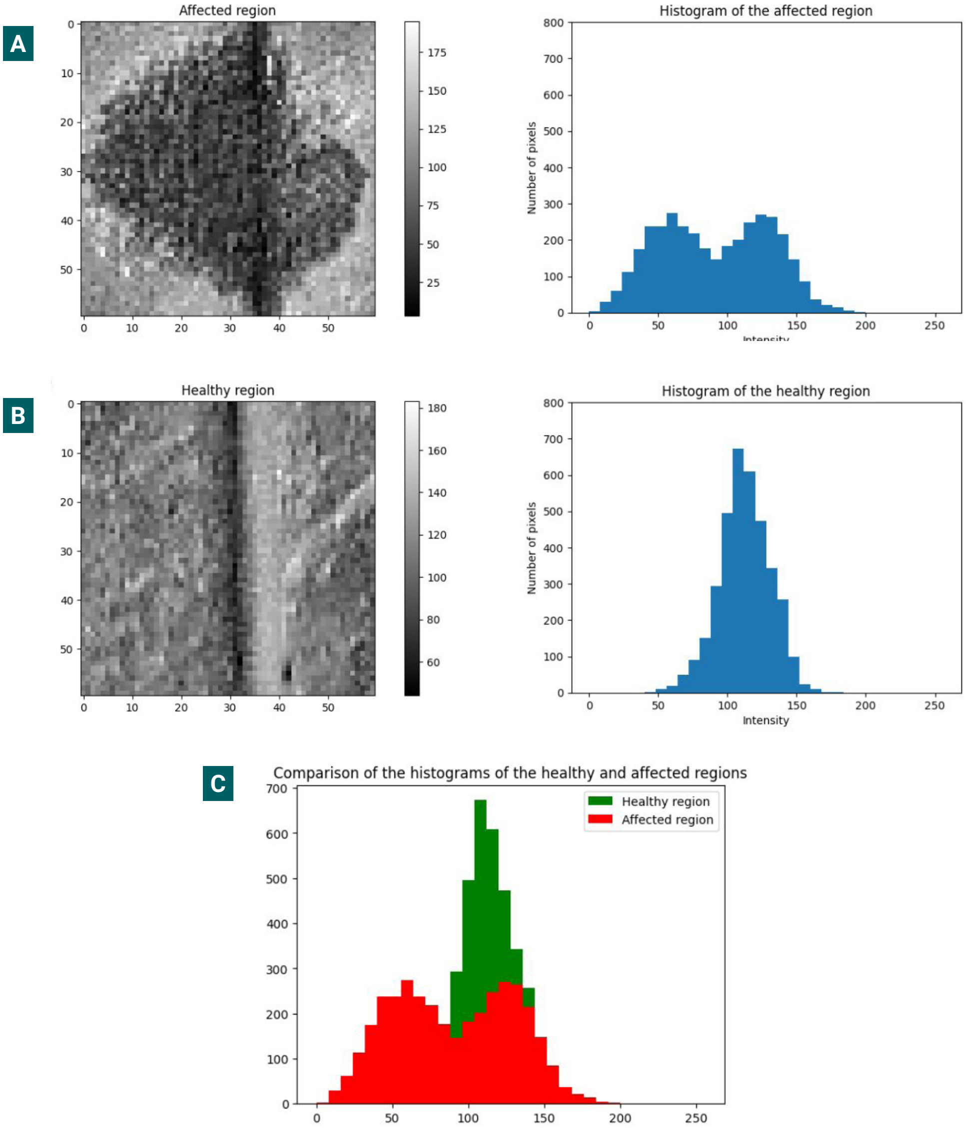

Figure 2. A: Cropped grayscale image of a lesion caused by *N. personata* and its pixel luminance histogram. **B**: Cropped grayscale image of a healthy leaf area and its pixel luminance histogram. **C**: Overlaid histograms showing lesion pixels in red and healthy pixels in green.

