

82.02- Proyecto Final Analítica

Entrega 5

Conteo de cultivos de maíz (Stand Count)

Grupo 5

Integrantes del grupo:

- Azul de los angeles Makk - 61589
- Sofía González del Solar - 62292
- Nicole Reiman - 62407

Fecha de entrega: 10/06/2024



1er cuatrimestre – 2024

Índice

Resumen ejecutivo	3
Sección I	4
1. Introducción	4
2. Definición del problema	4
3. Estado del arte	5
4. Herramientas y Metodología	7
5. Entregables y Outputs del Proyecto	8
6. Plan de Trabajo	8
7. Caso de negocio	10
a. KPIs de Negocio	10
b. Enfoques primarios para evaluar el desempeño del modelo	11
c. Enfoques secundarios para evaluar el desempeño del modelo	12
d. Costos y Escenarios	13
e. Posibles Escenarios Resultantes de la Aplicación de Modelos de Conteo	15
Sección II	17
8. Introducción	17
9. Hipótesis de trabajo	18
10. Datos disponibles	18
a. Imágenes	18
b. Etiquetas (labels) con metadatos	18
c. Archivo JSON con metadatos	19
11. Adquisición de datos	20
12. Limpieza y detección de anomalías	21
a. Muestreo aleatorio	21
b. Búsqueda de anomalías	21
i. Evaluación de Irregularidades en la Cantidad de Bounding Boxes	22
ii. Etiquetado erróneo	23
iii. Overlapped Bounding Boxes	24
iv. Ajustes de bounding boxes	25
c. Resumen y Conclusiones del la sección	25
13. Características generales	26
a. Tamaño de imagen	26
b. Hora de vuelo	28
c. Ubicación del campo	29
d. Fecha de vuelo	30
e. Análisis de etiquetas	31
f. Resumen y Conclusiones de la sección	32
14. Features	33
a. Resumen y Conclusiones de la sección	36
15. Bounding Boxes	36
a. Cantidad	36
b. Distribución	37
16. Identificación de clusters	38

17. Conclusiones de la sección II	39
a. Calidad de los datos	39
b. Tamaño de las imágenes	39
c. Features de las imágenes	39
Sección III	41
18. Introducción	41
19. YOLO (You Only Look Once)	41
a. ConvNets y características	41
b. Entrenamiento en YOLO	41
i. Predicción de cajas delimitadoras	41
ii. Filtrado por umbral y supresión no máxima	42
c. Ventajas de YOLO y Conclusión de la sección	42
20. Elección de versión de YOLO.	43
21. Preparación del dataset	43
22. División Train, Validation y Test	44
a. Conjuntos de datos en YOLOv5	44
b. División Equitativa del Dataset	47
c. Conclusiones de la sección	49
23. Exploración de modelos	49
a. Disminución del dataset	49
b. Hiperparámetros en YOLOv5	50
i. Hiperparámetros en el Comando de Entrenamiento	50
ii. Hiperparámetros Internos del Modelo	51
iii. Hiperparámetros de Detección	51
c. Implementación de los modelos	52
i. Variación de hiperparámetros del comando de entrenamiento	53
ii. Variación de Hiperparámetros Internos del Modelo	53
iii. Variación de Hiperparámetros de Detección	56
iv. Conclusiones de la sección	58
24. Implementación del modelo	59
a. Variaciones en etiquetado manual	59
b. Análisis de resultados en imágenes	60
25. Experimentación de resultados	60
a. Resultados Preliminares	60
b. Resultados por segmento	62
c. Conclusiones de la sección	64
26. Data Augmentation	65
27. Posprocesamiento	66
28. Implementación	67
29. Impacto de los resultados en el business case	68
30. Conclusion	70
Feedback del cliente	71
Líneas de Investigación Futura	71
Bibliografía	73

Resumen ejecutivo

El proyecto tiene como objetivo el desarrollo de una solución automatizada para el conteo de plantas de maíz en campos agrícolas utilizando algoritmos de Machine Learning. Esta solución busca mejorar la precisión, eficiencia y escalabilidad del proceso de conteo de plantas, en comparación con los métodos manuales tradicionales.

En la agricultura, el conteo preciso de plantas es crucial para la optimización de recursos y la gestión de cultivos. El método manual presenta varios desafíos, como la subjetividad en la percepción humana, el tiempo y esfuerzo requeridos, y las limitaciones en la extensión de los conteos manuales. Estos factores pueden llevar a estimaciones inexactas y decisiones subóptimas en la gestión de los cultivos.

Se desarrolló un algoritmo de Machine Learning utilizando la arquitectura de Yolo, en colaboración con la empresa Eiwa, para contar automáticamente las plantas de maíz mediante el análisis de imágenes tomadas por drones. Esta solución pretende proporcionar un método más eficiente y preciso para el conteo de plantas, optimizando así la toma de decisiones agrícolas y reduciendo los costos operativos.

El proyecto se desarrolló en varias etapas, incluyendo la integración y limpieza de datos proporcionados por Eiwa, asegurando la corrección de cualquier inconsistencia; la exploración profunda del conjunto de datos para identificar patrones y posibles irregularidades, utilizando técnicas descriptivas y visuales; la selección y ajuste de algoritmos de Machine Learning, entrenamiento de modelos y evaluación de su desempeño; la evaluación de la eficacia y eficiencia de los modelos implementados, utilizando métricas desarrolladas específicamente para este propósito; y la preparación de una presentación profesional adaptada a la audiencia específica.

El proyecto culmina en una herramienta que procesa imágenes para contar la cantidad de plantas de maíz presentes en ellas. Esta herramienta permitirá a Eiwa ofrecer a sus clientes un análisis más eficiente de los ensayos agrícolas, facilitando la evaluación de métodos y condiciones agrícolas más efectivas. El modelo entrenado alcanzó una precisión de 0,8414 y un recall de 0,8301 en el conjunto de prueba. Este desempeño resultó en un aumento del 16% en las ganancias de Eiwa en comparación con los métodos manuales y un ahorro del 73% en las horas dedicadas al conteo, manteniendo un alto nivel de exactitud.

El desarrollo de esta herramienta no solo optimiza la gestión de cultivos, sino que también proporciona una base para futuras investigaciones y mejoras en el campo de la agricultura. Además, la colaboración con Eiwa garantizará la relevancia y aplicabilidad de la solución en entornos agrícolas reales, adaptándose a las necesidades específicas del mercado.

Sección I

Planteamiento, desarrollo y propuesta de la solución del problema

1. Introducción

En el ámbito de la agricultura, el conteo preciso de plantas es una práctica fundamental que incide directamente en la optimización de los recursos y el éxito de los cultivos. Este proceso, cuando se realiza de manera manual, enfrenta una serie de desafíos que afectan su precisión, eficiencia y escalabilidad. Factores como la subjetividad inherente a la percepción humana, la laboriosidad del proceso y las limitaciones en la extensión del conteo manual en campos agrícolas extensos, hacen imperativo buscar soluciones alternativas que mejoren la gestión general de los cultivos.

El presente proyecto surge en respuesta a estas necesidades, proponiendo el desarrollo de un algoritmo de Machine Learning capaz de contar automáticamente las plantas de maíz en campos agrícolas mediante el análisis de imágenes tomadas por drones. En colaboración con la empresa Eiwa, especializada en el análisis de datos agropecuarios, se propone abordar este desafío con el objetivo de proporcionar una solución innovadora y eficiente que simplifique el proceso de conteo, a la vez que optimice la toma de decisiones agrícolas informadas y reduzca los costos operativos asociados al conteo manual. Mediante esta asociación, se permite trabajar directamente con clientes reales, adaptando la solución a las necesidades específicas del mercado y garantizando su relevancia y aplicabilidad en entornos agrícolas reales.

2. Definición del problema

El conteo de plantas es una práctica frecuente en la agricultura y se utiliza como una técnica de optimización para mejorar las técnicas de cultivo y la gestión de recursos. Esto incluye el riego, la distancia entre cultivos, el uso de fertilizantes y pesticidas, y su impacto en el resultado final. Sin embargo, los agricultores enfrentan diversos desafíos en el conteo de plantas en sus terrenos, lo cual es crucial para el éxito de sus cultivos. Tanto si se realiza recorriendo los campos como mediante la visualización de fotografías, el conteo de plantas presenta dificultades significativas que afectan la precisión, la eficiencia y la gestión general de los cultivos.

En primer lugar, el conteo manual¹ a través de imágenes, es propenso a errores debido a la subjetividad inherente a la percepción humana. Las condiciones de iluminación, la fatiga visual y otros factores pueden influir en la capacidad de una persona para contar con precisión el número de plantas presentes en una parcela. Esto puede resultar en estimaciones inexactas que no reflejan adecuadamente la densidad real de plantas, lo que a su vez puede llevar a decisiones de gestión ineficientes.

Además, este proceso resulta laborioso y consumidor de tiempo, especialmente a mayor extensión de cultivos. Adicionalmente, otro desafío importante es su limitación en cuanto a la escalabilidad. A medida que aumenta el tamaño del campo agrícola, se vuelve cada vez más difícil y costoso llevar a cabo el conteo manualmente. Esto puede resultar en áreas del campo que no se cuentan o se cuentan de manera deficiente, lo que nuevamente conduce a estimaciones inexactas y decisiones de gestión subóptimas.

¹ En este proyecto, "conteo manual" se define como el conteo realizado a través de la visualización de imágenes.

Una de las principales funciones de Eiwa es experimentar en los grandes terrenos de sus clientes. Estos experimentos consisten en sembrar bajo diferentes condiciones y comparar los resultados, lo que les permite ofrecer una amplia gama de análisis y asesoramiento personalizado a los clientes. Una de las formas de comparación entre ensayos de siembra es la cantidad de plantas que surgen de cada uno. Para evaluar este aspecto, Eiwa contrata verificadores que cuentan, a través de fotografías, la cantidad de plantas en cada parcela de cada ensayo. Este enfoque ayuda a los agricultores a tomar decisiones informadas sobre los momentos óptimos y las condiciones más favorables para la siembra, con el objetivo de maximizar tanto la cantidad como la calidad de los cultivos.

En este contexto, se propone desarrollar una herramienta que automatice el conteo de plantas, permitiendo a Eiwa brindar a sus clientes agricultores la capacidad de tomar decisiones más informadas y eficientes, maximizando tanto la cantidad como la calidad de los cultivos.

3. Estado del arte

Para desarrollar la solución propuesta, se realizó una investigación sobre los casos de uso existentes en relación con los objetivos propuestos. El propósito fue comprender mejor las metodologías y enfoques utilizados en el conteo de plantas de maíz. Esta investigación incluyó el análisis de diversos modelos de stand count implementados en distintos contextos agrícolas.

El Departamento de Ingeniería de Biosistemas y Ciencia de los Alimentos de la Universidad de Zhejiang en Hangzhou, China, trabajó con el Departamento de Ingeniería Agrícola e Ingeniería de Biosistemas de la Universidad Estatal de Iowa en Ames, Iowa, EE. UU. para desarrollar un modelo². Este modelo se basa en imágenes de video recopiladas en parcelas de maíz en diferentes etapas de crecimiento para realizar el recuento de plantas.

Este método, basado en redes neuronales convolucionales, utiliza la red YoloV3 y un filtro de Kalman para contar las plántulas de maíz en línea. Los resultados muestran que el método es preciso y confiable, logrando una accuracy de más del 98%. Además, la utilización de una red de detección de una sola etapa de alta velocidad, como YoloV3-tiny, posibilita la realización de conteos en tiempo real, lo que representa una ventaja significativa en términos de eficiencia y costo.

Otro modelo a tener en cuenta es U-Net³, que también fue utilizado para el conteo de plantaciones de maíz. U-Net es una arquitectura de red neuronal convolucional (CNN) que se ha utilizado ampliamente en tareas de segmentación de imágenes, como la identificación y delimitación precisa de objetos en imágenes médicas y de computer vision.

El modelo U-Net se caracteriza por su estructura en forma de U, compuesta por el "codificador" y el "decodificador". El codificador, en la parte superior de la U, utiliza capas convolucionales y de agrupación para extraer características clave de la imagen, reduciendo su resolución pero aumentando la profundidad de las características. Por otro lado, el decodificador, en la parte inferior de la U, utiliza capas de convolución transpuesta para aumentar la resolución de las características extraídas, fusionándolas con las del codificador mediante conexiones de salto. El modelo se entrena con un conjunto de datos etiquetado, ajustando sus pesos para minimizar la

² Wang, L., Xiang, L., Tang, L., & Jiang, H. (2021). A Convolutional Neural Network-Based Method for Corn Stand Counting in the Field. *Sensors*, 21(2), 507. <https://doi.org/10.3390/s21020507>

³ Vong, C. N., Conway, L. S., Zhou, J., Kitchen, N. R., & Sudduth, K. A. (2021). Early corn stand count of different cropping systems using UAV-imagery and deep learning. *Computers and Electronics in Agriculture*, 186, 106214. <https://doi.org/10.1016/j.compag.2021.106214>

discrepancia entre predicciones y etiquetas. Posteriormente, se utiliza para segmentar automáticamente las plántulas⁴ de maíz en imágenes, asignando etiquetas a cada píxel para identificar la presencia de plántulas.

Otros modelos que podrían utilizarse para realizar una adaptación a este caso particular son:

- *CSRNet (Red de Densidad de Recursos Compartidos)*: Utiliza un enfoque convolucional para estimar la densidad de multitudes en imágenes. Este sistema emplea una red neuronal convolucional (CNN) con una arquitectura truncada de VGG-16⁵ en la parte delantera, que actúa como extractor de características. Además, utiliza capas convolucionales dilatadas para estimar el mapa de densidad. El objetivo es proporcionar una estimación precisa de la densidad de objetos en una imagen, lo que es útil para tareas como la gestión de multitudes y la vigilancia.
- *SaCNN (Red Neural Convolucional Adaptativa a la Escala)*: Se enfoca en manejar el cambio de escala y perspectiva en imágenes para la detección de multitudes. SaCNN utiliza una red neuronal convolucional que se adapta dinámicamente a diferentes escalas en la imagen. Para ello, emplea una arquitectura similar a VGG-16 para la extracción de características y luego combina mapas de características de diferentes capas para asegurar la robustez ante variaciones de escala. Esto es útil en escenarios donde los objetos pueden aparecer a diferentes distancias y tamaños en la imagen.
- *MSCNN (Red de Múltiples Escalas)*: Está diseñada para extraer características relevantes a la escala de una imagen mediante el uso de una red CNN multinivel. MSCNN utiliza múltiples módulos inspirados en Inception para la extracción de características a diferentes escalas. Esto permite capturar información significativa de diferentes contextos espaciales, lo que mejora la capacidad del sistema para reconocer y contar objetos en una variedad de escalas y contextos.
- *CrowdNet (Red de Multicolumna)*: Emplea una arquitectura de red de dos partes para predecir mapas de densidad de multitudes. La primera parte consta de múltiples capas de CNN, mientras que la segunda parte utiliza una red superficial para la predicción final. El uso de una red profunda con una arquitectura similar a VGG en la primera parte permite capturar características complejas de la imagen, mientras que la red superficial proporciona una predicción más precisa del mapa de densidad.
- *DeepCrowd (Multitudes Profundas)*: Utiliza un enfoque basado en regresión para contar el número total de objetos en una imagen. Emplea una combinación de CNN y capas totalmente conectadas para aprender una función que mapea parches de imagen al recuento total de objetos. Este enfoque es útil cuando el objetivo es obtener un recuento preciso de objetos en una imagen sin necesariamente generar un mapa de densidad.

Como se analizó, existen múltiples formas funcionales que se utilizan hoy en día para realizar conteo automático de plantaciones, todas ellas usan una arquitectura compuesta por redes neuronales convolucionales. En este proyecto se apuntará a usar YOLO ya que es conocido por su eficiencia y rapidez, además de tener una gran comunidad que se actualiza constantemente y publica modelos pre entrenados. Si YOLO no alcanza el rendimiento esperado, se evaluará la implementación de otras arquitecturas mencionadas anteriormente, como CSRNet, SaCNN, MSCNN, CrowdNet o DeepCrowd, que también han demostrado ser eficaces en tareas de conteo y estimación de densidad en diferentes contextos.

⁴ Plantas jóvenes de maíz que están en las primeras etapas de crecimiento después de haber germinado.

⁵ VGG-16 es una arquitectura de red neuronal convolucional (CNN). La designación "16" se refiere al hecho de que esta red tiene 16 capas en total, incluidas 13 capas convolucionales y 3 capas completamente conectadas (también conocidas como capas densas).

4. Herramientas y Metodología

El desarrollo del proyecto se realizará en Python. Se planea emplear Redes Convolucionales, comúnmente referidas como CNN (Convolutional Neural Networks) o ConvNets. Estas son arquitecturas de redes neuronales profundas diseñadas específicamente para el procesamiento de imágenes.

Se planea experimentar con las siguientes herramientas:

YOLO (You Only Look Once)

Algoritmo de detección de objetos en imágenes y videos que se distingue por su eficiencia y precisión. A diferencia de otros métodos, YOLO emplea una única red neuronal convolucional para llevar a cabo la tarea de detección. Su enfoque distintivo radica en la predicción simultánea de múltiples cajas delimitadoras y las probabilidades de clase asociadas a estas cajas.

En términos prácticos, esto significa que YOLO analiza la imagen únicamente una vez, en lugar de realizar múltiples escaneos, lo que resulta en una velocidad de detección notablemente rápida. En el contexto del proyecto, YOLO podría identificar cada planta de maíz presente en las imágenes, proporcionando no solo información precisa sobre su ubicación, sino también asignando una probabilidad a cada detección, lo que añade una capa de confianza a los resultados obtenidos.

Una ventaja que tiene el algoritmo de YOLO es la posibilidad de utilizar transfer learning, es decir, tomar un modelo pre-entrenado en un conjunto de datos grande y general, y luego ajustar o "transferir" ese conocimiento al conjunto de datos específico otorgado por la empresa.

Data Augmentations

Técnicas utilizadas en el procesamiento de imágenes y otros tipos de datos para aumentar la diversidad del conjunto de datos original mediante la aplicación de transformaciones aleatorias o controladas. Estas transformaciones se aplican a las imágenes originales para crear nuevas instancias de datos que conservan las características importantes, pero que presentan variaciones en aspectos como la rotación, el escalado, el recorte, el cambio de brillo y contraste, entre otros.

Color Space Conversion and Contouring

Proceso técnico utilizado en procesamiento de imágenes y computer vision. La conversión de espacio de color implica cambiar la representación del color de una imagen de un sistema de color a otro, como de RGB a HSV, para realizar análisis o manipulaciones específicas. El contorneado, por otro lado, implica detectar y resaltar los contornos de objetos en una imagen. Estas técnicas pueden ser de utilidad para la detección de plantas. Cambiar el dominio de color y realizar corridas del modelo sin cada uno de los colores RGB podría ser de utilidad para evaluar como mejora la performance del modelo.

Google collab y Github

Se utilizará Google Collab como plataforma para desarrollar el modelo ya que permite, en el caso de tener limitaciones en el procesamiento, acceso gratuito a recursos de hardware acelerado y tiene integración con el ecosistema de Google.

A su vez, con el objetivo de manejo de versiones, trabajo colaborativo y la presentación ordenada del código realizado se utilizará un repositorio de GitHub.

5. Entregables y Outputs del Proyecto

Durante el desarrollo del proyecto, se llevaron a cabo diversas entregas. Cada entrega mencionada se atribuye a una sección del presente informe.

La actual sección corresponde al primer entregable, donde se introduce a la empresa seleccionada y se detalla la problemática a abordar. En este apartado se establecen los objetivos del proyecto así como el plan de trabajo a seguir y las herramientas que se utilizaron para alcanzar dichos objetivos. Por último se presentan los indicadores clave de rendimiento (KPIs) que se espera afectar y se propone el business case del proyecto, el cual sirve como una herramienta clave para respaldar la toma de decisiones y obtener el respaldo necesario para la implementación exitosa del proyecto.

En la segunda sección se introducen los datos disponibles y se lleva a cabo un Análisis Exploratorio de Datos (EDA). El EDA proporciona información esencial sobre el dataset, revelando patrones visuales, distribución de características y correlaciones entre las imágenes y los datos asociados. Este análisis es crucial para detectar desafíos potenciales en el conjunto de datos, como variaciones en el tamaño, iluminación, calidad de imagen o la presencia de elementos no deseados en las fotografías, que podrían haber afectado negativamente la precisión del modelo. También permite identificar características relevantes para el modelo, como el color de las plantas, la textura del suelo o la forma de las hojas de maíz, lo que contribuye significativamente a mejorar la capacidad predictiva del modelo y optimizar su rendimiento en la tarea de conteo de plantaciones de maíz.

En la tercera sección del proyecto se revisa y ajusta el enfoque de solución inicial, teniendo en cuenta el EDA y la inmersión en el problema. Se detallan las metodologías a implementar y se explica el método de experimentación. Se centra principalmente en el desarrollo de la solución propuesta, presentada de manera clara y directa. A su vez, se presentan los resultados obtenidos a través de la implementación de la solución y se realizan conclusiones en relación con las hipótesis generadas. Finalmente, se presenta el modelo predictivo final y la evaluación de su rendimiento, teniendo en cuenta los KPIs establecidos en la primera sección.

Por último, en el presente informe se realizó un rearmado completo para brindar coherencia y sentido al trabajo desarrollado. Tal incluye las correcciones del desarrollo de las tres secciones mencionadas según las observaciones de la cátedra y de los compañeros. Además se añadieron módulos adicionales al informe. Esto incluye una conclusión final, donde se repasa el business case con los resultados obtenidos durante el desarrollo del proyecto, se responde a las hipótesis iniciales y se revisa la descripción del problema junto con los KPIs sugeridos. Asimismo, se incluye una sección de Potenciales Próximos Pasos, en donde se detallan los pasos a seguir para la implementación y la medición del éxito, así como también se sugieren mejoras al enfoque utilizado.

6. Plan de Trabajo

El progreso del proyecto se desglosa en las siguientes etapas:

1. *Recopilación de datos y preprocesamiento:* Se inicia con la unión de todos los labels otorgados por la empresa, seguida de su preprocesamiento, corrigiendo cualquier inconsistencia.
2. *Análisis exploratorio de datos:* Esta fase implica una exploración profunda de la base de datos para identificar patrones, tendencias y posibles irregularidades. Se utilizarán técnicas

descriptivas y visuales para examinar la distribución de los datos y las relaciones entre las variables.

3. *Implementación y evaluación de modelos:* Se seleccionarán algoritmos de optimización o de aprendizaje automático considerados más adecuados para resolver el problema. Luego de entrenar y evaluar los modelos, se ajustarán los parámetros para obtener el mejor desempeño posible. Además, se utilizarán gráficos para representar las decisiones tomadas por los modelos.
4. *Validación y comparación de resultados:* Una vez implementados los modelos, se evaluará su eficacia y eficiencia mediante diferentes medidas desarrolladas en el inciso III, y se considerará su eficiencia computacional. Se llegará a una conclusión sobre el enfoque más efectivo para identificar y contar las plantaciones de maíz en las parcelas.
5. Elaboración de la presentación para los clientes: Se ajustarán los detalles del proyecto con el objetivo de preparar una presentación profesional adaptada a la audiencia específica.

Es importante destacar que cada etapa del proyecto debe completarse antes de pasar a la siguiente, ya que la superposición no es posible y se requiere un resultado claro al finalizar cada fase. Por esta razón, se utilizará la metodología de gestión de proyectos en cascada -véase Figura I-, que es ideal para proyectos con resultados claramente definidos desde el principio.

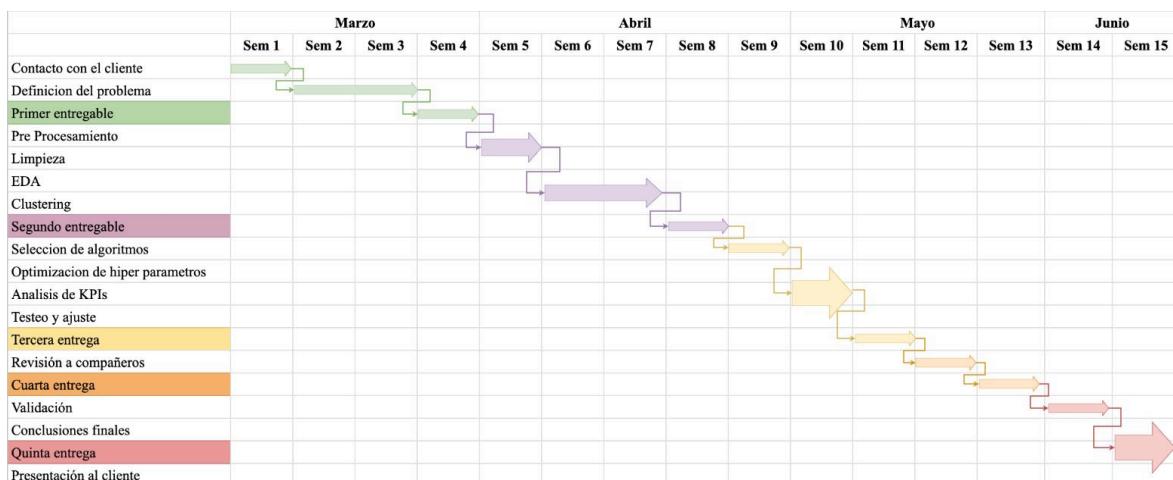


Figura I: Diagrama de Gantt del proyecto

A su vez, véase la Tabla I para observar un análisis de riesgos con respecto al cumplimiento de objetivos.

Riesgo	Probabilidad de ocurrencia	Gravedad del impacto	Plan de mitigación
Limitación dada una cantidad de imágenes escasa	Baja	Bajo	Informar a Eiwa sobre la limitación con el fin de obtener más registros o

			generar información adicional mediante métodos automatizados.
Tiempo muy elevado de entrenamiento/ Poca capacidad de procesamiento	Alta	Medio	Notificar a Eiwa sobre la limitación y solicitar acceso a su infraestructura en la nube para obtener un mayor poder computacional.
Bajo rendimiento dado el benchmark	Media	Alto	Emplear técnicas de optimización de algoritmos y ajuste de hiper parámetros para mejorar el rendimiento del modelo y acercarlo a los estándares del benchmark establecido, junto con la exploración de posibles mejoras en el hardware o la infraestructura utilizada para ejecutar el modelo.
Etiquetado manual de mala calidad	Media	Alto	Informar a Eiwa sobre los errores encontrados para que realicen un control de calidad del trabajo de los verificadores. Disminuir el threshold de IOU utilizado para clasificar a una predicción correcta.

Tabla I: Análisis de riesgo

7. Caso de negocio y KPIs

En esta sección, se detallan los principales KPIs (Key Performance Indicators) para medir el impacto del modelo de conteo de plantas automatizado. Además, se presentará el Business Case, demostrando cómo el modelo automatizado puede reducir significativamente los costos operativos y mejorar la eficiencia, justificando así la inversión y destacando los beneficios económicos y operativos para la empresa.

a. KPIs de Negocio

Los KPIs de negocio que se utilizarán para medir el impacto en el negocio antes y después de la implementación del modelo son los siguientes:

- Tiempo (en horas) dedicado al conteo de plantas en imágenes

$$T = \frac{N}{R} + C$$

Siendo T el tiempo total en horas que demora el planteo, N el número de parcelas a contar, R el número de parcelas que se pueden contar por hora y C el tiempo adicional asociado al conteo.

- Costo (por hora) asociado al conteo de plantas en imágenes

$$Costo \text{ } por \text{ } hora = \frac{Costo \text{ } total}{Tiempo \text{ } total}$$

Los costos a considerar para el costo total son la mano de obra, en caso de conteo no automatizado y costo de utilizar una instancia p3 o g5 de Amazon EC2⁶ en caso de realizar el conteo de manera automatizada. Además se tendrá en cuenta el costo de la mano de obra necesaria para realizar las verificaciones que sean necesarias.

b. Enfoques primarios para evaluar el desempeño del modelo

En el problema que se está abordando, existen dos enfoques principales para evaluar el desempeño del modelo. El primero implica medir qué tan precisa es la predicción en términos de la cantidad real de plantas detectadas. El segundo enfoque consiste en verificar si las características identificadas por el modelo en las imágenes realmente corresponden a plantas. Es esencial destacar esta distinción, ya que durante la experimentación con algoritmos, un modelo podría demostrar un rendimiento destacado al aproximarse mucho al número real de plantas en el recuento, pero al mismo tiempo podría estar identificando elementos en las imágenes que no son plantas. Este último caso constituye un problema significativo si no se detecta, por lo tanto, es necesario evaluar el modelo no sólo en términos de precisión en el recuento, sino también en cuanto a la precisión de la detección de plantas en las imágenes.

Los principales KPIs que se utilizarán para evaluar el rendimiento del modelo son los siguientes:

- *Intersection over Union (IoU)* : se calcula como el área de intersección entre la predicción y la real dividida por el área de unión entre ambas. La intersección es el área donde las predicciones del modelo y la verdad fundamental⁷ (la información correcta, en este caso, las áreas reales donde hay plantas) se superponen y la unión se refiere al área total cubierta tanto por las predicciones del modelo como por la verdad fundamental.

$$IoU = \frac{\text{Área de intersección}}{\text{Área de unión}}$$

Antes de proceder con las siguientes métricas, es fundamental establecer claramente el significado de un *falso positivo* y un *falso negativo* en este contexto. Las siguientes métricas se derivan del resultado del IoU, donde se establece un umbral que determina cuándo se considera que el modelo ha identificado correctamente una planta y cuándo no.

De acuerdo con esta lógica, se define como *falsos positivos* a aquellos casos en los que el modelo predijo la presencia de una planta en un área específica donde no había ninguna (siempre

⁶ Las instancias EC2 g5 y p3 de AWS se definirán más adelante en la sección 7.d *Costos y Escenarios*.

⁷ En este caso, la verdad fundamental es el conteo manual realizado por los verificadores que, como se desarrolla más adelante, tampoco es perfecto. Es la única referencia posible para evaluar al modelo. Si la precision y el recall no alcanzan 1, puede ser debido a que la referencia no es una representación perfecta de la realidad. En algunos casos, el modelo podría predecir mejor que el verificador manual.

respetando el umbral establecido). Por otro lado, los *falsos negativos* se refieren a los casos en los que el modelo no ha detectado la presencia de una planta en un área donde efectivamente existía una.

Por otro lado, se define como *verdaderos positivos* aquellos casos en los que el modelo detectó correctamente la presencia de una planta. Por último, los *verdaderos negativos* contemplan los casos en los que el modelo no predice la presencia de una planta y efectivamente no la había; **estos últimos corresponden al fondo de las imágenes y no se utilizan para calcular las métricas finales.**

Ya aclarado esto, se continúa con la presentación de las métricas:

- *Recall*: Proporción de instancias positivas que fueron correctamente identificadas por el modelo. La utilización de esta métrica es óptima cuando es importante evitar los falsos negativos y cuando la detección de casos positivos es crítica. Un alto recall significa que el modelo es capaz de identificar la mayoría de las plantas presentes en las imágenes. El recall se calcula de la siguiente manera:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- *Precisión*: Proporción de predicciones correctas positivas sobre el total de predicciones realizadas por el modelo. Un valor alto de precisión indica que el modelo tiene una baja tasa de falsos positivos, es decir, que cuando predice la presencia de una planta, es probable que sea correcto. Se rige por la siguiente fórmula:

$$\text{Precisión} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- *F1-Score*: Es una medida que combina precisión y recall en un solo número. Resulta óptima su utilización en caso de haber un desequilibrio entre las clases y querer encontrar un equilibrio entre falsos positivos y falsos negativos. La precisión también se conoce como valor predictivo positivo y el *recall* también se conoce como sensibilidad en la clasificación binaria de diagnóstico. La fórmula es un promedio armónico entre ambas medidas, formulado de la siguiente manera:

$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2tp}{2tp + fp + fn}$$

c. Enfoques secundarios para evaluar el desempeño del modelo

Para asegurar una evaluación integral del desempeño del modelo, se consideran enfoques secundarios. Estos enfoques incluyen:

- *Tiempo de Inferencia*: tiempo del modelo en procesar y analizar los datos de entrada y producir un resultado adicional. Se medirá en minutos y se respetarán las mismas condiciones para todos los algoritmos a experimentar.
- *Escalabilidad*: cómo el rendimiento del modelo varía con el tamaño del conjunto de datos. Un modelo escalable debería mantener un rendimiento constante o aceptable incluso cuando se enfrenta a conjuntos de datos más grandes, sin comprometer significativamente su tiempo de inferencia o eficiencia.
- *Robustez*: capacidad del modelo para mantener un rendimiento consistente incluso cuando se enfrenta a datos de entrada atípicos o ruidosos. Un modelo robusto es capaz de generalizar

bien a nuevos datos y resistir condiciones inesperadas en el entorno de producción. La capacidad de mantener un rendimiento constante, incluso cuando se enfrenta a datos de entrada atípicos o ruidosos, es fundamental para garantizar la eficacia del modelo en entornos de producción. Teniendo en cuenta que la definición, dimensión, características de suelo y luminosidad pueden variar entre las imágenes, la presencia de data drift⁸ puede comprometer la robustez del modelo, ya que los cambios en las características de los datos pueden introducir sesgos o errores que afectan su capacidad para generalizar a nuevas situaciones. Por lo tanto, evitar el data drift es esencial para preservar la capacidad del modelo de conteo de plantas para adaptarse a cambios en el entorno, garantizando así su utilidad y fiabilidad a lo largo del tiempo.

- *Análisis visual cualitativo:* además de las métricas cuantitativas, es importante realizar un análisis visual de las predicciones del modelo en comparación con las imágenes satelitales originales. Esto puede proporcionar información sobre la capacidad del modelo para identificar correctamente las áreas de plantas y su distribución en el campo agrícola.

d. Costos y Escenarios

Realizar las verificaciones manualmente implica un alto costo en tiempo y capital humano. En Argentina, el servicio de conteo manual se cobra por parcela, presentando un costo que oscila entre 4.636 y 6.954 pesos para un total de 150 parcelas, conteniendo cada foto una parcela. Un verificador requiere entre 2 y 3 horas para llevar a cabo el servicio, lo que supone un costo aproximado de 2.318 pesos por hora de trabajo del verificador. Alternativamente, utilizando modelos de machine learning, el conteo tardaría entre 1 y 2 horas en una instancia p3 de Amazon EC2, con un costo de \$3.06 USD por hora, equivalente a 3.014 pesos al 6 de abril de 2024 para un total de 5.000 parcelas. Cabe destacar que la instancia mencionada justo con sus tiempos y costos es una referencia proporcionada por la empresa.

	Costo en pesos argentinos (por parcela)	Tiempo de conteo en segundos (Por parcela)	Costo total estimado 2024	Tiempo necesario (En horas)
Conteo manual	\$92,78	48	\$27.816.000	4.000
Modelo de machine learning (EC2: p3)	\$0,918	0,058	\$275.400	4,83

Tabla II: Comparación conteo manual y modelo de machine learning en tiempos y costos.

⁸ El data drift es el fenómeno en el que las características o distribución de los datos de entrada cambian con el tiempo de manera significativa. En este contexto, esto podría manifestarse en cambios en la definición, dimensión, características de suelo y luminosidad de las imágenes a lo largo del tiempo. Estos cambios pueden introducir sesgos o errores en el modelo, lo que compromete su capacidad para generalizar a nuevas situaciones.

Según Eiwa, los costos asociados al trabajo manual en Argentina son inferiores en comparación con Brasil y Estados Unidos. Teniendo en cuenta que el número total de parcelas que se espera como demanda en 2024 es aproximadamente 300.000 y que los costos son considerablemente elevados, la reducción de estos gastos resulta indispensable.

El propósito de este modelo es minimizar al máximo las verificaciones manuales. Desarrollar un modelo eficiente y rápido reduciría significativamente los recursos necesarios para su ejecución, disminuyendo así los costos asociados en gran medida.

En un principio el costo dependerá de la cantidad de parcelas procesadas. Como se mencionó anteriormente, al día de la fecha, los costos de Eiwa por el conteo manual se registran de la siguiente manera:

$$\text{Costo verificación por parcela} = \$2.318 \text{ ars} * 2 * 3 / 150 = \$92,78 \text{ ARS}$$

Donde \$2.318 es el costo por hora del verificador, 2 son las horas que tarda cada verificador en realizar el conteo, 3 son los verificadores requeridos para contar cada parcela y 150 es la cantidad de parcelas que el verificador cuenta en las 2 horas.

Eiwa ofrece el servicio de conteo de plantas en parcelas a sus clientes y por esto les cobra:

Precio al cliente por parcela = 0,5 usd = 500 ars (Se toma un valor de tipo de cambio de 1 usd = 1000 pesos argentinos).

Por lo tanto el ingreso actual de la empresa por brindar este servicio se rige por la siguiente fórmula:

$$\text{Ingreso Actual} = (\$500 - \$92,72) * \text{cantidad de parcelas} = \$407 * \text{cantidad de parcelas}$$

Escenario	Tiempo de conteo en segundos (Por parcela)	Costo de conteo en pesos (Por parcela)	Ingreso potencial	Ingreso anual potencial total (en pesos)	Aumento porcentual
Actual (manual)	48	\$92,72	\$407 * parcelas	\$ 122.184.000	N/A
EC2: p3	1,08	\$0,918	\$499,082 * parcelas	\$ 142.770.000	22,54%
EC2: g5	1,188	\$1,155	\$498,845 * parcelas	\$ 142.698.000	22,48%

Tabla III: Tiempo y costo estimado de correr el modelo a desarrollar con distintas instancias⁹

El ingreso potencial de la utilización de las EC2s se calculó bajo la suposición que Eiwa decide no reducir el precio al cliente. En la Tabla III se han propuesto 2 tipos de instancias EC2 de AWS para procesar las imágenes nuevas, suponiendo un modelo ya entrenado. A su vez, en estos escenarios calculados no se considera el costo de una posible revisión manual por medio de verificadores, optada por la empresa. Es esencial aclarar que los valores presentados son estimativos brindados por la empresa, los reales se conocerán luego de la programación del modelo.

⁹ Los tiempos de correr el modelo son estimativos dado a que al no tener modelo todavía se desconoce su eficiencia.

Las instancias EC2 P3 están optimizadas para cargas de trabajo de cómputo intensivo, como el aprendizaje automático, la computación de alto rendimiento (HPC) y las simulaciones de dinámica de fluidos computacional (CFD). Estas instancias utilizan GPUs NVIDIA Tesla V100¹⁰, ideales para tareas de aprendizaje profundo.

Por otro lado, las instancias EC2 G5 están diseñadas para aplicaciones gráficas y cargas de trabajo de uso intensivo de GPU que no necesariamente implican aprendizaje automático, pero que pueden beneficiarse de la aceleración de GPU. Estas instancias usan las GPUs NVIDIA A10G¹¹, adecuadas para gráficos ricos y tareas de streaming de video, así como para aplicaciones de gaming y visualización. Ambas instancias son viables para el proceso implicado, aunque hay que considerar que además del tipo de instancia, el costo dependerá del tamaño de la instancia elegida.

Los costos de entrenamiento inicial del modelo no fueron considerados debido a que se le proporcionará a la empresa el modelo ya entrenado.

e. Posibles Escenarios Resultantes de la Aplicación de Modelos de Conteo

Con el objetivo de plantear posibles escenarios resultantes, es de vital importancia conocer cuales son los indicadores y costos para el modelo actual (verificación manual). Para ello, se llevó a cabo un muestreo aleatorio de un total de 40 imágenes, sobre las cuales se dibujó el trazado de las bounding boxes colocadas en la predicción manual. Dentro de esta muestra, se identificaron un total de 205 verdaderos positivos (plantas correctamente identificadas), 9 falsos positivos (lugares donde se ha marcado una planta erróneamente) y 2 falsos negativos (plantas no identificadas). Los resultados revelados muestran una precisión y recall notablemente altos, que es atribuible al enfoque meticuloso adoptado durante la marcación de las imágenes. En particular, se le asignó a un empleado de Eiwa la tarea de encuadrar manualmente las plantas observadas en las imágenes mediante una herramienta en la computadora. Este método intensivo, aunque efectivo, conlleva un costo monetario sustancialmente elevado.

Escenarios	Recall	Precisión	Costo	Costo total en pesos argentinos EC2:p3	Costo total en pesos argentinos EC2:g5
Actual (Manual)	0,9903	0,9579	\$92,72 *cantidad de parcelas = \$27.816.000	N/A	N/A
Conservador (Automático)	0,6	0,5	\$92,72*cantidad de parcelas*0,35 +costo procesamiento	\$10.011.000	\$10.082.100
Moderado (Automático)	0,75	0,65	\$92,72*cantidad de parcelas*0,25 +costo procesamiento	\$7.229.400	\$7.300.500
Optimista (Automático)	0,9	0,85	\$92,72*cantidad de parcelas*0,05 +costo procesamiento	\$1.666.200	\$1.737.300

Tabla IV: Posibles escenarios resultantes de la aplicación de modelos de conteo

Como resultado, se exhiben cuatro posibles escenarios resultantes -véase Tabla IV-, siendo uno de ellos el escenario actual previamente analizado. Se debe tener en cuenta que para el cálculo del costo total se tuvo en cuenta que se anticipa que la demanda de imágenes a detectar será de 300.000 para este año 2024. A su vez, se consideraron los costos de EC2:p3 y EC2:g5 exhibidos en la Tabla III con el fin de calcular el costo total en pesos argentinos para cada escenario, optando por distintas

¹⁰ Tesla V100 es el motor más potente de Nvidia ideal para aplicaciones de inteligencia artificial y deep learning

¹¹ Tarjeta gráfica de desktop de Nvidia

instancias. El objetivo principal del planteamiento de escenarios alternativos es comprender el costo evitable con un buen rendimiento del modelo, a medida que este otorga resultados mejores, así como también comprender el costo monetario de un modelo que no satisface por completo las necesidades del cliente (llevar a cabo correctamente el conteo con el menor costo posible).

Las métricas de referencia tomadas para el análisis de escenarios son *precisión* y *recall*. La priorización de la precisión sobre el recall, o lo contrario, depende del caso que quiera resolver el modelo¹². En este caso ambas métricas tienen un peso similar dado a que los falsos negativos y los falsos positivos impactan de igual manera en el resultado final, por lo que se decide mantener el ratio que se obtiene en la verificación manual.

Cada escenario tiene un resultado asociado que representa el recall y precisión **mínima** que se debe obtener para ese escenario y asumir el costo que implica el mismo. En caso de que alguno de los valores (o ambos) resultantes del modelo sean tan bajos que no alcancen el escenario conservador, entonces se retomará con el escenario actual, el cual implica la verificación manual del 100% de las imágenes. En caso de que los dos indicadores sean muy distintos entre sí, siendo uno notablemente menor y cayendo en un escenario distinto, se tomará el escenario del indicador que haya sido más bajo.

Lo más importante para Eiwa es aproximarse lo más posible al número real de plantas en las imágenes por el menor costo posible. Con este objetivo, se implementará un proceso de verificación manual selectiva de imágenes tras la ejecución del modelo automatizado. La proporción de imágenes sujetas a revisión manual dependerá directamente de las métricas de rendimiento obtenidas:

- *Escenario Conservador*: Se revisará manualmente el 35% de las imágenes.
- *Escenario Moderado*: Se revisará manualmente el 25% de las imágenes.
- *Escenario Optimista*: Se revisará manualmente el 5% de las imágenes.

La empresa ha especificado estos porcentajes bajo el argumento de que equilibran adecuadamente el rendimiento del modelo con el porcentaje de verificación manual, permitiendo así una gestión eficiente de los recursos en función del desempeño observado del modelo. Es de suma relevancia conseguir un rendimiento igual o mayor al escenario conservador, ya que, de no ser así, el modelo no resultaría rentable en términos económicos. En caso que no se alcance el umbral, mínimo el costo para Eiwa sería el actual sumado al costo del modelo:

$$\text{costo total} = (\text{costo modelo} + \text{verificación manual}) * \text{cantidad de parcelas}$$

El procesamiento de imágenes mediante modelos de machine learning reduce los costos hasta en un 90% en comparación con el método completamente manual. Esto representa una ventaja económica considerable para Eiwa. Por lo tanto, aunque no se logre eliminar completamente la verificación manual, la reducción de su frecuencia puede contribuir significativamente a optimizar recursos y costos, permitiendo a la empresa mantener un alto nivel de precisión sin sacrificar la eficiencia económica.