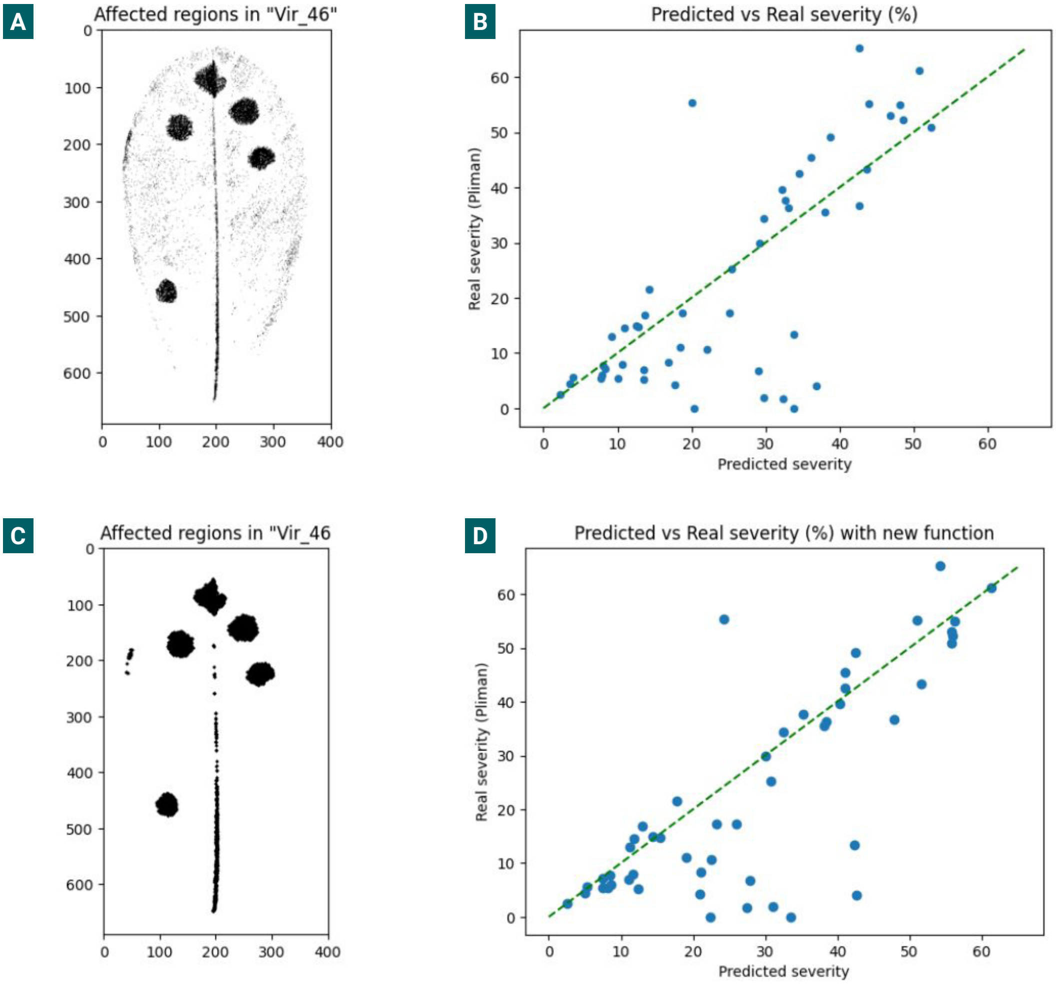

Figure 3. Comparison between two versions of the model for estimating LLS severity. **A.** Lesion segmentation from the first model, based on pixel luminance thresholding. **B.** Relationship between severity (%) predicted by the first model and that calculated with *Pliman*; the green line represents Deming regression. **C.** Lesion segmentation using a modified model incorporating morphological operations (erosion and dilation). **D.** Relationship between severity (%) predicted by the modified model and that calculated with *Pliman*; the green line shows the Deming regression.

Applying the K-means model enabled effective image segmentation, accurately distinguishing healthy tissue, affected areas, and background regions. This approach was used by Phadikar *et al*. (2012) for rice disease classification. The cluster visualization revealed a clear distinction between *N. personata*-affected areas and healthy tissues, allowing an objective assessment of disease severity (fig. 4A). The model was applied to all images in the dataset, enabling the calculation of the affected area percentage in each case.