¹²

Evidently AI. (s. f.). Cómo usar el umbral de clasificación para equilibrar la precisión y la exhaustividad [Página web]. Recuperado de <https://www.evidentlyai.com/classification-metrics/classification-threshold>

Sección II

Análisis Exploratorio de datos

8. Introducción

El análisis de esta sección se enfocará en comprender la calidad de los datos disponibles, identificar patrones y tendencias, y proponer una estrategia fundamentada en los resultados obtenidos. Además, se abordará el tema de la calidad de los datos, el planteo de hipótesis, el ajuste de enfoque y el cálculo del Business Case estimado. Se prestará especial atención al tratamiento de datos y variables, así como a la medición de KPIs que permitan evaluar el impacto del modelo propuesto en el negocio, como el tiempo dedicado al conteo de plantas, el costo asociado a dicho proceso y la eficiencia del modelo en términos de precisión y tiempo de inferencia. El procesamiento de los datos posteriormente descrito y la realización de las gráficas se encuentra disponible en el siguiente repositorio de [Github](#).

9. Hipótesis de trabajo

La utilización de técnicas de procesamiento de imágenes y algoritmos de aprendizaje automático en el conteo de plantas de maíz en campos agrícolas permitirá reducir el tiempo, los costos y los recursos necesarios para realizar el conteo de plantas de maíz en grandes extensiones de terreno, manteniendo o superando los indicadores del conteo manual.

10. Datos disponibles

a. Imágenes

La colección consta de 2.000 imágenes de campos de maíz tomadas por drones en una amplia gama de condiciones ambientales y diversos lugares geográficos. Esto resulta en notables diferencias en la topografía del terreno entre las imágenes. Además, la colección fue capturada a distintas alturas de vuelo, lo que afecta la calidad y resolución de cada fotografía. Estas imágenes también muestran cómo las plantas de maíz interactúan con su entorno, revelando situaciones en las que la falta de luz solar o la presencia de sombras plantean desafíos adicionales.

A modo de ejemplo, véase Figura II, donde se muestra un campo de maíz en pleno crecimiento, con una buena definición y una textura detallada. La perspectiva aérea ofrece una vista amplia del cultivo, así como también condiciones climáticas ideales y un suelo con poco rastrojo y sin malezas.



Figura II: Imagen de referencia de una parcela.

La base comprende principalmente tres países de gran extensión geográfica: Argentina, Brasil y Estados Unidos. Los datos analizados comprenden a un total de 42 regiones, siendo 10 de

Argentina, 6 de Estados Unidos y 26 de Brasil, cuyas características se detallarán posteriormente. Es relevante que el modelo pueda identificar las plantas de maíz en cada uno de ellos a pesar de sus diferencias que radican principalmente en la composición de los suelos.

b. Etiquetas (labels) con metadatos

El proceso comienza con la captura de imágenes. Cada vuelo de un dron captura un total de una imagen que luego es, mediante un programa, dividida en partes para su mejor procesamiento. De cada vuelo, al que se reconoce como *layout*, surgen un total de 20 imágenes resultantes que no se superponen entre sí.

Paralelamente, también manualmente se le asignan etiquetas a cada layout, que pretenden ser descriptivos de la imagen. De esta manera, funcionan para comprender qué características suelen presentar las imágenes, teniendo en cuenta que algunas pueden implicar desafíos futuros para el modelo, como por ejemplo la presencia de plantas dobles o imágenes borrosas. Tales etiquetas asociadas a cada imagen están contenidas en un archivo de Excel, en donde se le asignan varias a un mismo layout y es representativo para las 20 imágenes que surgen de tal layout. El objetivo de las etiquetas es, en definitiva, ofrecer información adicional y esencial para el análisis y procesamiento automatizado. Cada etiqueta en la colección representa una variable binaria que describe características específicas de la imagen, con la excepción del país de procedencia y región:

- *País de procedencia*: Identifica el país donde se capturó la imagen.
- *Región*: identifica la región geográfica del campo.
- *Suelo oscuro*: Indica si el suelo en la imagen muestra tonalidades oscuras o una falta de luminosidad. Esto puede deberse a la composición del suelo, la presencia de sombras u otros factores.
- *Suelo rojo*: Determina si el suelo exhibe tonalidades rojas. Esta característica puede ser indicativa de la presencia de óxidos de hierro u otros minerales en el suelo.
- *Alta densidad de plantas*: Este criterio señala si la imagen muestra una concentración significativa de plantas de maíz, lo que sugiere una densidad de siembra alta en el campo.
- *Sombra*: Informa sobre la presencia de sombras en la imagen. Las sombras suelen ser generadas por las mismas plantas según su perspectiva en relación a la luz solar.
- *Rastrojo*: Identifica la presencia de rastrojos, es decir, los restos de cultivos previos que permanecen en el campo después de la cosecha. Los rastrojos pueden afectar diversos aspectos agronómicos, como la retención de humedad y la fertilidad del suelo.
- *Maleza*: Indica si hay presencia de maleza en la imagen. La maleza se refiere a las plantas no deseadas que crecen junto con los cultivos y pueden competir con ellos por recursos como agua, luz y nutrientes.
- *Borroso*: Señala si la imagen presenta un grado de borrosidad, es decir, si los objetos o elementos en la imagen no están claramente definidos o son difíciles de distinguir.
- *Imagen oscura*: Indica si la imagen en su conjunto es oscura, es decir, si la iluminación general es baja o si hay una falta de brillo en la imagen.
- *Imagen clara*: Determina si la imagen es clara o muy iluminada, con una iluminación suficiente que permite una buena visibilidad de los detalles.
- *Plantas crecidas*: Marca si las plantas de maíz en la imagen están desarrolladas, es decir, si han alcanzado un tamaño considerable en su ciclo de crecimiento.

- *Plantas dobles*: Señala la presencia de plantas dobles, es decir, dos plantas de maíz que han crecido juntas en una misma ubicación. Esto puede ocurrir debido a prácticas de siembra o condiciones ambientales.

c. Archivo JSON con metadatos

Una vez obtenidas las 20 imágenes resultantes de un vuelo, se procede con la demarcación de las imágenes en donde los verificadores manualmente demarcan las fotos ubicando bounding boxes, utilizando el programa Darwin.

Una **bounding box**, en términos conceptuales, constituye un rectángulo circundante que encapsula un objeto específico en una imagen, precisando su ubicación y clase (por ejemplo: maíz). En tanto en ámbitos académicos como profesionales, las cajas delimitadoras se utilizan primordialmente en tareas de detección de objetos, un campo de estudio enfocado en la identificación y localización de múltiples elementos dentro de una imagen. En el caso de la delimitación de plantas de maíz, estas cajas se aplicarían para identificar y contabilizar individualmente cada planta dentro de la imagen, facilitando así un análisis cuantitativo de la distribución y densidad de los cultivos. El archivo de JSON con metadatos contiene información detallada asociada a cada imagen y es crucial para el análisis, ya que incluye la variable que se necesita predecir: la ubicación de las bounding boxes:

- *file_name*: El nombre de la imagen a la que hace referencia.
- *url*: La URL de acceso a la imagen a la que hace referencia.
- *annotations*: Una lista de anotaciones para cada imagen, que proporcionan datos esenciales para el entrenamiento de modelos de detección. Estas anotaciones incluyen:
 - *bounding_box*: Un conjunto de cuatro coordenadas que definen un rectángulo encerrando a una planta visible en la imagen. Cada bounding box está compuesta por: x: La coordenada horizontal del extremo izquierdo de la bounding box sobre el eje X de la imagen. y: La coordenada vertical del extremo superior de la bounding box sobre el eje Y de la imagen. w: El ancho del rectángulo, extendiéndose hacia la derecha desde el punto X. h: La altura del rectángulo, extendiéndose hacia abajo desde el punto Y.
 - *id*: Un identificador único asignado a cada bounding box para rastrear y referenciar específicamente cada anotación.
 - *name*: La etiqueta asociada con la bounding box, como "1 plant", que confirma la presencia de una planta dentro del rectángulo especificado.

11. Adquisición de datos

Las imágenes de los campos de maíz fueron capturadas mediante el uso de drones operados por Eiwa, en colaboración con una empresa especializada en tecnología de drones. Este enfoque permite obtener vistas aéreas detalladas de los campos.

Las etiquetas binarias para cada imagen, que describen características específicas como el tipo de suelo y la densidad de plantas, fueron meticulosamente etiquetadas a mano por personal de Eiwa.

Por último, las bounding boxes, fueron colocadas manualmente por empleados de Eiwa. Es de suma importancia aclarar la proveniencia de estos datos, ya que este procedimiento depende de la precisión humana y puede introducir variaciones en los datos, como inconsistencias en el tamaño y la posición de las cajas. Por lo tanto, es crucial implementar rigurosas verificaciones de calidad para asegurar que las bounding boxes reflejen con precisión las ubicaciones de las plantas, minimizando el impacto de los errores humanos en el entrenamiento de modelos de detección.

12. Limpieza y detección de anomalías

a. Muestreo aleatorio

Como primer paso en el proceso de limpieza de datos y detección de anomalías, se realizó un muestreo aleatorio de 50 imágenes, cada una con sus respectivas bounding boxes delineadas. A continuación en Figura III se presentan algunas de las imágenes obtenidas:

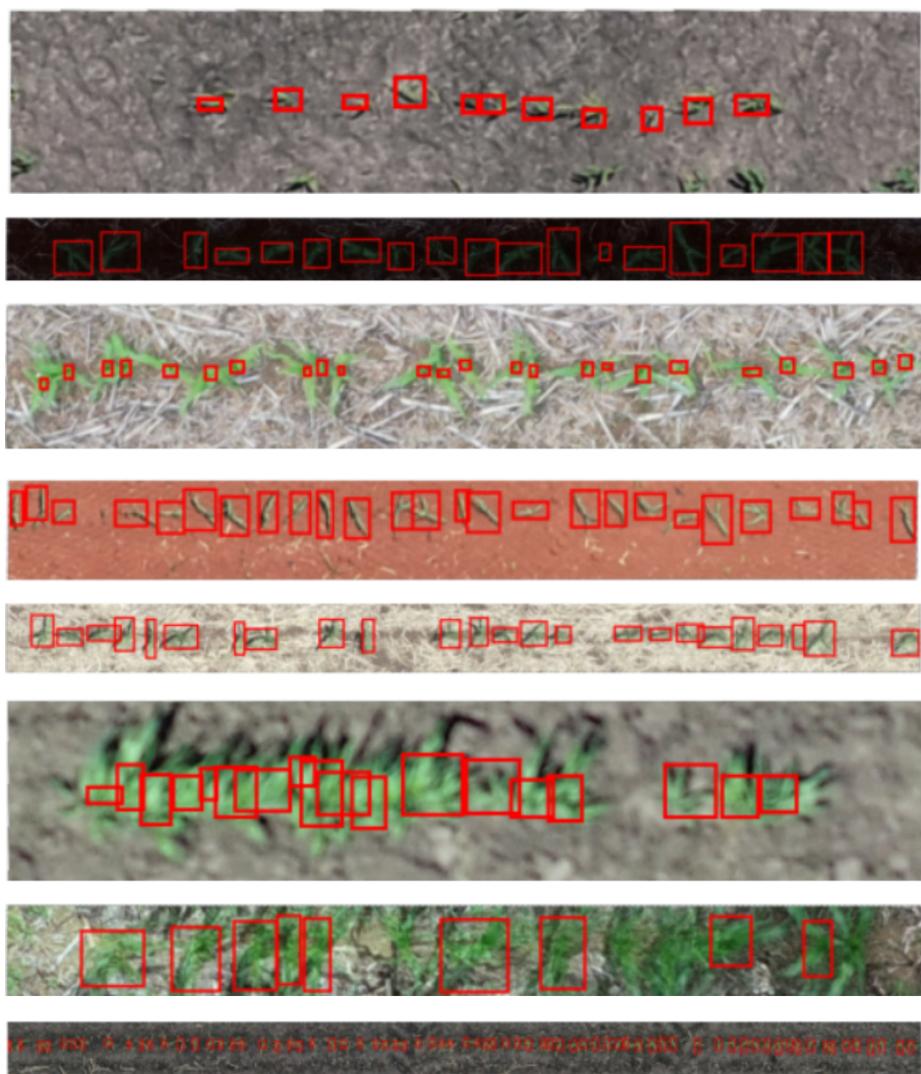


Figura III: Ejemplo de muestreo aleatorio de imágenes

Este análisis inicial destaca la diversidad de las imágenes del conjunto. Se observan distintos tipos de suelos, distintas iluminaciones, variaciones de tamaño y calidad, entre otros. La mayoría de

estos aspectos son considerados en el archivo de etiquetas. Además, este primer paso enfatiza la importancia de la intervención humana en el proceso de etiquetado de plantas.

b. Búsqueda de anomalías

En este apartado se exponen las estrategias utilizadas para la identificación de datos anómalos o errores en la colocación de las bounding boxes.

i. Evaluación de Irregularidades en la Cantidad de Bounding Boxes

El análisis inicial se enfocó en identificar imágenes que presentaran un número de bounding boxes notablemente menor o mayor en comparación con el promedio, buscando así detectar posibles errores. Durante esta experimentación, se determinó que no resulta adecuado comparar la cantidad de bounding boxes por imagen con un valor estadístico general, ya que el número de estos se encuentra influido por las dimensiones de cada imagen. En líneas generales, las imágenes más anchas, o de aspecto alargado, tienden a contener más plantas (véase imagen a.I. del anexo). Por lo tanto, un número reducido de bounding boxes en una imagen alargada podría señalarse como una anomalía, mientras que en una imagen más cuadrada, esto podría considerarse normal.

Para ajustar esta variabilidad y proporcionar una evaluación más precisa, se decidió normalizar la cantidad de bounding boxes en relación al ancho de cada imagen. Esta metodología permite detectar más eficazmente aquellas imágenes que, efectivamente, presentan una cantidad pequeña de bounding boxes para sus dimensiones.

A partir de las imágenes analizadas, se identificó que algunas anomalías eran atribuibles a errores humanos, mientras que otras no lo eran. A continuación, en la Figura IV se presentan ejemplos de imágenes con escasas bounding boxes, donde la escasez se debe efectivamente a la poca presencia de plantas en la imagen:



Figura IV: Ejemplo de imágenes con Bounding Boxes bien colocados¹³

¹³ <https://darwin.v7labs.com/api/v2/teams/eiwa/uploads/fac001f5-1bf8-48ef-a421-9477013d152f>

Por otro lado, se descubrieron imágenes con una deficiencia de bounding boxes atribuible a errores humanos. Se consideraron como errores aquellas imágenes en las que existen plantas sin su bounding box correspondiente. Algunos ejemplos pueden observarse en la Figura V.



Figura V: Ejemplo de imágenes con Bounding Boxes sin colocar¹⁴

En respuesta a este hallazgo, se tomó la decisión de eliminar 23 imágenes con insuficiente cantidad de bounding boxes colocados debido a errores en su etiquetado. La eliminación se realizó para evitar la introducción de ruido en el modelo de predicción. Aunque la solución ideal hubiera sido añadir manualmente las bounding boxes faltantes para preservar toda la información útil para el modelo, esto no se llevó a cabo debido a la complejidad del proceso dado las limitaciones de tiempo. Es importante aclarar que las imágenes detectadas en este paso que no presentaban errores humanos se mantuvieron en la base de datos.

ii. Etiquetado erróneo

Se exploró la posibilidad de que errores humanos, como imprecisiones al usar el mouse, hayan influido en la colocación de bounding boxes. Para identificar estas anomalías, se realizó una búsqueda de bounding boxes cuyas áreas fueran significativamente más pequeñas o más grandes comparadas con el resto proveniente de la **misma imagen**. Es crucial tener en consideración que el área de las bounding boxes está correlacionada con las imágenes; por ejemplo, en una imagen tomada a mucha distancia del suelo, probablemente se observen plantas más pequeñas, y por consiguiente, bounding boxes más pequeñas. Se estableció un umbral para detectar estas anomalías: se marcaron las bounding boxes cuya área fuese un 500% mayor que la mediana (preferida sobre la media, que puede estar muy influenciada por valores atípicos) o que fuese menor o igual al 15% de la misma. Estos umbrales fueron seleccionados luego de probar con distintas alternativas.

Durante la revisión de las imágenes atípicas, se identificaron que algunas anomalías eran efectivamente errores, mientras que otras se debían a la presencia de plantas de tamaño inusualmente pequeño o grande en las imágenes. Se confirmó que un total de 15 bounding boxes habían sido colocadas por error. A continuación en la Figura VI, se presentan ejemplos en los cuales la bounding box erróneamente colocada está destacada en rojo:

<https://darwin.v7labs.com/api/v2/teams/eiwa/uploads/668a71f9-e967-4346-a288-df840283d112>
<https://darwin.v7labs.com/api/v2/teams/eiwa/uploads/0d621d8c-7298-40aa-9c55-2f3360dba3ac>

¹⁴ <https://darwin.v7labs.com/api/v2/teams/eiwa/uploads/760281f0-d8e8-4d85-9952-b79f2842219a>
<https://darwin.v7labs.com/api/v2/teams/eiwa/uploads/f651c4f6-5675-4621-992c-3ddcb6aa4730>
<https://darwin.v7labs.com/api/v2/teams/eiwa/uploads/8ecd1e1b-8d19-43f4-8e8f-d702595868ab>

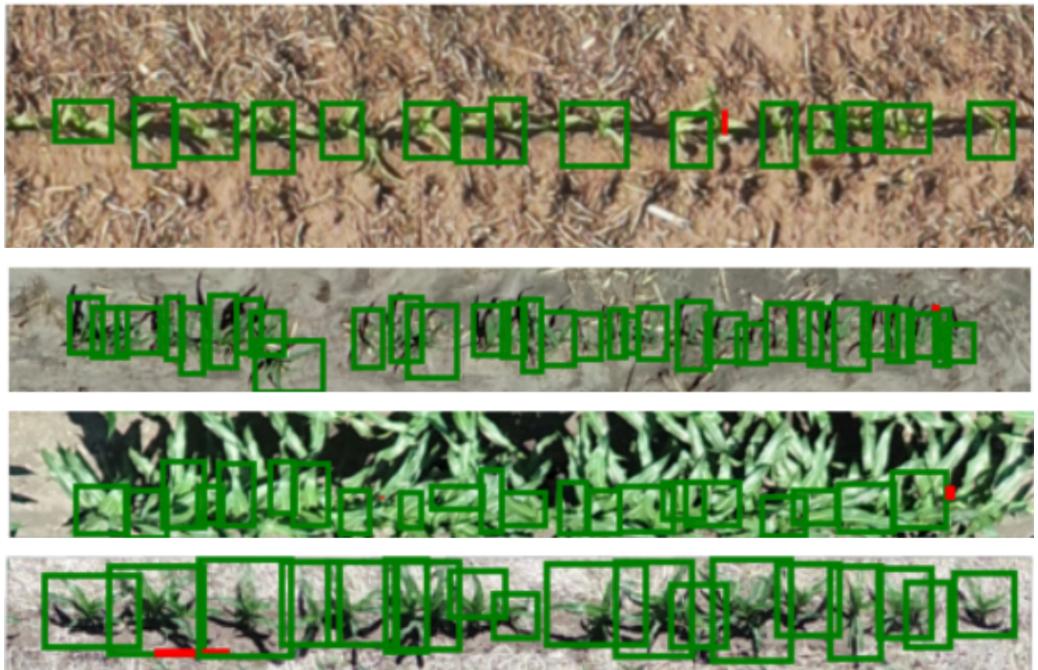


Figura VI: Ejemplo de imágenes con Bounding Boxes colocadas por error

Como resultado del análisis, se optó por eliminar únicamente las bounding boxes que habían sido colocadas de manera errónea. Las imágenes afectadas se conservaron en la base de datos, pero sin las bounding boxes incorrectas.

iii. Overlapped Bounding Boxes

Se realizó un análisis para identificar casos en los que el área de una bounding box estuviera completamente contenida dentro de otra, lo cual podría indicar que una misma planta fue marcada accidentalmente dos veces. Sin embargo, tras una revisión detallada de las imágenes involucradas, se determinó que en todos los casos analizados, las bounding boxes marcaban efectivamente dos plantas distintas, no la misma. En estos casos se trataba de plantas superpuestas. Por lo tanto, no se procedió a eliminar ninguna bounding box. Para ejemplos de superposiciones, véase Figura VII.

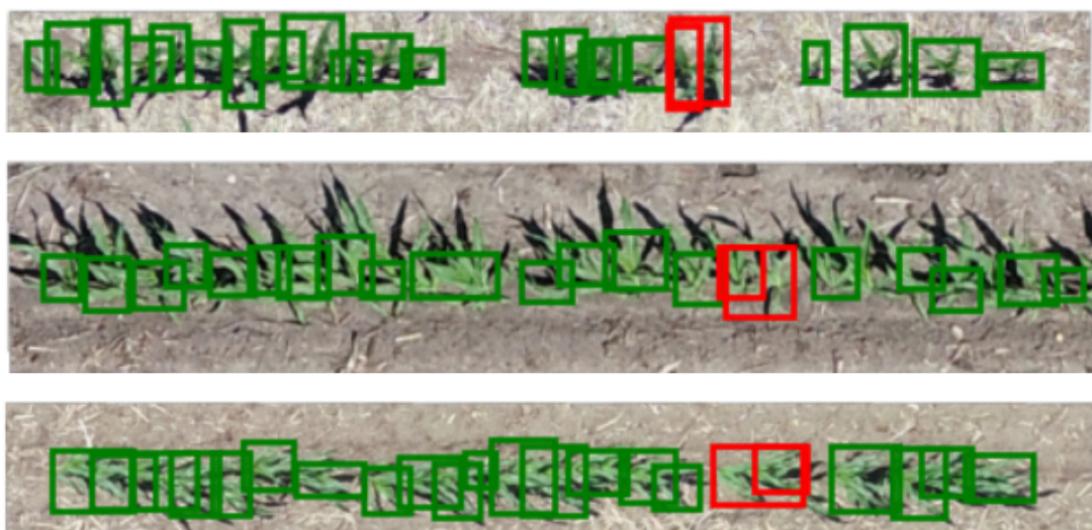


Figura VII: Ejemplo de imágenes con Bounding Boxes totalmente solapadas

iv. Ajustes de bounding boxes

Por último, se realizó un análisis para determinar la presencia de bounding boxes cuya área se extendiera más allá de los límites de las imágenes. Este es un escenario posible debido a la herramienta que utiliza Eiwa para el etiquetado de las plantas.¹⁵ Se encontró que un porcentaje de 22,4% de bounding boxes contaba con al menos un porcentaje de su área fuera de la imagen -véase Figura a.III del anexo-. Ante este descubrimiento, se decidió proceder con su ajuste. Se recortaron estas bounding boxes en los sectores donde traspasaba la imagen, asegurando que ahora coincidan exactamente con los bordes de la misma, sin extenderse más allá.

Algunos ejemplos de imágenes con bounding boxes que terminaban fuera de ella pueden ser observados en la Figura VIII.



Figura VIII: Ejemplo de imágenes con Bounding Boxes cuya área la exceden.

c. Resumen y Conclusiones del la sección

Se inició con la visualización de un muestreo de 50 imágenes para realizar una revisión preliminar de los datos y anticipar posibles errores. Luego se procedió con la búsqueda de errores. Durante este análisis, se detectaron varios tipos de errores, incluyendo bounding boxes que faltaban, se superponían o se extendían más allá de los límites de las imágenes. Tras identificar estas anomalías, se procedió a eliminar las bounding boxes incorrectas y a ajustar aquellas que excedían los bordes, asegurando así la integridad del conjunto de datos. Además, se tuvieron en cuenta factores como el tamaño y las variaciones en las imágenes para una detección efectiva de anomalías.

Si bien se realizó toda la limpieza de datos esencial, las situaciones que hubieran requerido etiquetado manual de bounding boxes no se llevaron a cabo debido a la complejidad del proceso y la insignificancia numérica de los casos afectados.

El proceso de limpieza realizado es crucial para garantizar la precisión y fiabilidad de los datos que alimentan el modelo predictivo, aspectos esenciales para la calidad del análisis futuro.

¹⁵ Eiwa utiliza una herramienta específica para dibujar las bounding boxes, la cual coloca la imagen sobre un fondo transparente. Esto permite extender las bounding boxes más allá de los límites de la imagen original, abarcando también el área transparente.

13. Características generales

Como resultado de la limpieza realizada, el dataset resultante está compuesto por un total de 1.978 imágenes. Las dimensiones de las imágenes varían significativamente, con anchos que oscilan entre los 565 y 3.044 píxeles y altos que van desde 22 hasta 209 píxeles. Como mencionado previamente, existe un total de 42 ubicaciones distintas en las imágenes recopiladas, distribuidas en diferentes países -véase la Tabla V-.

País	Cantidad regiones	Regiones
Argentina	10	ARBA, Las Cejas, Chacabuco, BA, Pergamino, Manuel Ocampo, Carlos Casares, BA, Corral de Bustos, La Cruz, Rio Cuarto, Las Varillas
Brasil	26	Pontão RS, Tibagi, Palma Sola, Imbituva, Guarapuava, Santo Augusto, Castro pr, Mamborê, Londrina FX1 - 5, Paulínia - SP, Ipiranga, State of Mato Grosso, State of Paraná, Toledo, Medianeira, Araguari, Santa Terezinha Itaipu, Campo Mourão, Floresta, Ivatuba, Santa Mariana, Aral Moreira, Rolândia, Palotina, Sidrolândia, Sertanópolis
Estados Unidos	6	Wentworth - MO, Galva - IA, Dallas Center - IA, Spencer - SD, Vermillion - SD, Iowa

Tabla V: Países y regiones presentes en el corpus

a. Tamaño de imagen

Como se mencionó anteriormente, las dimensiones de altura y anchura de las imágenes son variables. Al analizar estas medidas mediante un histograma, se observaron distribuciones significativamente diferentes, probablemente influenciadas por alguna característica categórica. Esto llevó a investigar cuál categoría podría estar causando dichas diferencias. Tras evaluar diversas hipótesis, se determinó que la principal variable afectando las distribuciones era el país de origen de las imágenes, como se ilustra en la Figura IX.

El histograma del alto de las fotos revela que las imágenes de Argentina exhiben alturas que fluctúan considerablemente, con un pico pronunciado alrededor de las 85 píxeles en el eje x. Por otro lado, las fotografías de Brasil muestran una distribución más concentrada, con un pico distintivo alrededor de los 100 píxeles en el eje x. En contraste, las imágenes de Estados Unidos también presentan una distribución concentrada con un pico alrededor de los 140 píxeles en el eje x, aunque con una cantidad total de fotos menor en comparación con Brasil.

En el histograma del ancho de las fotos, se observa que Argentina presenta una distribución con alta varianza. Aunque la mayoría de las imágenes tienen anchos elevados, entre 2.000 y 3.000 píxeles, también se identifican algunos picos en valores más bajos, como 1.000 y 1.300 píxeles. Por otro lado, Brasil muestra una distribución con múltiples picos, la mayoría de ellos entre 750 y 1.500 píxeles, lo que indica la presencia de varias medidas comunes de ancho entre las imágenes. En

contraste, Estados Unidos presenta un pico principal alrededor de los 1.000 píxeles, aunque también se observan algunas imágenes más anchas, superando los 2.000 píxeles.

En conclusión, las fotografías de Brasil y Estados Unidos parecen tener medidas más estandarizadas tanto en alto como en ancho, con picos más definidos y menos dispersión en comparación con Argentina. Esto podría reflejar diferencias en los protocolos de captura de imágenes o en los tipos de cámaras utilizadas en cada país para tomar fotografías de los cultivos. Este aspecto es crucial a tener en cuenta, ya que será necesario monitorear continuamente cómo se toman las fotografías para evitar el data drift¹⁶.

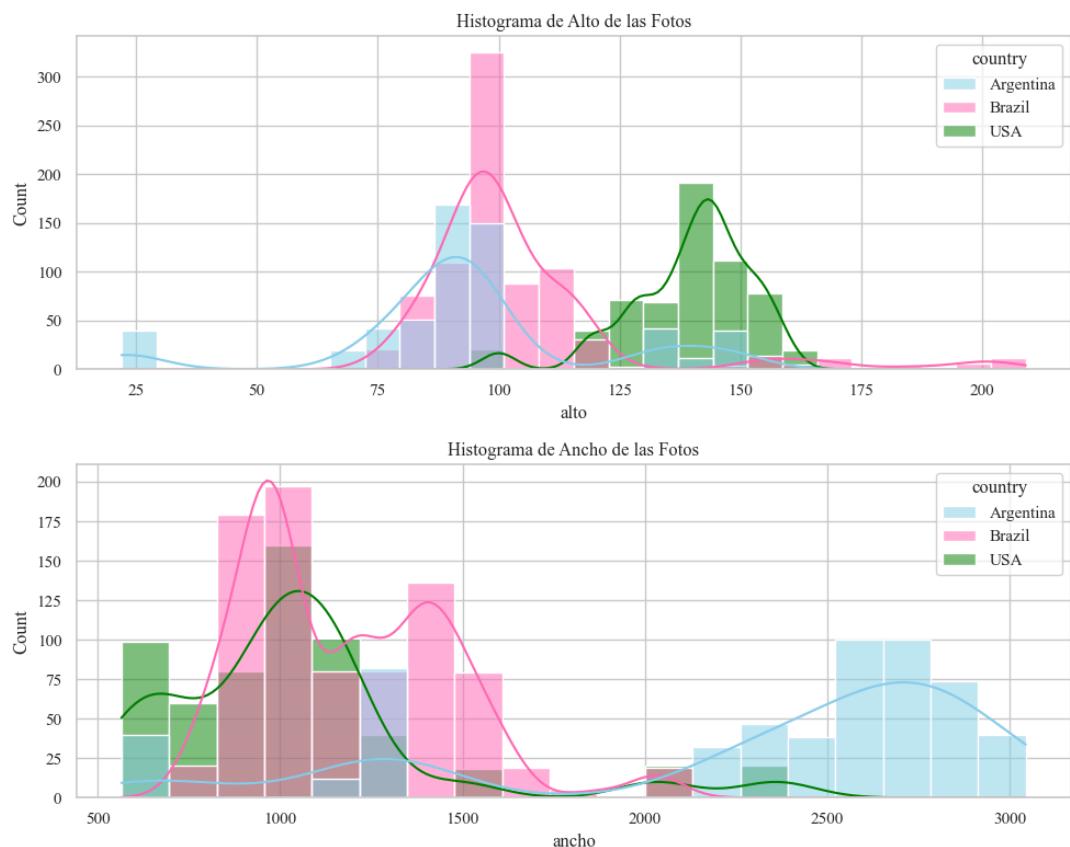


Figura IX: Histograma de ancho y alto de imágenes, por país en píxeles.

Posteriormente, el análisis se centra en un histograma que representa el cociente entre el ancho y el alto de imágenes provenientes de tres países distintos: Argentina, Brasil y Estados Unidos -véase Figura X-. Se observa que en Estados Unidos, la distribución de las relaciones ancho/alto de las imágenes muestra una concentración significativa alrededor de los valores de 5 a 10, lo que sugiere la prevalencia de formatos de imagen cercanos al cuadrado o ligeramente rectangulares. En Brasil, los datos revelan dos picos notables en las relaciones ancho/alto, uno alrededor de 10 a 15, similar a Argentina, y otro pico más estrecho y alto, aproximadamente en el rango de 15, indicando la presencia de dos formatos comunes de imágenes: uno cuadrado y otro significativamente más ancho que alto. Por otro lado, las imágenes argentinas exhiben un pico prominente en las relaciones ancho/alto alrededor de 25 a 30, sugiriendo que estas imágenes tienden a ser considerablemente más anchas que altas, posiblemente con un formato panorámico. Teniendo en cuenta que tener imágenes tan largas

¹⁶ Fenómeno en el que las propiedades estadísticas de los datos de entrada utilizados para entrenar un modelo de aprendizaje automático cambian con el tiempo.

puede ser problemático a la hora de armar el modelo, existe una probabilidad que las imágenes de este grupo resulten las más difíciles.

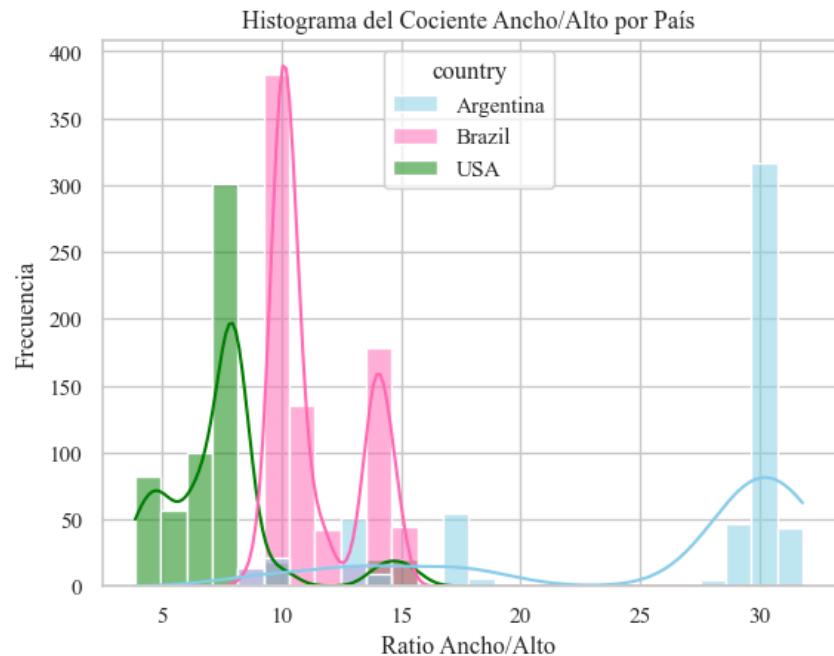


Figura X: Histograma del resultado del cociente entre ancho y alto por país.

b. Hora de vuelo

En cuanto al horario en el que se ha realizado la captura de los layouts -véase Figura XI-, se observan patrones distintos en los tres países del dataset. En Brasil se observan fotos en tres franjas horarias, a las 00:00, entre las 8 y las 11 y entre las 14 y las 16. Se puede observar que Brasil no contiene fotos tomadas al mediodía, lo cuál resulta interesante porque se cree que es en ese horario en el que el modelo tendrá los mejores resultados debido al aumento de luz.

De Argentina, por otro lado, también se tienen fotos tomadas a media noche, además de 9 a 17.

Por último, Estados Unidos no presenta fotos tomadas de noche pero si hay fotos tomadas en todas las horas entre las 9 y 18, lo cuál implica una mayor variabilidad en la iluminación.

La iluminación durante diferentes momentos del día puede influir significativamente en la calidad de las imágenes, siendo un factor crítico en el análisis de imágenes para aplicaciones de Machine Learning. Los momentos pico pueden coincidir con períodos de mejor calidad de imagen, lo que minimiza las interferencias como sombras o deslumbramientos que podrían afectar la precisión del análisis, especialmente en tareas como el conteo de cultivos en aplicaciones agrícolas de precisión.

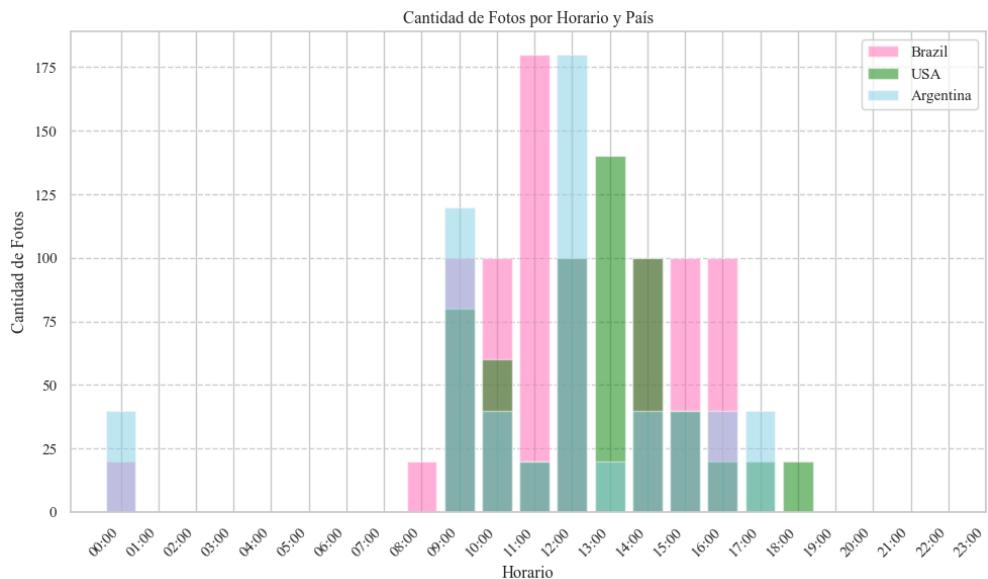


Figura XI: Histograma de cantidad de fotos por país y horario

Tanto Argentina como Brasil registran layouts que se han capturado en horarios de la medianoche, lo que señala que ambos grupos presentaron imágenes notablemente oscuras -véase Figura XII-.



Figura XII: Imagen tomada de un layout capturado durante la noche.

Es esperable que el modelo obtenga un peor rendimiento en las imágenes nocturnas en comparación con las diurnas, ya que la distinción entre las plantas es difícil incluso para el ojo humano. Para evitar un mal desempeño en estas imágenes, es necesario incluir fotos de este tipo en los conjuntos de entrenamiento, validación y prueba. Posteriormente, se deberá evaluar la calidad de la detección en estos casos específicos.

c. Ubicación del campo

Distribución de países de las imágenes. Los clientes de Eiwa que usan el servicio de Stand Count se localizan en tres países; Argentina, Brasil y Estados Unidos -véase Figura XIII-. Es relevante entender cuántas fotos fueron tomadas en cada país porque algunos atributos pueden afectar al funcionamiento del modelo como el tipo de suelo, la exposición solar y las condiciones climáticas. Es observable que Brasil es el país con mayor cantidad de imágenes, así como también el país con mayor cantidad de regiones. Tanto Argentina como Estados Unidos presentan una cantidad similar de imágenes después de realizar el filtrado.

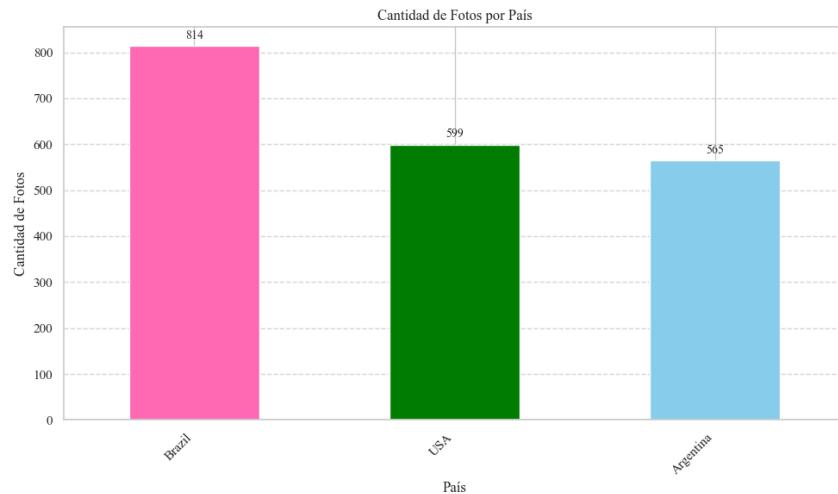


Figura XIII: Histograma de cantidad de fotos tomadas por país.

d. Fecha de vuelo

En cuanto a las fechas de vuelo, existe una clara relación entre la estación del año y el mes en el que fue tomado el vuelo -véase Figura XIV-, principalmente relacionado con la búsqueda de mayor luz durante el día y las fechas de plantación. De esta manera, para el caso de Argentina el registro se concentra principalmente en los meses de noviembre, enero y febrero, meses de los cuales hay mayor cantidad de horas de luz. Asimismo, es similar la tendencia en Brasil -al ubicarse en el mismo hemisferio- solo que con una mayor amplitud durante el año donde los meses donde se han capturado imágenes son octubre, noviembre, diciembre, enero, febrero, marzo y abril. Es posible que la amplitud de meses esté ligada a la distinta cantidad de latitudes que abarca el país. Finalmente, Estados Unidos al ubicarse en el hemisferio opuesto concentra sus datos predominantemente en el mes de agosto y en menor medida en mayo y julio, comprendiendo las estaciones primavera y verano en el hemisferio correspondiente. Es posible que en alguna estación en particular, debido a la cercanía de la tierra con el sol, el modelo funcione mejor para predecir la cantidad de plantas.

Es importante considerar esta información, ya que todas las fotos fueron tomadas en meses cálidos en sus respectivos países de origen. Si esto cambia en el futuro, será necesario revisar cómo performance el modelo con los nuevos datos para evitar el data drift.

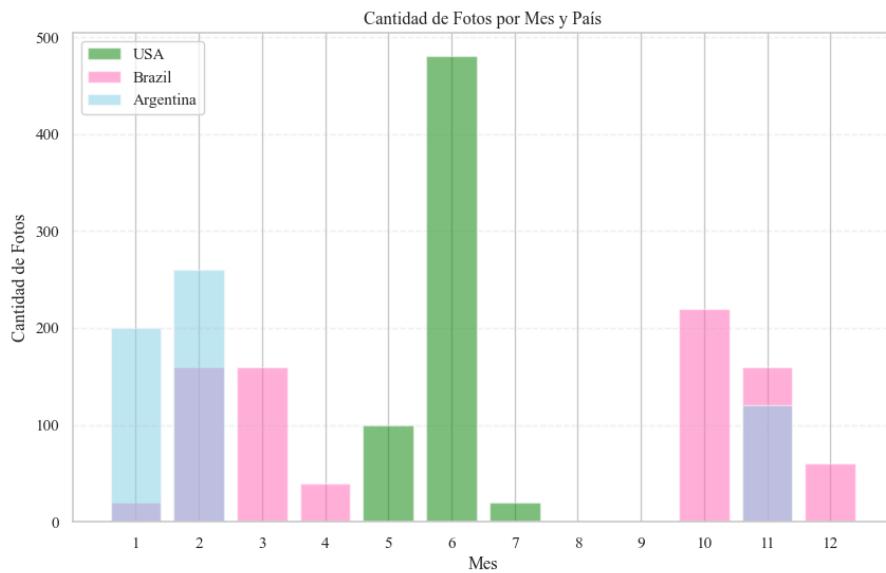


Figura XIV: Histograma de cantidad de fotos por mes por país.

e. Análisis de etiquetas

Retomando las etiquetas mencionadas en el apartado 10.b. *Etiquetas (labels) con metadatos* de la presente sección, se analiza la presencia de las mismas en las imágenes disponibles -véase Figura XV-. Se destaca la frecuencia de las "stubble" (rastrojo) y "shadow" (sombras), lo cual sugiere la prevalencia de condiciones típicas de campo resultantes de una cosecha previa y los efectos de iluminación que pueden influir en el análisis de las imágenes. Asimismo, en tercer lugar se ubica "double plants", siendo esto un desafío tanto para el verificador como para el modelo interpretar la presencia de dos plantas superpuestas. Asimismo, las etiquetas relacionadas con la calidad de la imagen, como "blurry", "dark image" y "bright image", indican la variaciones en la calidad de las imágenes y condiciones.

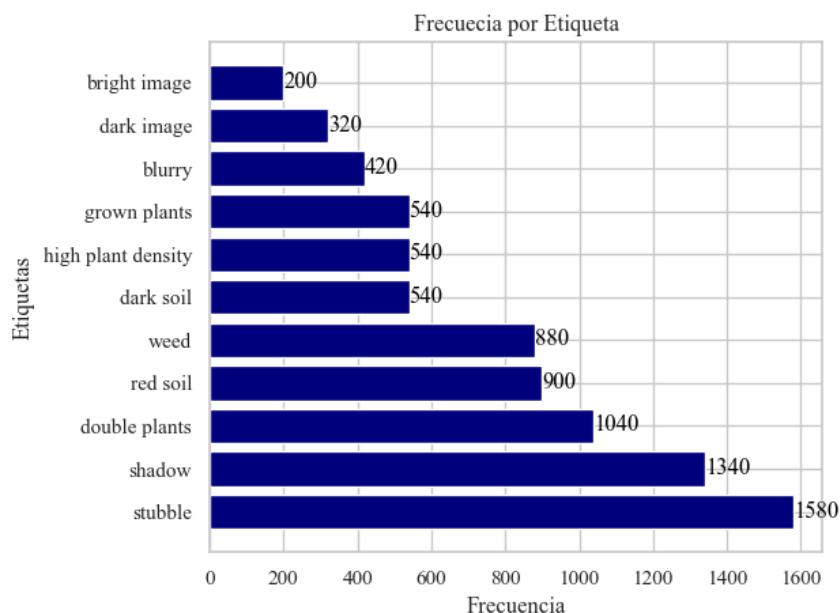


Figura XV: Histograma de frecuencias por etiqueta.