At first glance, a significant improvement is observed compared to the results obtained with the initial functions. However, discrepancies remain between the previously calculated severity and the severity estimated using the developed model (fig. 4B). When images with considerable discrepancies between *Pliman* and the *LLS-SevEst* model were closely analyzed, it was observed that some photosynthetic regions were not classified as lesions by *Pliman*. Since many authors emphasize the high efficiency of *Pliman* in determining disease severity, any methodological error, likely introduced by the user during the creation of palettes in the initial image processing stages, can lead to overall classification errors (Del Ponte, 2023).

When images exhibiting this discrepancy were excluded, the model fit improved significantly, suggesting that K-means is an efficient tool for image segmentation in foliar disease quantification, enabling automated evaluation of LLS severity (fig. 5).

Future research should focus on improving the model by integrating deep learning techniques, such as convolutional neural networks, and expanding the dataset to include more diverse leaf images from different genotypes and environmental conditions (Ferentinos, 2018; Mohanty *et al*., 2016). In addition, allowing threshold flexibility and manual adjustment remains crucial to ensure accuracy across a variety of scenarios. Although the primary objective of LLS-SevEst is to support research and development activities, particularly fungicide efficacy trials and resistance evaluations, it should be emphasized that this is a preliminary model. The absence of a large, diverse dataset and reliance on unsupervised methods limit its generalizability. Furthermore, no formal statistical validation (e.g., accuracy, precision, recall) was conducted. Future versions must incorporate more advanced AI approaches, such as convolutional neural networks, along with rigorous validation metrics. Thus, the current version of LLS-SevEst represents an early-stage tool with potential for development rather than a definitive or benchmark model for disease severity estimation.

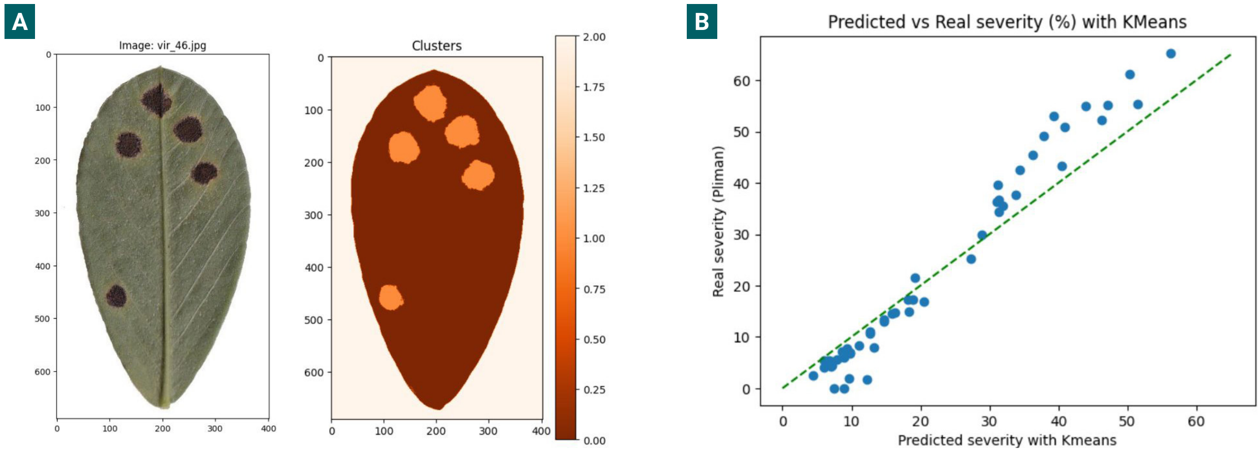

Figure 4. Leaf lesion segmentation using the K-Means clustering algorithm. **A.** Original image (left) and segmentation by K-Means (right), where white corresponds to the "background" cluster, red to the "healthy" cluster, and orange to the "lesion" cluster. **B.** Comparative scatterplot between severity (%) estimated by the K-Means model and that calculated with *Pliman*. The green line represents the Deming regression.