Al comprar el desglose de las etiquetas en función a la proporción en las que las mismas se han presentes por país -véase Figura XVI- es notable que en Argentina, las etiquetas más frecuentes son "shadow" (sombra) y "dark soil" (tierra oscura), presentes en el 80% y 73% de las imágenes respectivamente. "Stubble" (rastrojo) también es común, apareciendo en el 62% de las imágenes, mientras que "weed" (malezas) y "grown plants" (plantas crecidas) tienen proporciones menos significativas, con un 37% y 25% respectivamente. Por otro lado, las etiquetas menos comunes incluyen "bright image" (imagen con brillo), "dark image" (imagen oscura) y "double plants" (plantas dobles/superpuestas).

En Brasil, la etiqueta "red soil" (tierra roja) domina con una proporción del 100%, seguida por rastrojo con un 98%. Además, sombra y alta densidad de plantas están presentes en el 58% y 56% de las imágenes respectivamente. En el caso de Estados Unidos, las etiquetas más comunes son sombra y rastrojo, presentes en el 60% y 57% de las imágenes.

En líneas generales es notable, una alta proporción de imágenes con sombra puede requerir ajustes para manejar variaciones de iluminación, mientras que una alta frecuencia de tierra roja en Brasil indica la necesidad de distinguir plantas del suelo de manera efectiva en ese contexto. A su vez, la proporción de imágenes oscuras o borrosas suelen presentarse en la menor medida para los tres casos, siendo un indicador positivo.

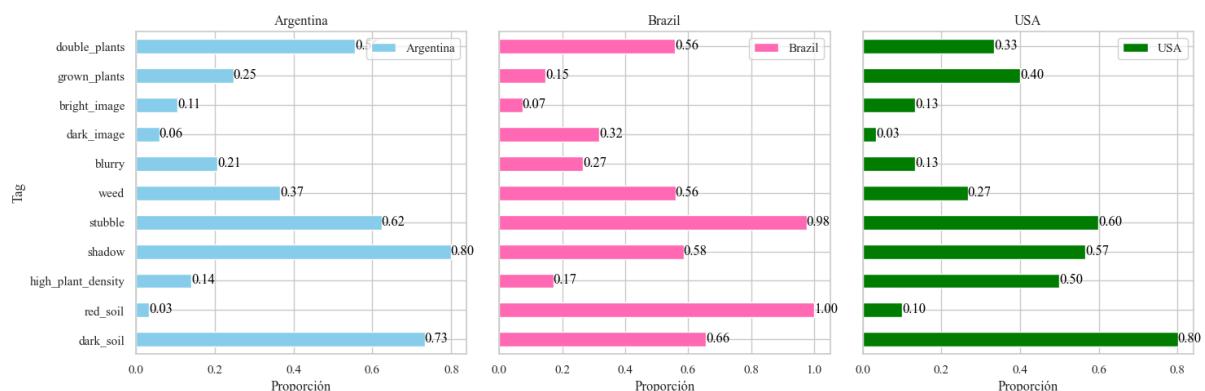


Figura XVI: Histograma de la proporción en la que cada etiqueta se utiliza por país.

f. Resumen y Conclusiones de la sección

Finalmente, véase Figura XVII para un ejemplo de cómo es que se ven imágenes de cada país, las cuales se han tomado aleatoriamente y reflejan el análisis realizado en la presente sección. En Argentina, las imágenes exhiben cultivos con variaciones en la densidad de plantas, junto con la presencia de suelo oscuro, lo que podría coincidir con la etiqueta tierra oscura predominantemente presente en el país. Además, se observa cierta variabilidad en la luz y las sombras.

En Brasil, el suelo rojizo es prominente, en línea con la etiqueta suelo rojo y una gran presencia de rastrojo, coincidente con las etiquetas. Tanto en Argentina como en Brasil se observa una imagen oscura, lo cual puede estar relacionado a la captura de imágenes nocturnas de ambos países.

En Estados Unidos, se pueden ver plantas en varias etapas de crecimiento, y las condiciones de luz aparentan ser más consistentes a lo largo de las filas. Asimismo, se aprecia la presencia de áreas con sombras y otras con mayor densidad de plantas.

Las diferencias en las condiciones del suelo y la vegetación entre las regiones reflejan la diversidad que un modelo debe manejar para ser efectivo en una variedad de entornos agrícolas. Comprender estas diferencias visuales y su distribución puede ayudar a mejorar la precisión de la detección y el conteo de cultivos.

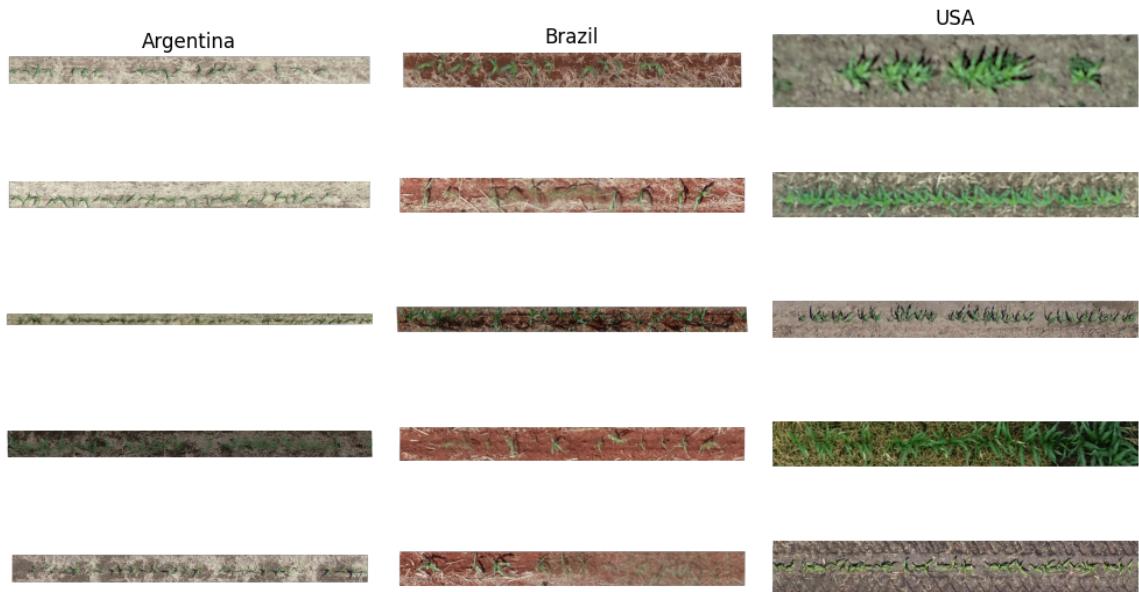


Figura XVII: Cinco imágenes tomadas aleatoriamente por país.

14. Features

En el proceso de automatización del conteo de plantas, se han identificado una serie de características visuales clave que se utilizan para analizar imágenes agrícolas. Estas características, conocidas como *features*, proporcionan información significativa sobre las propiedades de las plantas y su entorno en una imagen. A continuación, se detallan cada uno de los features calculados:

1. **Tonalidad media** (mean_hue): La tonalidad media se refiere al promedio de los colores presentes en la imagen en términos de su matiz. Este feature proporciona información sobre la distribución de los colores.
2. **Saturación media** (mean_hue): La saturación media representa la intensidad o pureza de los colores en la imagen. Un valor alto de saturación indica colores más vibrantes y saturados, mientras que un valor bajo puede sugerir tonos más apagados.
3. **Brillo medio** (mean_brightness): El brillo medio se refiere al nivel medio de luminosidad en la imagen. Este feature proporciona información sobre la iluminación general de la escena.
4. **Rojo, verde, azul** (red, green, blue): Estos features representan la intensidad de los componentes de color rojo, verde y azul en la imagen. Proporcionan información detallada sobre la distribución de colores en la imagen.
5. **Contraste** (contrast): El contraste se refiere a la diferencia de intensidad entre los píxeles vecinos en la imagen. Un alto valor de contraste indica una mayor variación entre los colores y puede ser útil para distinguir a las plantas en la imagen.
6. **Borrosidad** (bluriness): La borrosidad indica el grado de desenfoque presente en la imagen. Un valor alto de borrosidad sugiere una falta de nitidez en la imagen, lo que puede dificultar la identificación precisa de las plantas.

7. **Área** (area): El área se refiere a la medida de la superficie que ocupa dentro de un espacio bidimensional. Se calcula multiplicando la altura de la imagen por su ancho. Cuanto mayor sea el área de una imagen, mayor será la cantidad de píxeles y, por lo tanto, mayor será la cantidad de detalles y la información visual que contiene.
8. **Entropía** (entropy): La entropía es una medida de la incertidumbre o aleatoriedad en la distribución de los valores de píxeles en la imagen. Un valor alto de entropía indica una mayor variabilidad en los niveles de color y puede ser útil para identificar áreas con patrones complejos o texturas diferentes.

Con el objetivo de calcular los features de las imágenes, primero se carga cada imagen de la carpeta especificada. Luego, se convierte cada imagen al espacio de color HSV para analizar sus componentes de tonalidad, saturación y brillo. Se recorre cada píxel de la imagen y se extraen estos componentes, calculando posteriormente el valor medio de cada uno. Además, se convierte cada imagen al espacio de color RGB y se calculan los valores medios de sus componentes de rojo, verde y azul. Este proceso proporciona información sobre la distribución de color en cada imagen.

El modelo de color HSV (Hue, Saturation, Value) se distingue del modelo RGB (Red, Green, Blue) por su enfoque en la percepción humana del color. HSV descompone el color en tres componentes intuitivos: matiz, saturación y valor, que corresponden a la tonalidad, la pureza y la luminosidad, respectivamente. Esta representación facilita la especificación y manipulación de colores en términos más cercanos a la percepción visual humana, lo que lo hace útil en aplicaciones como diseño gráfico y tratamiento de imágenes. Por otro lado, el modelo RGB se basa en la mezcla aditiva de luces rojas, verdes y azules, lo que resulta en la producción de una amplia gama de colores en dispositivos electrónicos -véase Figura a.II. en el anexo-.

Para evaluar la textura y la nitidez de las imágenes, se convierten a escala de grises y se calcula el contraste y la borrosidad utilizando el método de Laplacian de OpenCV. Estos valores indican la variación de intensidad entre los píxeles y la nitidez general de la imagen. Finalmente, para medir la complejidad de la distribución de intensidades de los píxeles en cada imagen, se calcula su entropía mediante la construcción y normalización del histograma de la imagen en escala de grises y la aplicación de la fórmula de entropía de la información.

Se optó por realizar un análisis separado de las características de las imágenes en un heatmap de correlación -vease Figura XVIII- y del espacio de color RGB -Figura XIX-, con el fin de obtener una comprensión más profunda y detallada de los factores que influyen en el conteo de cultivos de maíz. Esta decisión se fundamenta en la naturaleza diferente de las características evaluadas en cada enfoque. Por un lado, el heatmap de correlación se centra en atributos como la borrosidad, el contraste y la entropía, que reflejan la textura, nitidez y complejidad de las imágenes, siendo cruciales para identificar patrones relevantes en el conteo de cultivos. Por otro lado, el análisis del espacio de color RGB se enfoca en aspectos relacionados con la distribución y la intensidad de los colores, ofreciendo una visión adicional de las propiedades visuales de las imágenes. Esta separación permite una evaluación más exhaustiva de los factores visuales que inciden en el proceso de conteo, facilitando así la identificación de características relevantes para el desarrollo de modelos de Machine Learning precisos y efectivos.

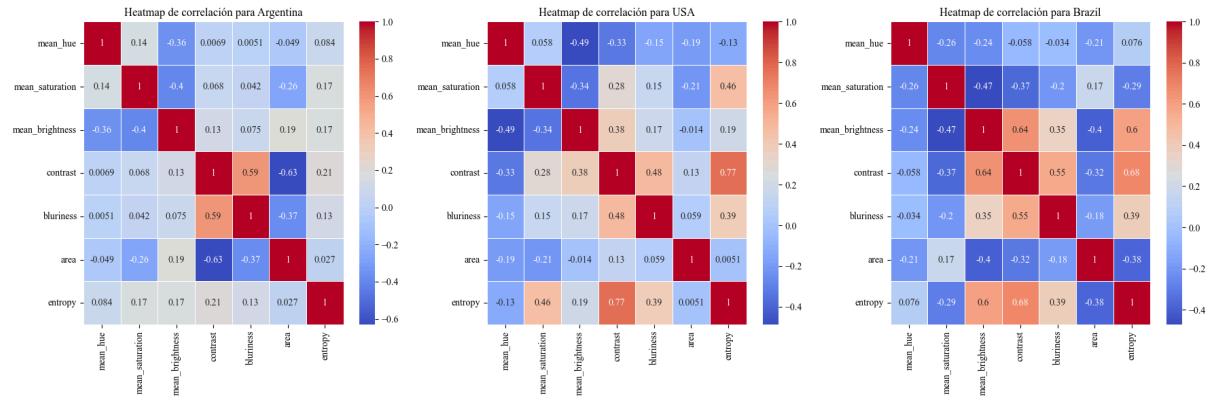


Figura XVIII: Heatmap de correlaciones de features por país.

En el caso de Argentina, las características no muestran correlaciones fuertes entre sí, excepto una moderada correlación negativa entre el área y contraste, lo que podría significar que a medida que aumenta el área de los cultivos detectados en las imágenes, hay menos contraste. Hay correlaciones positivas altas entre la borrosidad y el contraste, lo que indica que en este país a mayor borrosidad mayor es el contraste.

Por otro lado, existe una correlación negativa entre la saturación media y el brillo medio, así como con la saturación media y área, sugiriendo que las imágenes con mayor saturación suelen ser las de menor área y brillo medio. A su vez, se observa una correlación negativa entre el área y la borrosidad, lo que podría interpretarse como que las áreas más grandes tienden a ser las de mayor definición. Estados Unidos suele conservar estos patrones a excepción de una notable correlación negativa entre matiz y contraste. Contrariamente, presenta excepcionalmente una alta correlación entre entropía y contraste, lo que implica que a mayor contraste más pronunciada es la textura de la imagen.

Luego, para el caso de Brasil, las correlaciones más fuertes son positivas entre el contraste y el brillo medio, así como entre el contraste y la entropía, lo que indica que las áreas más contrastadas tienden a ser más texturizadas y posiblemente más brillantes, lo que podría reflejar características específicas del suelo o de las prácticas agrícolas en Brasil. En imágenes con alta borrosidad y bajo contraste, como podría ser el caso en algunos escenarios en Brasil, las plantas podrían ser más difíciles de detectar precisamente. En áreas con alta entropía, que indica una mezcla compleja de texturas, podría ser más difícil diferenciar entre cultivos y maleza, lo que es relevante para los datos de EE. UU. y Brasil.

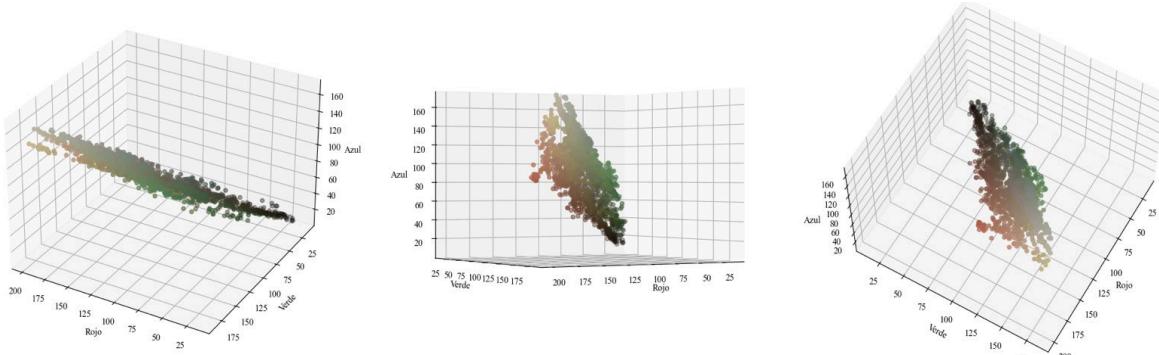


Figura XIX: Gráfico en 3 dimensiones de los valores verde, rojo y azul de las imágenes.

Por otro lado, la integración de los gráficos 3D de los valores RGB permite inferir varias características relevantes para el conteo de plantas de maíz en imágenes satelitales. Se observa una transición de colores desde tonos altos de rojo y azul hacia valores bajos en estos canales y tonos más altos en el verde, indicando una variabilidad asociada a la presencia de vegetación. Además, la densa agrupación en los rangos medios sugiere una abundancia de colores no saturados, mientras que la variabilidad significativa en el canal verde refuerza la sensibilidad de este canal a los diferentes tonos de verde de las plantas. Esto subraya la importancia de las características de color en la diferenciación de las plantas y su entorno, destacando la necesidad de aplicar técnicas de preprocesamiento para mejorar la distinción de las características relevantes en las imágenes, lo que puede mejorar la eficacia de los algoritmos de Machine Learning en el conteo de plantas de maíz.

a. Resumen y Conclusiones de la sección

En esta sección, se profundizó en la exploración de características complejas presentes en las imágenes disponibles. La exploración de estas características visuales revela que factores como la forma y el tamaño de la imagen influyen significativamente en la identificación de las plantas. La evaluación de las imágenes a través de un heatmap de correlación y del análisis del espacio de color RGB ha permitido obtener una comprensión más detallada de los factores que afectan el conteo de cultivos de maíz.

Esta exploración realizada no es trivial, dado que se centra en atributos que serán probablemente cruciales para los modelos de machine learning que se implementarán. Además, resaltar ciertas características en las imágenes, como la saturación o el contraste, puede ser muy valioso para facilitar la detección de plantas.

15. Bounding Boxes

a. Cantidad

En primer lugar, si bien se ha mencionado previamente en sección 13.b.i. del presente informe que existe una correlación entre el ancho de la imagen y la cantidad de bounding boxes, se asume que el mismo es variable. Por esta razón, se ha confeccionado un histograma de la cantidad de bounding boxes por imagen, independientemente de su ancho -véase Figura XX-. Es observable que la mayoría de las fotografías presentan entre aproximadamente 5 y 30 cajas delimitadoras, con un pico notable alrededor de 10 a 15 cajas por imagen. A su vez, la frecuencia decrece a medida que se incrementa el número de cajas delimitadoras por fotografía, lo que sugiere que hay menos imágenes con un gran número de objetos identificados. Se observa una escasez de fotografías con más de 50 cajas delimitadoras, indicando que es menos frecuente encontrar imágenes con una cantidad elevada de objetos, o que el proceso de etiquetado fue más selectivo en estas instancias y lo que demarca una significativa asimetría positiva.

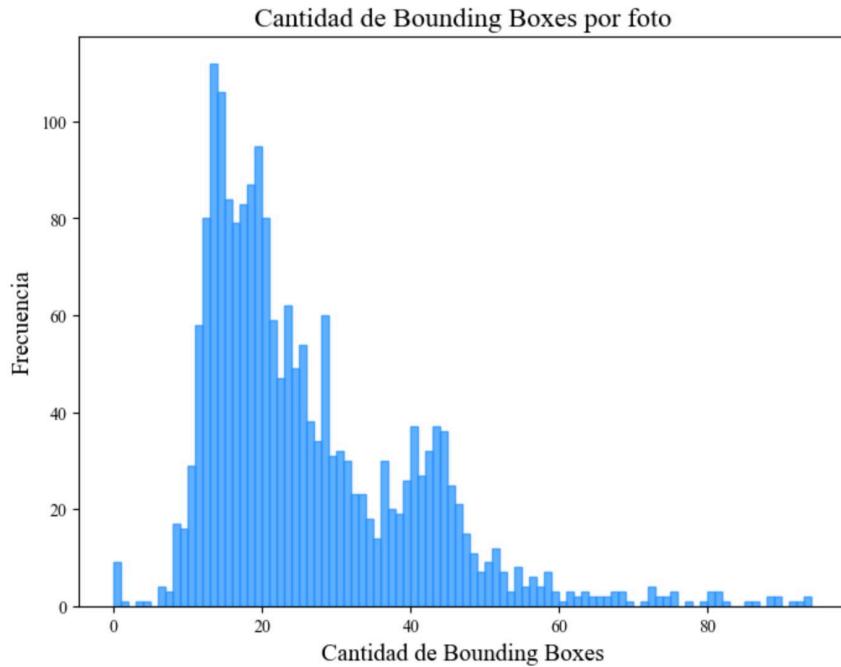


Figura XX: Histograma de frecuencias de cantidad de bounding boxes.

b. Distribución

A su vez, es observable la manera en la que se distribuyen las bounding boxes a lo largo de la imagen (tomando la totalidad de la imagen como un porcentaje) conservan cierto patrón de comportamiento -véase Figura XXI-. Existe una tendencia clara a que las plantas están distribuidas horizontalmente en la mitad inferior de las imágenes. Esto se refleja en la mayor intensidad de color en el rango del 20% al 60% de la proporción horizontal y del 40% al 60% de la proporción vertical. A su vez, en su gran mayoría se concentran en la mitad inferior de las imágenes. Esto se refleja en la mayor intensidad de color en el rango del 20% al 60% de la proporción horizontal y del 40% al 60% de la proporción vertical.

De esta forma, es posible afirmar que la mayor frecuencia de cajas delimitadoras se concentra hacia el centro de las imágenes. Esto podría indicar que las plantas tienden a ser capturadas en el medio del campo visual de las imágenes, lo que es común en las fotografías de campos agrícolas tomadas desde arriba.

Esta distribución podría deberse a prácticas estándar de muestreo o a la configuración de la cámara en drones o satélites que capturan las imágenes. Esta información es especialmente valiosa para la detección y el análisis de cultivos, ya que sugiere que los modelos podrían beneficiarse al prestar más atención a estas áreas específicas de las imágenes donde es más probable encontrar plantas y considerar que es verdaderamente atípica la presencia de bounding boxes en la parte superior de las imágenes, especialmente las esquinas. Además, esto podría ayudar a mejorar la precisión y eficiencia del algoritmo al permitir que se enfoque en las regiones de las imágenes donde las plantas son más frecuentes, optimizando así los recursos computacionales.

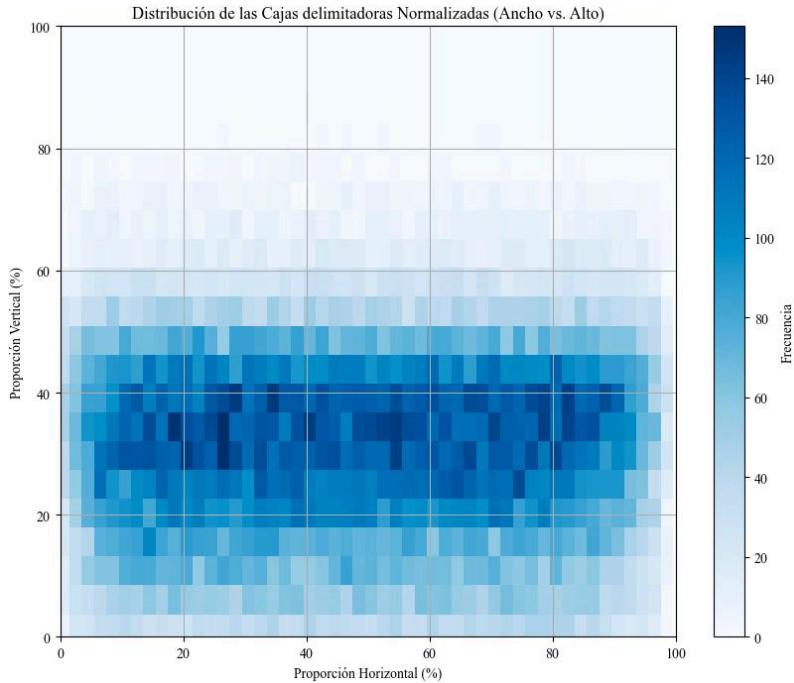


Figura XXI: Distribución de bounding boxes a lo largo de las imágenes, en proporción.

16. Identificación de clusters

El análisis se enfocó en dos estrategias de agrupamiento: una basada en características extraídas de imágenes mediante el modelo VGG16 pre-entrenado y otra utilizando características numéricas directas. Sin embargo, los resultados obtenidos no alcanzaron las expectativas del estudio.

El modelo VGG16 pre-entrenado es una red neuronal convolucional que ha sido previamente entrenada en un conjunto de datos extenso, como ImageNet, para reconocer características visuales en imágenes. Esta arquitectura, conocida por su profundidad y capacidad para capturar detalles complejos, ha aprendido a extraer características útiles como bordes, texturas y formas de objetos. Al estar pre-entrenado, el modelo VGG16 ofrece la ventaja de utilizar el conocimiento adquirido durante el entrenamiento en tareas de visión por computadora sin la necesidad de entrenarlo desde cero, lo que lo convierte en una herramienta eficiente para la extracción de características en aplicaciones de reconocimiento de imágenes.

Inicialmente, se aplicó el algoritmo de agrupamiento K-Means a las características numéricas de los datos, donde se observó un valor relativamente bajo del índice de silueta promedio (0.1156). Esto sugiere que los clústeres no están bien separados, pero tampoco demasiado superpuestos, indicando una estructura de clústeres poco clara en el conjunto de datos.

Posteriormente, se utilizó el modelo VGG16 para extraer características de las imágenes (embeddings), seguido de un análisis de agrupamiento similar. Sin embargo, los resultados tampoco fueron satisfactorios, ya que el índice de silueta promedio sugiere una estructura de clústeres poco clara y los clústeres no están bien definidos.

Una posible explicación de por qué estos métodos de agrupamiento no alcanzaron los resultados esperados radica en la naturaleza heterogénea y compleja de los datos. Las imágenes pueden contener una amplia variedad de características visuales, y las características numéricas

pueden capturar aspectos diversos y no necesariamente relacionados de los objetos representados en las imágenes. Esto puede conducir a una falta de coherencia en los grupos identificados por los algoritmos de agrupamiento.

Además, la alta dimensionalidad de los datos también podría haber influido en los resultados, dificultando la identificación de patrones significativos por parte de los algoritmos de agrupamiento. Aunque se intentó implementar el Análisis de correspondencias múltiples (MCA) para reducir la dimensionalidad, se encontraron limitaciones de procesamiento que impidieron su aplicación. Esto representa una oportunidad de mejora para futuros análisis.

17. Conclusiones de la sección II

En este apartado se realizará una conclusión general de lo mencionado a lo largo de la segunda sección, enfatizando en algunas cuestiones que pueden ser un desafío a la hora de entrenar el modelo de machine learning a futuro.

a. Calidad de los datos

Como se ha mencionado en repetidas ocasiones, el conjunto de imágenes disponible es extremadamente diverso. Este incluye imágenes de variada calidad, algunas borrosas, oscuras, con abundante rastrojo, o con plantas superpuestas. Se optó por conservar estas imágenes por el momento. Esta decisión se debe a que el modelo debe estar preparado para manejar cualquier tipo de imagen que pueda encontrarse en situaciones reales en el futuro. Entrenar el modelo únicamente con imágenes ideales no sería prudente, ya que esto no reflejaría las condiciones reales de operación y podría resultar en un rendimiento deficiente del modelo al enfrentarse a imágenes con estas características en escenarios reales.

Sin embargo, es crucial identificar y evaluar adecuadamente las imágenes con características deficientes. Si al evaluar el modelo se observa que alguna de estas características resulta ser ruidosa o interfiere con el desempeño del modelo, se considerará su eliminación para optimizar su rendimiento y robustez.

b. Tamaño de las imágenes

El tamaño variable de las imágenes es un factor crucial en el entrenamiento de un modelo de machine learning. Algunos modelos solo permiten que las imágenes de entrada tengan el mismo tamaño, o las convierten directamente a un tamaño uniforme. Esto puede requerir un preprocessamiento en el que todas las imágenes se estandaricen a un tamaño específico, lo que puede ser problemático. Al normalizar el tamaño de las imágenes, se suelen perder características importantes que son esenciales para la detección precisa de las plantas. Por lo tanto, elegir un tamaño de imagen de entrada adecuado es esencial.

c. Features de las imágenes

Se han examinado diversas características visuales en las imágenes del dataset que son esenciales para la distinción de las plantas. Para la siguiente fase del proyecto, podría ser beneficioso experimentar con la modificación de ciertos atributos de las imágenes, como la saturación o el contraste. Ajustar estos parámetros puede ser crucial para mejorar la precisión del modelo en la identificación de plantas, especialmente en imágenes que presentan condiciones subóptimas. Por

ejemplo, aclarar imágenes oscuras o reducir la luminosidad en imágenes demasiado claras podría optimizar la visibilidad de detalles importantes y facilitar el proceso de detección.

Además, se ha investigado la dificultad de incorporar información adicional que no está presente visualmente en las imágenes a modelos de machine learning que trabajan con estos datos. Una estrategia innovadora podría ser la sustitución de un canal de color poco utilizado para introducir datos significativos, como el país de procedencia. Este enfoque podría enriquecer los conjuntos de datos con metadatos útiles, potencialmente mejorando la capacidad del modelo para realizar inferencias más precisas y contextualizadas.

Sin embargo, es importante proceder con cautela y realizar pruebas para validar la efectividad de estas técnicas en el contexto específico de su aplicación, asegurándose de que los cambios implementados en las características de la imagen no introduzcan artefactos que puedan confundir al modelo o reducir su precisión general.

Sección III

Desarrollo del modelo y evaluación

18. Introducción

La presente sección se centra en el desarrollo y la posterior validación de modelos de Machine Learning para contar automáticamente las plantas de maíz en campos agrícolas utilizando imágenes tomadas por drones. Así como se desarrolló en las secciones anteriores, un modelo de machine learning para este fin evitaría el conteo manual y sus costos asociados.

19. YOLO (You Only Look Once)

Para lograr entender el desarrollo de la solución es crucial introducir YOLO (You only look once) el algoritmo elegido para abordar el problema. YOLO es una arquitectura de redes neuronales de detección de objetos en tiempo real, destacada por su velocidad. A diferencia de otras arquitecturas, que proponen regiones para luego clasificarlas, YOLO evalúa la imagen completa en una sola pasada, utilizando una cuadrícula para predecir cajas delimitadoras y probabilidades de clases simultáneamente. Esto permite altas velocidades de procesamiento, adecuadas para aplicaciones en tiempo real.

En esta sección se presentarán características fundamentales de la estructura y el funcionamiento de Yolo que ayudarán a entender la aplicación del algoritmo en el caso particular del proyecto.

a. ConvNets y características

Yolo utiliza redes neuronales profundas en donde cada red utiliza múltiples capas convolucionales para extraer características de la imagen. Estas capas, conocidas como ConvNets, aplican diversos filtros a la imagen para detectar patrones visuales tales como bordes, texturas, y formas. A medida que la información avanza a través de la red, las capas más profundas pueden interpretar características más complejas, compuestas por las simples detectadas en las capas anteriores.

b. Entrenamiento en YOLO

i. Predicción de cajas delimitadoras

Una vez que las características se han extraído y procesado a través de las capas convolucionales, YOLO utiliza una cuadrícula superpuesta en la imagen para localizar objetos. Cada celda de esta cuadrícula es responsable de predecir varias cajas delimitadoras. Así, el modelo puede detectar varios objetos y sus ubicaciones en una sola pasada de la red -véase Figura XXII-.

Cada caja delimitadora predicha por YOLO incluye cinco componentes clave: las coordenadas del centro de la caja (x, y), su ancho y alto, y un puntaje de confianza. Este último refleja no sólo si hay un objeto dentro de la caja, sino también qué tan bien la caja se ajusta al objeto.

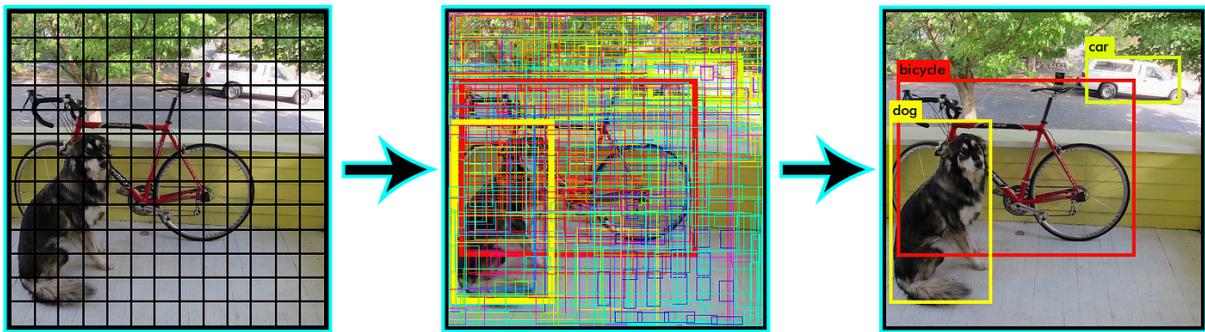


Figura XXII: imagen de referencia de identificaciones realizadas por YOLO.

ii. Filtrado por umbral y supresión no máxima

Finalmente, YOLO aplica un umbral de confianza para descartar cajas con baja probabilidad de contener un objeto significativo. Luego, utiliza una técnica llamada supresión no máxima para eliminar cajas redundantes o superpuestas que predicen el mismo objeto. Esto se hace conservando sólo la caja con la mayor confianza mientras elimina las otras, lo que ayuda a reducir los falsos positivos y mejora la precisión general del modelo (medido por IoU—Intersección over Union).

c. Ventajas de YOLO y Conclusión de la sección

Para concluir la sección es importante justificar por qué se ha elegido este algoritmo para abordar la solución del problema.

- *Velocidad:* En lugar de aplicar la detección objeto por objeto, YOLO evalúa toda la imagen a la vez. Esto elimina la necesidad de pasos separados para proponer regiones y luego clasificar esas regiones. Al reducir la cantidad de pasos y cálculos necesarios, YOLO aumenta enormemente su velocidad.
- *Mejor generalización:* La capacidad de YOLO para ver la imagen en su totalidad durante el entrenamiento y la inferencia le permite generalizar mejor cuando se enfrenta a nuevas escenas o variaciones en la apariencia de los objetos.
- *Soporte de la comunidad:* Al ser YOLO un algoritmo de código abierto, se beneficia de las contribuciones de una amplia comunidad de desarrolladores que constantemente añaden mejoras, correcciones y nuevas funcionalidades. Además, la comunidad alrededor de YOLO es muy activa y ofrece soporte a través de foros, grupos de discusión y plataformas como GitHub. Los usuarios pueden obtener ayuda rápidamente para resolver problemas y compartir experiencias.

Si bien los tres aspectos mencionados son de suma importancia, el factor que resultó determinante en la selección de esta arquitectura fue el último: la necesidad de adoptar un algoritmo que disponga de un amplio soporte y una comunidad extensa. Este requisito es de vital importancia, especialmente cuando se considera que, una vez entregado el modelo a Eiwa, la cual se encuentra en etapas iniciales de desarrollo de su área de ciencia de datos, será indispensable contar con una variedad de recursos accesibles. Estos recursos incluyen videotutoriales y disponibilidad de foros de discusión activos, los cuales son esenciales para facilitar el mantenimiento continuo y la actualización del modelo. Además, una comunidad robusta puede ofrecer apoyo rápido y efectivo ante cualquier desafío técnico que pueda surgir una vez entregado el modelo, asegurando así su longevidad y eficacia.

20. Elección de versión de YOLO.

Así como se mencionó anteriormente, YOLO cuenta con una comunidad que facilita una constante innovación y refinamiento, permitiendo que se mantenga a la vanguardia en tecnología de detección de objetos. Es por esto que YOLO cuenta con diversas versiones, cada una con sus distintas variantes.

Actualmente, YOLO ha desarrollado versiones que van desde la 1 hasta la 8, incorporando variaciones y mejoras en cada iteración. Dentro de las mismas versiones, existen variaciones de tamaño como "small" o "large", diseñadas para equilibrar la precisión y la velocidad según los diferentes requerimientos computacionales y aplicativos.

Para el proyecto en cuestión, se decidió utilizar la versión 5 lanzada en 2020. Tras una investigación sobre las distintas variantes, se seleccionó esta versión por su consolidación en la comunidad, respaldada por una amplia documentación que incluye modelos pre entrenados y tutoriales sobre su uso. Aunque las versiones 6, 7 y 8 son más recientes (2022-2023), su adopción aún es incipiente y se recomienda esperar a que su uso se estabilice. Por estas razones, se optó por la versión 5, demostrando ser más fiable y estable. Además, se inició el proyecto con la variante "small" de YOLOv5, ya que es la más rápida. Tras comprobar su buen desempeño, se decidió continuar con esta versión, dado que se necesitaba un modelo rápido debido a las limitaciones de tiempo.

La posibilidad de utilizar modelos pre entrenados es un gran distintivo de YOLO que fue aprovechado para el desarrollo de todos los modelos que se presentarán a lo largo del informe. A través del transfer learning se partió de pesos predeterminados y se optimizaron los hiper parámetros del modelo para adaptarlos al dataset de entrenamiento, logrando el mejor modelo posible con una ventaja inicial.

De todas las tareas que puede realizar un modelo de YOLO la que se necesita para este caso es object detection. Esto implica reconocer cuando hay objetos de las clases indicadas y crear la caja delimitadora del espacio que ocupan esos objetos. En el caso de este proyecto solo existe una clase, maíz. Por cada imagen el modelo creará un archivo con las coordenadas de las bounding boxes que luego se podrán contar para determinar la cantidad de plantas en la imagen.

21. Preparación del dataset

En el marco del proyecto en curso, se ha adaptado el dataset para asegurar su compatibilidad con YOLO. A continuación, se detalla el proceso de transformación de los datos.

En la sección 10.c *Archivo JSON con metadatos*, se describió el formato en que se encuentran las anotaciones de las imágenes, las cuales están almacenadas en un archivo JSON. Actualmente, dicho archivo contiene las coordenadas de las cajas delimitadoras (bounding boxes) según el siguiente esquema:

- x: Coordenada horizontal del extremo izquierdo de la bounding box sobre el eje X de la imagen.
- y: Coordenada vertical del extremo superior de la bounding box sobre el eje Y de la imagen.
- w: Ancho del rectángulo, extendiéndose hacia la derecha desde el punto x.
- h: Altura del rectángulo, extendiéndose hacia abajo desde el punto y.

Sin embargo, YOLO posee requisitos específicos para la organización y el formato de los archivos de anotaciones. Uno de estos requisitos es que cada imagen en el conjunto de datos debe tener un archivo de anotación de texto (.txt) correspondiente y separado. Luego, requiere un estilo específico de anotaciones diferente al actual, en donde x e y deben ser las coordenadas del centro de la caja delimitadora. Por último, también se requiere que todas las coordenadas estén normalizadas en valores entre 0 y 1 (Li et al., 2018).¹⁷

Por consiguiente, se realizó una reestructuración del archivo en la que se separaron las coordenadas de las bounding boxes de cada imagen en archivos .txt distintos, cada uno con el mismo nombre que la imagen correspondiente. Para convertir las coordenadas desde un sistema donde x e y son las esquinas superiores izquierdas a uno donde representan el centro y todas las coordenadas están normalizadas, se aplicaron los siguientes cálculos de transformación:

- $x_{nuevo} = \frac{x + \frac{w}{2}}{\text{AnchoImagen}}$
- $y_{nuevo} = \frac{y + \frac{h}{2}}{\text{AltoImagen}}$
- $w_{nuevo} = \frac{w}{\text{AnchoImagen}}$
- $h_{nuevo} = \frac{h}{\text{AltoImagen}}$

Con respecto a las imágenes, estas se encontraban ya correctamente organizadas para la adecuada utilización de YOLO.

22. División Train, Validation y Test

Como se evidenció en la sección anterior del proyecto, el conjunto de imágenes posee una amplia variedad de características. Es elemental realizar una correcta división de Train, Validation y Test para que el modelo aprenda de un conjunto de imágenes lo más variado posible y que pueda luego testear esas imágenes. Si no se realiza esta división con atención, podrían incluirse en el conjunto de testeo imágenes con características muy distintas a aquellas con las que se entrenó el modelo, afectando así la capacidad de generalización del mismo.

a. Conjuntos de datos en YOLOv5

Al utilizar YOLOv5 es importante resguardar un porcentaje de los datos tanto para validación como para testeо. Este enfoque se justifica por la metodología de entrenamiento del algoritmo.

Durante el entrenamiento, el conjunto de datos de entrenamiento se utiliza para ajustar los pesos de la red neuronal, optimizando el modelo en función de su capacidad para predecir correctamente las anotaciones de las imágenes. El conjunto de validación desempeña un papel esencial en este proceso: no influye directamente en el ajuste de los pesos, pero se utiliza para evaluar el rendimiento del modelo a medida que se entrena. Esto permite realizar ajustes en los hiperparámetros y prevenir el sobreajuste. Es de suma importancia destacar que los pesos que se seleccionan al final del entrenamiento son aquellos que mostraron la mejor performance sobre el conjunto de validación.

¹⁷ Li, Y., Du, H., Wang, X., Wang, L., Zhang, Y., & Zhao, H. (2018). A review on YOLO object detection. Journal of Signal and Information Processing, 9(4), 139-147. <https://www.scirp.org/journal/paperinformation?paperid=88545>

Dado que los pesos finales se eligen por su eficacia en el conjunto de validación, no sería adecuado utilizar estos mismos datos para pruebas. Hacerlo podría proporcionar una estimación inflada y no realista del rendimiento del modelo. Por esta razón, se reserva un conjunto de testeo, compuesto por datos completamente nuevos para el modelo, que permite evaluar objetivamente cómo se desempeñaría en condiciones reales y con datos que no han influido en las decisiones de entrenamiento y validación. Este enfoque garantiza una evaluación imparcial de la capacidad de generalización del modelo y su efectividad en escenarios no vistos durante el entrenamiento.

Para el entendimiento de las métricas del modelo es importante introducir el concepto de Average Precision. La precisión promedio (AP)¹⁸ resume una curva de precisión-recall (PR) en un solo valor que representa el promedio de todas las precisiones. Generalmente se entiende como la aproximación del área bajo la curva PR. AP varía entre 0 y 1, donde un modelo perfecto tiene puntajes de precisión, recall y AP de 1. El área bajo la curva, donde $p(r)$ es la precisión y el recall r se puede definir como:

$$AP = \int_0^1 p(r)dr$$

A continuación, se presenta la Tabla VI que ilustra las métricas de rendimiento del modelo, evaluadas de manera iterativa durante el proceso de ajuste de pesos en cada iteración de entrenamiento (epochs). En esta tabla se puede monitorear:

Métrica	Explicación
<i>GPU_mem</i>	La cantidad de memoria GPU que se está utilizando actualmente para el entrenamiento del modelo.
<i>box_loss</i>	La pérdida asociada con la predicción de las bounding boxes de los objetos. Una pérdida más baja indica que el modelo está prediciendo mejor las ubicaciones de las cajas delimitadoras.
<i>obj_loss</i>	La pérdida asociada con la predicción de la presencia de objetos en las cajas delimitadoras. Una pérdida más baja indica que el modelo está mejorando en la detección de si hay o no un objeto en una caja dada.
<i>cls_loss</i>	La pérdida asociada con la clasificación de los objetos dentro de las cajas delimitadoras. Una pérdida más baja indica que el modelo está mejorando en la identificación correcta de la clase del objeto. En este caso es 0, ya que hay una sola clase posible a predecir.
<i>Instances</i>	El número total de objetos que el modelo detectó.

¹⁸ Mu Zhu (26 de agosto, 2004) “Recall, Precision and Average Precision”. University of Waterloo. Department of Statistics & Actuarial Science.