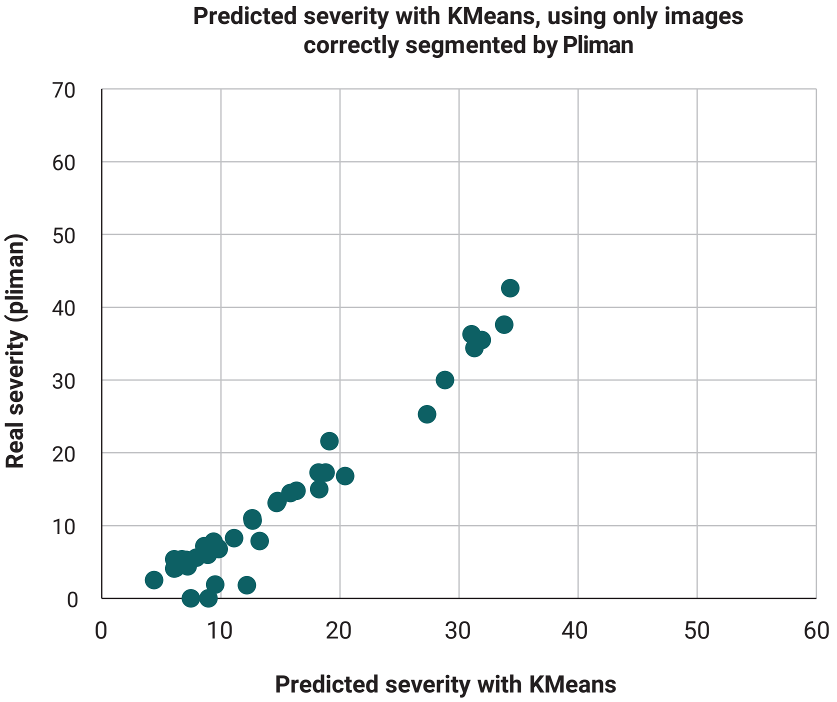

Figure 5. Comparative scatter plot between the severity percentage predicted by KMeans and the one previously calculated using *Pliman*. Only the images correctly segmented by *Pliman* were included in this analysis.

ACKNOWLEDGEMENTS

We want to thank INTA, Fundación Maní Argentino, and IUCBC.

DATA AVAILABILITY

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

CONFLICTS OF INTEREST

All authors declare that they have no conflicts of interest.

REFERENCES

BARBEDO, J.G.A. 2016. A review on the main challenges in automatic plant disease identification based on visible range images. Biosystems Engineering, 144, 52-60. https://doi.org/10.1016/j.biosystemseng.2016.01.017xz

BOCK, C.H.; BARBEDO, J.G.A.; DEL PONTE, E.M.; BOHNENKAMP, D.; MAHLEIN, A.-K. 2020. From visual estimates to fully automated sensor-based measurements of plant disease severity: Status and challenges for improving accuracy. Phytopathology Research, 2, 9.

CAZÓN, L.I.; PAREDES, J.A.; GONZÁLEZ, N.R.; CONFORTO, E.C.; SUAREZ, L.; DEL PONTE, E.M. 2025. Optimizing visual estimation of peanut late leaf spot severity with online training sessions and standard area diagrams. Eur J Plant Pathol Vol. 172. 451-465 pp. https://doi.org/10.1007/s10658-025-03016-1

CHEN, T.; ZHANG, J.; CHEN, Y.; WAN, S.; ZHANG, L. 2019. Detection of peanut leaf spots disease using canopy hyperspectral reflectance. Computers and Electronics in Agriculture, 156, 677-683.

DEL PONTE, E.M.; CAZÓN, L.I.; ALVES, K.; PETHYBRIDGE, S.; BOCK, C. 2021. How much do standard area diagrams improve accuracy of visual estimates of plant disease severity? A systematic review and meta-analysis. Tropical Plant Pathology. Avance online publication. https://doi.org/10.1007/s40858-021-00479-5

DEL PONTE, E.M. 2023. Training sessions. In R for plant disease epidemiology (R4PDE). (Available at: https://r4pde.net verified on March 26, 2025).