P (<i>precision</i>)	La Proporción de verdaderos positivos (objetos correctamente detectados) entre todos los positivos predichos (tanto verdaderos como falsos).
R (<i>Recall</i>):	La proporción de verdaderos positivos entre todos los positivos reales (los que realmente están en la imagen).
<i>mAP50</i>	Promedio de la precisión promedio (AP) calculado con un umbral de IoU de 50%. Una predicción es marcada como correcta en el entrenamiento si la superposición entre la caja predicha y la caja de verdad del terreno es al menos del 50%.

Tabla VI: Métricas monitoreadas en cada época de entrenamiento.

En cada iteración de ajuste de pesos, el modelo se somete a una evaluación utilizando el conjunto de validación. Esta metodología permite monitorear cómo las diferentes configuraciones de pesos influyen en el rendimiento del modelo, asegurando que cada ajuste contribuya de manera efectiva a mejorar la precisión y la eficacia del modelo frente a los datos de validación.

Durante el proyecto, se mantuvo un número fijo de 50 epochs para el entrenamiento. Esta cantidad se seleccionó tras observar que, aproximadamente después de 50 epochs, las métricas del modelo dejaban de mejorar significativamente. En otras palabras, el entrenamiento alcanzaba un "punto de saturación" donde continuar aumentando los epochs no justificaba el uso adicional de recursos computacionales, ya que las mejoras en las métricas eran mínimas. Esto se puede observar en la Figura XXIII, donde las métricas de rendimiento parecen estabilizarse y no presentan diferencias notables.

Mantener este balance entre la precisión del modelo y la eficiencia computacional es crucial para asegurar que el modelo no solo sea preciso, sino también eficiente en términos de recursos utilizados durante el entrenamiento y la evaluación.

Epoch	GPU_mem	box_loss	obj_loss	cls_loss	Instances	Size
45/49	4.68G	0.0983	0.1399	0	446	640: 100% 132/132 [01:49<00:00, 1.20it/s]
	Class Images	Instances		P	R	mAP50 mAP50-95: 100% 9/9 [00:04<00:00, 1.89it/s]
	all	198	4996	0.674	0.587	0.58 0.193
46/49	4.68G	0.09858	0.1387	0	594	640: 100% 132/132 [01:49<00:00, 1.20it/s]
	Class Images	Instances		P	R	mAP50 mAP50-95: 100% 9/9 [00:04<00:00, 2.08it/s]
	all	198	4996	0.671	0.589	0.583 0.194
47/49	4.68G	0.09765	0.1364	0	641	640: 100% 132/132 [01:48<00:00, 1.21it/s]
	Class Images	Instances		P	R	mAP50 mAP50-95: 100% 9/9 [00:04<00:00, 2.00it/s]
	all	198	4996	0.665	0.586	0.575 0.193
48/49	4.68G	0.09773	0.1396	0	493	640: 100% 132/132 [01:48<00:00, 1.22it/s]
	Class Images	Instances		P	R	mAP50 mAP50-95: 100% 9/9 [00:06<00:00, 1.47it/s]
	all	198	4996	0.677	0.588	0.582 0.196
49/49	4.68G	0.09878	0.1382	0	499	640: 100% 132/132 [01:48<00:00, 1.22it/s]
	Class Images	Instances		P	R	mAP50 mAP50-95: 100% 9/9 [00:06<00:00, 1.41it/s]
	all	198	4996	0.674	0.59	0.58 0.196

Figura XXIII: Imagen de ejemplo de la supervisión del entrenamiento.

b. División Equitativa del Dataset

Para la división del conjunto de datos se decidió asignar un 70% de los datos al entrenamiento, un 10% a la validación y el 20% restante a testeo -véase Tabla VII-.

Conjunto	Porcentaje del total de imágenes	Cantidad de imágenes por conjunto
Entrenamiento	70%	1.385
Validación	10%	198
Testeo	20%	395

Tabla VII: Porcentaje de imágenes en cada conjunto.

Con el objetivo de asegurar una distribución equitativa de aquella imágenes con características especiales, se utilizó el archivo 'Layout Tags'. Este archivo, presentado con detalle en el apartado 10.b *Etiquetas (labels) con metadatos*, cataloga cada imagen junto con cualquier característica no ideal que pueda tener, como ser oscuridad o borrosidad.

Es crucial recordar que una misma imagen puede tener múltiples etiquetas, lo que dificulta el proceso de división proporcional. Por ejemplo, una imagen puede ser simultáneamente oscura y borrosa. Este factor añade una capa de complejidad al intentar mantener una proporción equitativa en la distribución de las imágenes entre los conjuntos de entrenamiento, validación y prueba, ya que no se puede dividir simplemente basándose en una sola característica sin considerar las demás.

Para abordar la problemática se realizó un recuento de cuántas veces cada etiqueta aparecía vinculada a una imagen, permitiendo identificar las etiquetas menos comunes, que se asumieron como las más explicativas debido a su rareza. Una vez que las etiquetas fueron ordenadas por relevancia, se procedió a asignar a cada imagen, de aquellas que poseían etiquetas, la etiqueta más relevante. Este método asume que las etiquetas más inusuales proporcionan una mayor singularidad y relevancia para el entrenamiento del modelo.

$$\text{Relevancia de la etiqueta} = \frac{\text{Cantidad de fotos con la etiqueta}}{\text{total de fotos etiquetadas}}$$

En la Tabla VIII se presenta la frecuencia de aparición de cada etiqueta en las imágenes. Se observa que, entre las imágenes que poseen etiquetas, la etiqueta "Imagen clara" es la más inusual, mientras que las imágenes con rastrojo son las más frecuentes. Por lo tanto, si una imagen posee ambas etiquetas, es decir, es clara y tiene rastrojo, se priorizará la etiqueta "Imagen clara" como identificador principal.

Número de Etiqueta	Etiqueta	Frecuencia
9	Imagen clara	160
8	Imagen oscura	330
7	Imagen borrosa	330
1	Suelo oscuro	430
3	Alta densidad de plantas	436

10	Plantas crecidas	440
6	Maleza	686
2	Suelo rojo	705
11	Plantas dobles	814
4	Presencia de Sombra	1057
5	Rastrojo	1250

Tabla VIII: Frecuencia de Imágenes por Etiqueta.

A continuación se ejemplifica con una situación concreta. La Figura XXIV está etiquetada con los siguientes identificadores: 2, 3, 5, 6, 8 y 10. Sin embargo, para la correcta división del conjunto de datos, esta imagen retendrá únicamente la etiqueta 2, considerada la más significativa entre todas las presentes.



Figura XXIV: Imágen de ejemplo con múltiples etiquetas

Una vez resuelto el problema de la múltiple etiquetación de las imágenes, se procedió a dividir el conjunto de datos en grupos de entrenamiento, validación y prueba utilizando la biblioteca scikit-learn. Se utilizó el hiperparámetro *stratify = priority_tags* en *train_test_split*, configurado para considerar los tags más raros asignados previamente a cada imagen. El parámetro *stratify* garantiza que la proporción de cada etiqueta sea consistentemente representativa en los tres conjuntos.

En la Figura XXV se observa la frecuencia de etiquetas prevaleciente en cada conjunto de datos. Se evidencia cómo cada uno de ellos no solo contiene todas las etiquetas, sino que también lo hace en una proporción similar.

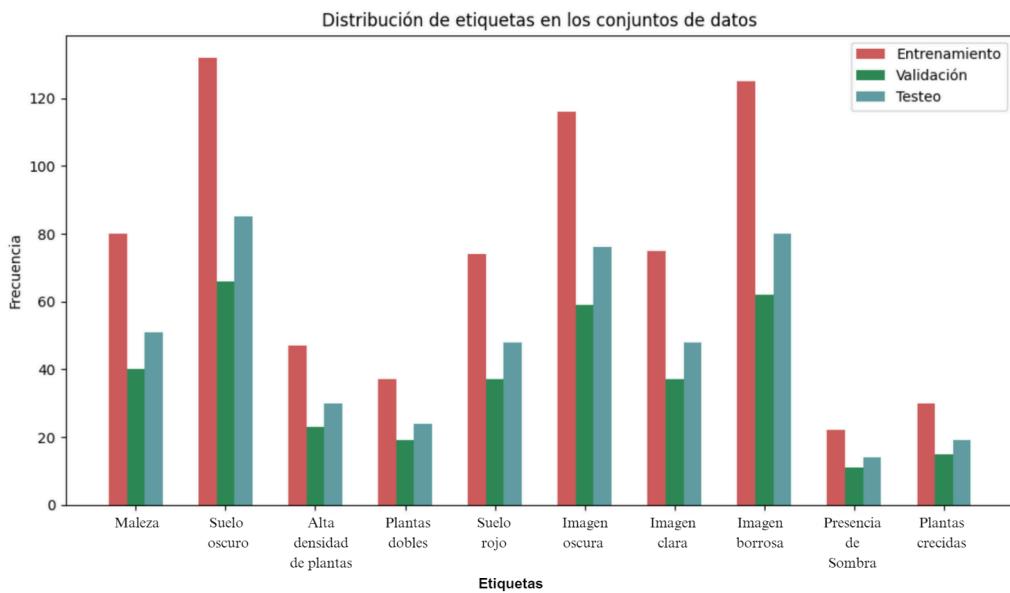


Figura XXV: Frecuencia de aparición de etiquetas por cada conjunto de datos

c. Conclusiones de la sección

A modo de conclusión, la adecuada partición de los datos en conjuntos de entrenamiento, validación y prueba es crucial para el éxito del modelo YOLOv5. Este enfoque no solo asegura que el modelo se entrene con una amplia variedad de imágenes, sino que también permite una evaluación objetiva y confiable de su desempeño. Utilizar correctamente estos conjuntos evita el sobreajuste y garantiza que los ajustes en los pesos del modelo durante el entrenamiento se basen en datos que reflejen un rendimiento genuino y no sesgado.

23. Exploración de modelos

En esta sección, se exploran diversas configuraciones del modelo YOLOv5, con el objetivo de optimizar su rendimiento para adaptarlo de manera efectiva a las necesidades específicas del proyecto. Esta experimentación detallada es esencial para identificar la configuración que mejor se desempeña en el recall y la precisión. A través de una serie de pruebas y ajustes de hiperparámetros, se buscó desarrollar una solución robusta y confiable que se alinee con los objetivos del proyecto.

a. Disminución del dataset

Al iniciar la exploración de modelos, se identificó que realizar pruebas con el conjunto de datos completo representaba un desafío debido al alto costo y tiempo computacional requeridos, limitados por nuestros recursos y plazos. Por este motivo, se optó por reducir el conjunto de datos para las pruebas preliminares.

Para asegurar la representatividad de esta versión reducida del conjunto de datos, se aplicó la misma lógica utilizada en la partición original (punto 22.b *División Equitativa del Dataset*), buscando conservar una amplia variedad de etiquetas y características en los conjuntos de entrenamiento, validación y prueba. Los conjuntos de datos reducidos incluían 400 imágenes para entrenamiento, 80 para testeo y 40 para validación. Las imágenes para cada conjunto reducido fueron cuidadosamente

seleccionadas de sus respectivos conjuntos completos, manteniendo la integridad de los datos sin mezclar imágenes entre los tres conjuntos.

b. Hiperparámetros en YOLOv5

Antes de abocarse a la exploración de los distintos modelos, es necesario aclarar los tipos de hiperparámetros que existen en YOLOv5. El proceso de configuración y ajuste del modelo se maneja a través de varios tipos de hiperparámetros, que se pueden categorizar principalmente en tres grupos: hiperparámetros utilizados en el comando de entrenamiento, hiperparámetros internos del modelo e hiperparámetros usados en la detección (Ultralytics, 2024)¹⁹.

i. Hiperparámetros en el Comando de Entrenamiento

Estos hiperparámetros se especifican directamente en la línea de comando al iniciar un proceso de entrenamiento. Algunos de ellos pueden ser observados en la Tabla IX.

Hiperparámetro	Explicación	Valor Default
<i>batch-size</i>	Determina cuántas imágenes se procesan antes de actualizar los pesos del modelo. Un batch-size más grande puede mejorar la estabilidad del entrenamiento pero requiere más memoria. Un tamaño más pequeño puede hacer que el entrenamiento sea más ruidoso, pero puede ayudar a escapar de mínimos locales durante la optimización.	16
<i>epochs</i>	Indica cuántas veces el modelo verá todo el conjunto de datos de entrenamiento. Más epochs pueden permitir un aprendizaje más completo, pero también aumenta el riesgo de sobreajuste si no se monitorea y ajusta adecuadamente.	300
<i>img-size</i>	Ajusta la resolución de las imágenes que se alimentan al modelo. Imágenes más grandes pueden capturar más detalles, pero también son más exigentes computacionalmente.	640
<i>weights</i>	Define los pesos iniciales con los que comenzará el entrenamiento. Se pueden utilizar pesos de un modelo previamente entrenado para acelerar la convergencia y mejorar la precisión inicial, especialmente cuando se dispone de un conjunto de datos limitado.	N/A

Tabla IX: Hiperparametros utilizados en el entrenamiento.

¹⁹ Ultralytics. (2024). Configuration - Ultralytics YOLO Docs. Recuperado de <https://docs.ultralytics.com/es/usage/cfg/#export-settings>

ii. Hiperparámetros Internos del Modelo

Estos hiperparámetros están integrados dentro de los archivos de configuración de YOLOv5 y permiten un ajuste más granular de la arquitectura de la red. Algunos de ellos pueden ser observados en la Tabla X.

Hiperparámetro	Explicación	Valor Default
<i>lr0 (Learning Rate Initial)</i>	La tasa de aprendizaje inicial, determina la magnitud de los ajustes realizados a los pesos del modelo en cada paso de entrenamiento.	0,01
<i>box</i>	Este parámetro pondera la pérdida asociada con las dimensiones de las bounding boxes durante el entrenamiento. El valor de box influye en la importancia que el modelo da a los errores en la predicción de las dimensiones de las cajas delimitadoras respecto a otras componentes de la pérdida, como la clasificación de objetos o la confianza de la detección. Un valor más alto significa que el modelo pondrá más énfasis en acertar el tamaño y la posición exactos de las cajas durante el entrenamiento.	0,05
<i>Iou_t (IOU Threshold)</i>	El umbral de Intersección sobre la Unión (IoU) es utilizado durante el entrenamiento para determinar qué tan bien deben coincidir las bounding boxes predichas con las cajas delimitadoras verdaderas para que se consideren predicciones correctas.	0,2
<i>hsv_h (Hue)</i>	Modifica el tono de la imagen. Un cambio en el tono puede simular condiciones de iluminación diferentes o variaciones en la apariencia del objeto.	0,015
<i>degrees</i>	Rota las imágenes un número de grados especificado.	0

Tabla X: Hiperparametros de Yolov5

Estos son solo algunos ejemplos de la amplia gama de hiperparámetros disponibles en YOLOv5. Además, existen hiperparámetros que permiten voltear las imágenes vertical o horizontalmente, mejorando la robustez del modelo ante variaciones en la orientación de los objetos. También cuenta con técnicas avanzadas de data augmentation como mosaic, mixup y copy-paste que permiten la superposición de objetos o cambios en el contexto, lo cual es esencial para entrenar modelos que operen eficientemente en entornos dinámicos y diversos. Sin embargo, es crucial manejar estas técnicas con precaución para evitar la introducción de ruido innecesario que podría comprometer la precisión del modelo.

iii. Hiperparámetros de Detección

Estos hiperparámetros se utilizan durante la fase de detección. Algunos de ellos pueden ser observados en la Tabla XI.

Hiperparámetro	Explicación	Valor Default
<i>conf-thres</i>	Es el umbral de confianza para decidir si una predicción se considera un objeto. Predicciones con una confianza por debajo de este umbral se descartan. Un umbral más alto reduce los falsos positivos pero puede aumentar los falsos negativos.	0,25
<i>iou-thres</i>	Este es el umbral para el IoU en el proceso de Non-Maximum Suppression (NMS), que se utiliza para resolver conflictos entre cajas delimitadoras que se solapan. Un umbral más bajo puede resultar en más cajas delimitadoras por objeto, mientras que uno más alto puede fusionar cajas de objetos adyacentes.	0,7

Tabla XI: Hiperparametros utilizados para la detección.

c. Implementación de los modelos

Se realizaron pruebas con distintos modelos, variando los hiperparámetros mencionados anteriormente. Se intentó realizar una optimización de hiperparámetros automática pero, dado que trabajar con imágenes implica una alta carga computacional para evaluar los modelos, no fue posible debido a las limitaciones de computación. Este desafío subraya la necesidad de un profundo entendimiento de cada hiperparámetro para realizar ajustes conscientes y orientados a mejorar el rendimiento del modelo.

Para mantener el orden en la experimentación, se planteó un esquema que se representa en la Figura XXVI. En primer lugar, se elegirán los hiperparámetros del comando de entrenamiento. Luego, utilizando los parámetros que mejor hayan funcionado, se seleccionarán los hiperparámetros internos del modelo. Finalmente, se determinarán los hiperparámetros de detección. Como resultado, se obtendrá el modelo resultante que englobará los hiperparámetros que hayan demostrado ser más efectivos en cada paso.



Figura XXVI: Diagrama de exploración de hipermáparámetros

Es importante tener en cuenta que se estableció un umbral de IoU del 30% para la medición de resultados en este proyecto. Esto significa que se considera una predicción correcta cuando la caja predicha coincide en un 30% o más con la caja real. Este valor se eligió tras probar distintos valores, y se determinó que era el más adecuado. Esta decisión prioriza la detección de la presencia de una planta sobre la precisión exacta de su ubicación, ya que el objetivo principal es el conteo total de plantas. Además, este enfoque permite manejar variaciones en las anotaciones.

Todas las pruebas y detecciones realizadas a lo largo del proyecto se han llevado a cabo bajo este ajuste de umbral, asegurando consistencia en la evaluación y permitiendo un balance adecuado entre precisión y flexibilidad en las predicciones.

i. Variación de hiperparámetros del comando de entrenamiento

Con respecto a la variación de hiperparámetros del comando de entrenamiento, se pusieron a prueba dos configuraciones principales de modelos. La primera configuración utilizó un batch-size de 10 y un img-size de 640, mientras que la segunda empleó un batch-size de 16 y un img-size de 832. Como se explicó en la sección 22.a *Conjuntos de datos en YOLOv5*, la cantidad de epochs utilizados fue de 50.

Configuración	Batch Size	Img Size	Epochs
1	10	640	50
2	16	832	50

Tabla XII: variación de hiperparámetros en el comando de entrenamiento

Se seleccionaron estos valores para explorar el equilibrio entre velocidad de procesamiento y capacidad de detalle. En el primer modelo, el tamaño de imagen más pequeño permite un procesamiento más rápido, aunque con menor atención a los detalles finos. Sin embargo, al usar un batch-size más pequeño, se intenta compensar parcialmente esta limitación, permitiendo un enfoque más detallado en el conjunto más reducido de imágenes procesadas por lote. Por otro lado, el segundo modelo, con un img-size mayor, está diseñado para captar más detalles, pero el incremento en batch-size facilita un procesamiento más eficiente de un mayor número de imágenes, aunque potencialmente con menos detalle por imagen individual en comparación con un batch-size más pequeño. Para la evaluación de los modelos, véase la Tabla I.

Modelo	Precision	Recall	F1-Score
Modelo 1	0,6623	0,7404	0,6992
Modelo 2	0,6034	0,7309	0,6587

Tabla XIII: Comparativa de performance de Modelo 1 y Modelo 2.

Es evidente que el Modelo 1 supera al Modelo 2 en términos de precisión, recall y F1-Score. Esto indica que el Modelo 1 es más efectivo en la identificación precisa de instancias positivas. Por lo tanto, se ha decidido mantener el Modelo 1 debido a su superior desempeño en estas métricas clave.

ii. Variación de Hiperparámetros Internos del Modelo

Ya se decidió que los hiperparámetros del comando de entrenamiento serán: *batchsize* = 10 e *imagesize* = 640. En este apartado se pondrán a prueba distintos valores para los hiperparámetros internos del modelo.

Al analizar las imágenes de las bounding boxes predichas, se identificó una debilidad en el modelo: el exceso de bounding boxes. Actualmente, el modelo coloca múltiples bounding boxes por objeto, resultando en la detección repetida de una misma planta -véase Figura XXVII-.



Figura XXVII: Ejemplos de múltiples bounding boxes para una misma planta

Para mejorar la detección se probó ajustar los hiperparámetros que influyen en la detección directa de los objetos, los cuales pueden observarse en la Tabla XIV.

Hiperparámetro	Valor Default	Valor Modificado	Resultado Esperado
cls	0,5	0	Este hiperparámetro pondera la pérdida asociada con la clasificación de clases durante el entrenamiento. Al modificar este valor a 0, se elimina la contribución de la clasificación de clases en la función de pérdida, lo cual es útil cuando no se requiere diferenciar entre varias clases de objetos. Esto simplifica el modelo al enfocarse exclusivamente en la localización de objetos.
box	0,05	0,08	Este hiperparámetro afecta cuánto se penaliza el error en las predicciones de las dimensiones de las bounding boxes. Al aumentar este valor, se da mayor importancia a la precisión de las bounding boxes, lo que es crucial cuando el modelo genera múltiples detecciones para un mismo objeto.
Iou_t	0,2	0,25	Este umbral determina qué tan precisas deben ser las predicciones de las bounding boxes para considerarse acertadas. Al incrementar el iou_t de 0,2 a 0,25, se establece un criterio más estricto para la coincidencia entre las bounding boxes predichas y las reales.

Tabla XIV: Comparativa de performance de Modelo 1 y Modelo 2.

Una vez modificados estos valores se entrenó un nuevo modelo, en donde se han obtenido los resultados presentes en la Tabla XV..

Modelo	Precision	Recall	F1-Score
Modelo con hiperparámetros internos por default	0,6623	0,7404	0,6992
Modelo modificando hiperparámetros internos	0,7403	0,7608	0,7504

Tabla XV: Resultados obtenidos producto del nuevo modelo.

Se ha registrado una mejora tanto en la precisión como en el recall en comparación con el último modelo, lo que indica que el ajuste de hiperparámetros ha sido efectivo. A continuación, se presentan las mismas tres imágenes que se mostraron en el apartado anterior -véase Figura XXVIII-. En las imágenes superiores, se observan las predicciones realizadas antes del ajuste de hiperparámetros y, en las inferiores, después de dicho ajuste. Se aprecia una notable reducción en la cantidad de bounding boxes superpuestas, aunque aún persisten algunas.