FERENTINOS, K.P. 2018. Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311-318.

GIORDANO, D.F.; PASTOR, N.; PALACIOS, S.; ODDINO, C.M.; TORRES, A.M. 2021. Peanut leaf spot caused by *Nothopassalora personata*. Tropical Plant Pathology, 46, 139151.

GONZALEZ, R.C.; WOODS, R.E. 2018. Digital Image Processing (4th ed.). Pearson.

HUNTER, J.D. 2023. Matplotlib [Software]. (Available at: https://matplotlib.org/ verified on March 26, 2025).

ITSEEZ. 2023. OpenCV [Software]. (Available at: https://opencv.org/ verified on March 26, 2025).

KLUYVER, T.; RAGAN-KELLEY, B.; PÉREZ, F.; GRANGER, B. E.; BUSSONNIER, M.; FREDERIC, J.; WILLING, C. 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows. In: Loizides, F.; Schmidt, B. (Eds.). Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press. 87-90. https://doi.org/10.3233/978-1-61499-649-1-87

MARINELLI, A.; MARCH, G.J. 2005. Viruela. In: Marinelli, A.; March, G.J. (Eds.). Enfermedades del maní en Argentina. Ediciones Biglia. 13-39 pp.

MCKINNEY, W. 2023. Pandas [Software]. (Available at: https://pandas.pydata.org/ verified on March 26, 2025).

MOHANTY, S.P.; HUGHES, D.P.; SALATHÉ, M. 2016. Using deep learning for image-based plant disease detection. Front. Plant Sci. 7:1419. doi: 10.3389/fpls.2016.01419

MONGUILLOT, J.H.; BERNARDI LIMA, N.; PAREDES, J.A.; GIORDANO, D.F.; ODDINO, C.; RAGO, A.M.; CARMONA, M.; CONFORTO, E.C. 2023. Caracterización de aislados de *Nothopassalora personata* agente causal de la viruela tardía del maní. 38° Jornada Nacional de Maní.

NUMPY DEVELOPMENT TEAM. 2023. NumPy [Computer software]. (Available at: https://numpy.org verified on March 26, 2025).

ODDINO, C.; GIORDANO, F.; PAREDES, J.; CAZÓN, L.; GIUGGIA, J.; RAGO, A. 2018. Efecto de nuevos fungicidas en el control de viruela del maní y el rendimiento del cultivo. Ab Intus, 1, 9-17.

OLIVOTO, T. 2022. Lights, camera, *Pliman*! An R package for plant image analysis. Methods in Ecology and Evolution, 13(4), 789-798.

OMRA, E.S.E. 2016. Early sensing of peanut leaf spot using spectroscopy and thermal imaging. Archives of Agronomy and Soil Science, 63, 883-896.

PEDELINI, R. 2021. MANÍ: Guía práctica para su cultivo. FMA press.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; *et al*. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

PHADIKAR, S.; SIL, J.; NAYAK, J. 2012. Rice diseases classification using feature selection and rule generation techniques. Computers and Electronics in Agriculture, 90, 76-85.

PYTHON SOFTWARE FOUNDATION. 2023. Python Language Reference (Version 3.x). (Available at: https://www.python.org verified on March 26, 2025).

R CORE TEAM. 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. (Available at: https://www.R-project.org/ verified on February 10, 2024).

SHOKES, F.M.; CULBREATH, A.K. 1997. Early and late leaf spots. In: KOKALIS-BURELLE N.; PORTER, D.M.; RODRÍGUEZ-KÁBANA, R.; SMITH D.H.; SUBRAHMANYAM, P. (Eds.). Compendium of peanut diseases, 2nd ed. APS Press. 17-20 pp.

XU, L.; CAO, B.; NING, S.; ZHANG, W.; ZHAO, F. 2023. Peanut leaf disease identification with deep learning algorithms. Mol Breeding 43, 25. https://doi.org/10.1007/s11032-023-01370-8