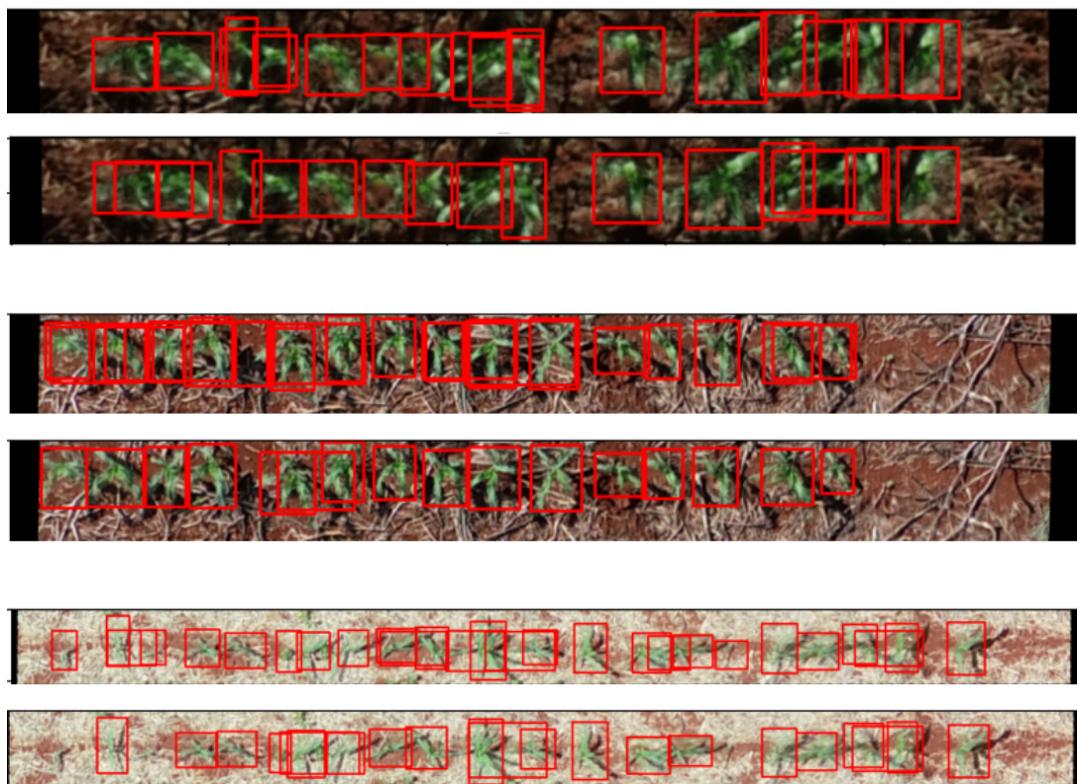


Figura XXVIII: Ajuste de bounding boxes superpuestas

Aunque se ha mejorado la detección redundante de objetos, también se ha observado que, mientras anteriormente la mayoría de las plantas eran detectadas, ahora algunas quedan sin detectar. En la Figura XXIX, se presenta un ejemplo en el que la imagen superior muestra las detecciones realizadas por el modelo antes del ajuste de hiperparámetros, y la imagen inferior muestra las detecciones después de este ajuste.

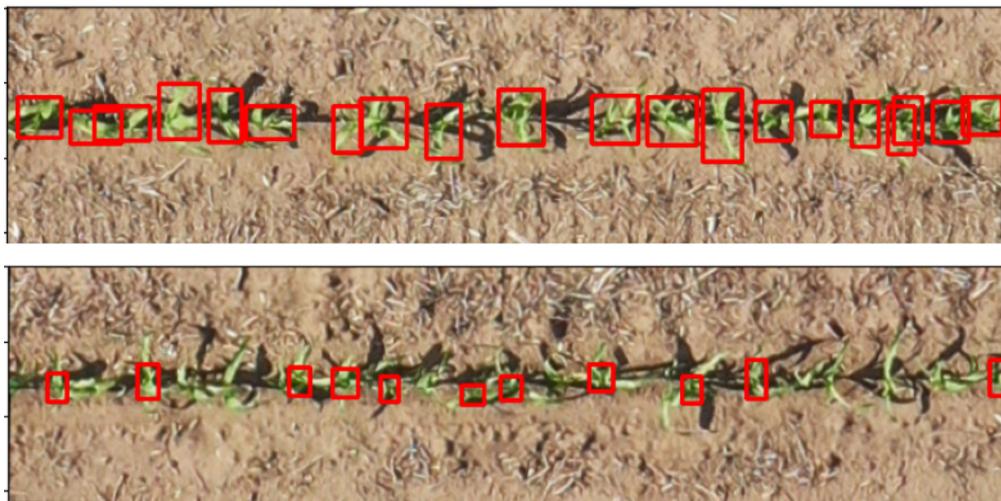


Figura XXIX: Imagen antes y después de ajuste de hiperparámetros.

Debido a estos resultados, no se descarta ninguno de los modelos; tanto el configurado con hiperparámetros por defecto como el ajustado seguirán siendo considerados. Además, como se presentó en el punto 22.e *Hiperparámetros de Detección*, se podría resolver el problema de superposición de cajas mediante la variación de este último conjunto de hiperparámetros. Por lo tanto, se esperará a la próxima instancia para decidir entre los dos modelos abordados en este apartado.

iii. Variación de Hiperparámetros de Detección

Por último, en la fase de detección de objetos, es posible modificar los hiperparámetros de detección para optimizar el rendimiento del modelo.

Dado que la principal debilidad del modelo es la excesiva superposición de las cajas delimitadoras, se decidió experimentar con el ajuste del hiperparámetro iou-thres. Este es crucial, ya que define el IoU para determinar cuándo las cajas se superponen demasiado y deben ser fusionadas o descartadas. Ajustar este valor permite un control más fino sobre las cajas que se intersectan, reduciendo las detecciones redundantes y mejorando la claridad en la detección de objetos. Al aumentar este umbral, el modelo se vuelve más estricto en aceptar superposiciones, resultando en menos cajas superpuestas y una representación más clara de los objetos detectados.

Dado que la superposición excesiva de cajas delimitadoras puede mitigarse ajustando el hiperparámetro iou-thres, se optó por mantener los valores base de los hiperparámetros internos del modelo sin modificaciones adicionales. Esta decisión se tomó tras observar que los ajustes en los hiperparámetros internos resultaban en que el modelo dejara de detectar ciertas plantas, un problema del cual no es posible recuperarse. Por el contrario, es preferible permitir que el modelo genere múltiples bounding boxes para la misma planta, ya que este comportamiento puede corregirse en esta etapa del proceso. Al mantener el iou-thres en un nivel que permita cierta superposición, se asegura una detección más confiable y completa, evitando perder detecciones cruciales, mientras que las cajas delimitadoras redundantes pueden ser filtradas posteriormente. Se experimentó con cuatro valores de iou-thres: 0,5, 0,4, 0,3 y 0,2.

Detección	Precision	Recall	F1 - Score
Iou-thres = 0,5	0,6623	0,7404	0,6992
Iou-thres = 0,4	0,7353	0,7347	0,7350
Iou-thres = 0,3	0,7868	0,7266	0,7555
Iou-thres = 0,2	0,8002	0,7201	0,7580

Tabla XVI: Resultados obtenidos de las métricas de referencia en función a distintos valores de iou-thres

El modelo que mostró los mejores resultados fue el ajustado con un IoU de 0,2. Esto significa que cuando el modelo predice dos cajas que se superponen más de un 20%, conservará solo una de ellas, específicamente la que tenga una mayor probabilidad de contener una planta. De este modo, el modelo evita devolver múltiples cajas con más de un 20% de superposición, lo que mejora la precisión y reduce la redundancia en las detecciones.

Es crucial no confundir el IoU de detección y el IoU de evaluación. El IoU de detección, ajustado a 0,2, se emplea durante la predicción para decidir cuándo conservar o descartar una caja predicha. Por otro lado, el IoU de evaluación, fijado en 0,3, se utiliza para medir las métricas del modelo (definido en el apartado 23.c *Implementación de los modelos*). Estas dos configuraciones de IoU juegan roles diferentes en el proceso de detección y evaluación.

En la Figura XXX a continuación, se presenta una comparativa de imágenes antes y después de aplicar este hiperparámetro. Las imágenes superiores muestran la situación previa, y las inferiores reflejan la situación actual tras el ajuste.

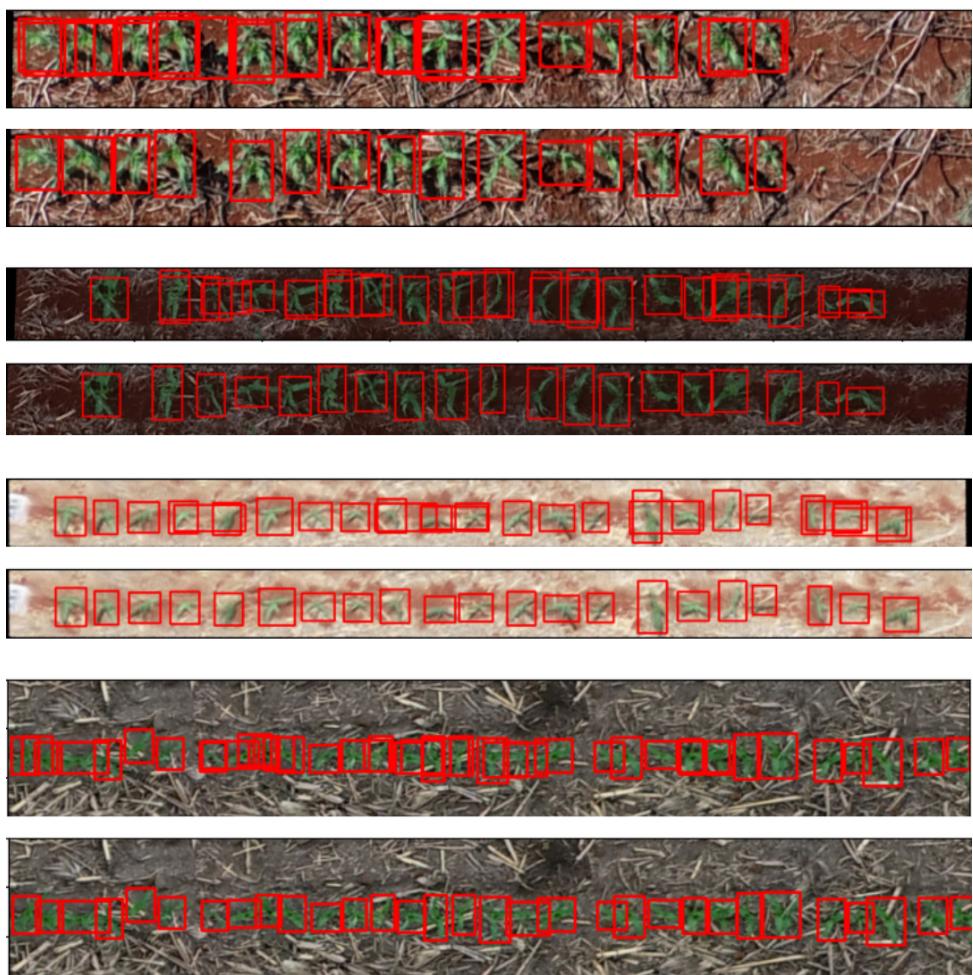


Figura XXX: Imágenes antes y después de aplicar $IoU-thres=0,2$

Es posible observar que ya no hay múltiples cajas detectando un mismo objeto; ahora, cada caja delimitadora identifica correctamente una planta individual. Con esto, se ha superado la mayor debilidad identificada en el modelo hasta el momento.

iv. Conclusiones de la sección

Este capítulo proporciona una visión detallada sobre cómo manejar efectivamente la exploración de modelos en un contexto de recursos limitados y alta demanda computacional. A través de la disminución estratégica del dataset y un enfoque en la configuración de hiperparámetros, se han podido realizar ajustes significativos que impactan directamente en el desempeño del modelo de detección de objetos.

Los resultados de las pruebas demuestran que ajustes cuidadosos en los hiperparámetros no solo pueden mejorar la precisión, sino también reducir la redundancia en las detecciones, un problema común en modelos de detección complejos. Al finalizar este proceso, se ha seleccionado un conjunto de configuraciones que proporcionan un equilibrio adecuado entre precisión y recall. Este enfoque subraya la importancia de una comprensión profunda de los hiperparámetros y la necesidad de adaptabilidad y precisión en el desarrollo de soluciones de computer vision.

Tipo de Hiperparámetro	Comando		Internos	Detección
Hiperparámetro	Batch Size	Img Size	-	IOU
Configuración elegida	10	640	default	0,2

Tabla XVII: Configuración elegida para los hiperparámetros del modelo.

24. Implementación del modelo

Manteniendo las configuraciones óptimas identificadas anteriormente, se procede a entrenar el modelo usando el conjunto completo de datos de entrenamiento. Además, la validación y las pruebas también se realizan utilizando la totalidad de los conjuntos correspondientes.

Luego de realizar el testeо, se observaron las imágenes en donde el modelo tuvo un peor rendimiento. Durante este proceso se detectaron algunos casos interesantes que vale la pena mencionar.

a. Variaciones en etiquetado manual

A continuación -véase Figura XXXI-, se muestra una comparativa de imágenes con sus respectivas bounding boxes. La imagen superior ha sido etiquetada manualmente por verificadores contratados por Eiwa. Por otro lado, la foto que se ubica en la parte inferior muestra la detección sobre la misma imagen realizada por el modelo.

En esta imagen en particular, el modelo registró una precisión del 0%, lo que podría sugerir inicialmente que no detectó ninguna planta. Sin embargo, es evidente que todas las plantas fueron correctamente identificadas. La razón de este aparente desajuste es que las bounding boxes predichas no alcanzaron el umbral mínimo de evaluación del 30% de coincidencia con las cajas reales. Este fenómeno se debe a que las cajas etiquetadas manualmente fueron realizadas de manera muy pequeña, lo cual afectó la evaluación de la precisión del modelo bajo los criterios establecidos.

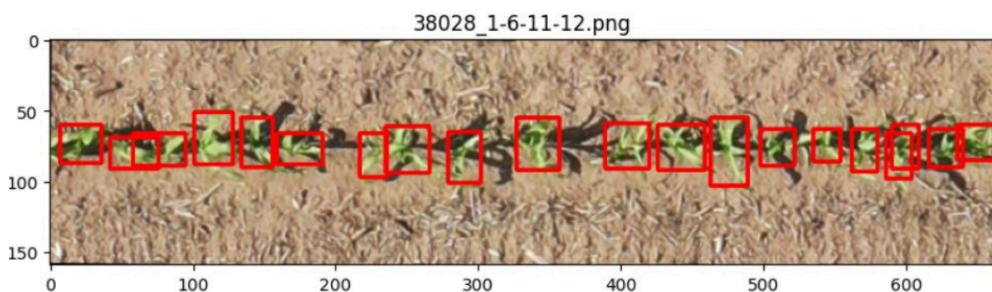
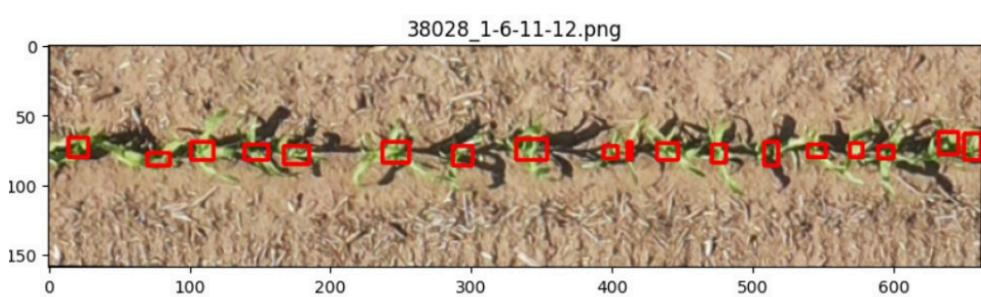


Figura XXXI: Bounding boxes mal etiquetadas

b. Análisis de resultados en imágenes

A continuación se presentan dos imágenes -véase Figura XXXII y Figura XXXIII- que demuestran la capacidad del modelo para detectar y distinguir plantas en condiciones de visibilidad desafiantes. Por un lado, en la Figura XXXII, tomada en un entorno oscuro, muestra cómo el modelo sigue siendo capaz de identificar las plantas. Esto indica que el modelo ha sido eficazmente entrenado para manejar situaciones de baja luminosidad, donde la distinción visual entre objetos puede ser difícil. Por otro lado, en la Figura XXXIII, se presenta un grado de borrosidad que podría complicar la identificación precisa de objetos. A pesar de esto, el modelo logra detectar la mayoría de las plantas, como se evidencia por las cajas delimitadoras rojas.



Figura XXXII: Predicción del modelo en una imagen oscura



Figura XXXIII: Predicción del modelo en una imagen borrosa.

Otro caso que ha resultado interesante a destacar es el reconocimiento de una planta por su morfología, así como es visualizado en la Figura XXXIV. En la misma, es posible observar que en la esquina izquierda detecta una planta. La misma fue detectada no por su forma visible, sino por la presencia de su sombra. Aunque esta detección no coincide con las bounding boxes originales y por tanto no se considera acertada, sugiere que el modelo ha desarrollado una capacidad notable para interpretar y utilizar las sombras como una característica relevante en la identificación de objetos.



Figura XXXIV: Detección de planta por medio de su sombra.

25. Experimentación de resultados

a. Resultados Preliminares

Los resultados del modelo para todo el conjunto de testeо fueron los siguientes:

Precision	Recall	F1-Score
0,8414	0,8301	0,8357

Tabla XVII: Resultado del modelo para el conjunto de testeо

Es posible notar que los resultados obtenidos se encuentran dentro del escenario moderado planteado en la sección 5.d *Costos y Escenarios* (para llegar al escenario optimista era necesario recall 0,9 y precisión 0,85). En consecuencia deberá incluirse un conteo manual del 25% de las imágenes, las que correspondan con el puntaje de confianza más bajo.

Con el objetivo de evaluar la performance del modelo en distintos escenarios, además de las métricas ya discutidas, se introdujo la métrica "Confidence score" o puntaje de confianza:

$$- \text{ Confianza} = \text{Probabilidad } \Pr(\text{Objeto}) \times \text{IOU}$$

Dónde Probabilidad $\Pr(\text{Objeto})$ es la probabilidad de que Bounding Box contenga un objeto, basada en la predicción del modelo.

El puntaje de confianza es una métrica que provee YOLO y útil para filtrar las detecciones generadas por el modelo. Un puntaje de confianza alto indica que el modelo está seguro de que la detección es precisa, mientras que un puntaje bajo indica que el modelo no está tan seguro. Utilizando este puntaje, se pueden establecer umbrales de confianza para decidir cuáles detecciones aceptar y cuáles descartar. En este caso se optó por descartar las predicciones con un puntaje de confianza menor a 0,25. El intervalo para la media de la confianza, tomando los resultados de testeo del modelo seleccionado y tomando un nivel de significación de $\alpha = 0,05$ es el siguiente:

$$\begin{aligned} (A \leq \underline{x} \leq B) &= 1 - \alpha \\ (-z_{0,975} \leq \frac{\bar{x} - \underline{x}}{\frac{\sigma}{\sqrt{n}}} \leq z_{0,975}) & \\ (-z_{0,975} * \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \underline{x} \leq z_{0,975} * \frac{\sigma}{\sqrt{n}}) & \\ (\bar{x} - z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \bar{x} + z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}) & \\ (0,506 - 1,96 * \frac{0,0013}{\sqrt{9966}} \leq \bar{x} \leq 0,506 + 1,96 * \frac{0,0013}{\sqrt{9966}}) & \\ (0,5042 \leq \bar{x} \leq 0,5094) & \end{aligned}$$

Donde:

σ : Desvío de la confianza estimada por el modelo para la bounding box identificada.

n : Cantidad total de bounding boxes identificadas por el modelo para el conjunto de testeo.

\bar{x} : Media de la confianza estimada por el modelo para la bounding box identificada.

De esta manera, se especifica que con un 95% de probabilidad la media de la confianza calculada por el modelo se encontrará entre 0,5042 y 0,5094. El intervalo de confianza establece la probabilidad de que el verdadero valor medio de la clase se encuentre dentro de este rango.

Si bien la cantidad de bounding boxes del conjunto de test es amplio, el intervalo estrecho sugiere que las predicciones del modelo YOLO son consistentes en términos de confianza. La media de confianza calculada cae en un rango muy específico y ajustado, lo que indica que las predicciones tienden a ser precisas y confiables en términos de estimaciones de confianza.

b. Resultados por segmento

Previo a la implementación del modelo, existía la hipótesis de que el modelo podría tener un rendimiento deficiente con las fotos de Argentina, dado que estas son notablemente más anchas que las de otros países. Por lo tanto, se decidió analizar los resultados de las pruebas según el país de origen de cada imagen -véase tabla VI-.

País	F1-Score	Recall	Precision	Confidence score	Cantidad de fotos
Argentina	0,80	0,79	0,81	0,46	95
Brazil	0,93	0,95	0,92	0,59	171
USA	0,80	0,78	0,81	0,43	130

Tabla XVIII: Métricas obtenidas para cada país en el conjunto de testeo

A pesar de que el modelo tuvo mejores resultados en Brasil que en el resto de los países, en Argentina el resultado tanto de precisión como de recall fue mejor que el de Estados Unidos, por lo que se puede afirmar que el ancho de las fotos no generaron un problema para el modelo.

Se analizó, por otra parte, el puntaje de confianza del modelo en cada uno de los países, por estación del año -véase Figura XXXV- pero no se llegó a la conclusión de que exista una estación en la que el modelo funcione mejor que en el resto, sino que depende de cada país. Las fotos de Brasil obtuvieron un confidence score promedio mayor al del resto de los países, llegando a su mejor rendimiento en otoño. Luego Argentina, cuyo puntaje de confianza fue inferior al de Brasil y superior al de USA en todas las estaciones, obtuvo el mejor confidence score en primavera. Por último, el mejor resultado respecto al puntaje de confianza para Estados Unidos fue en Verano.

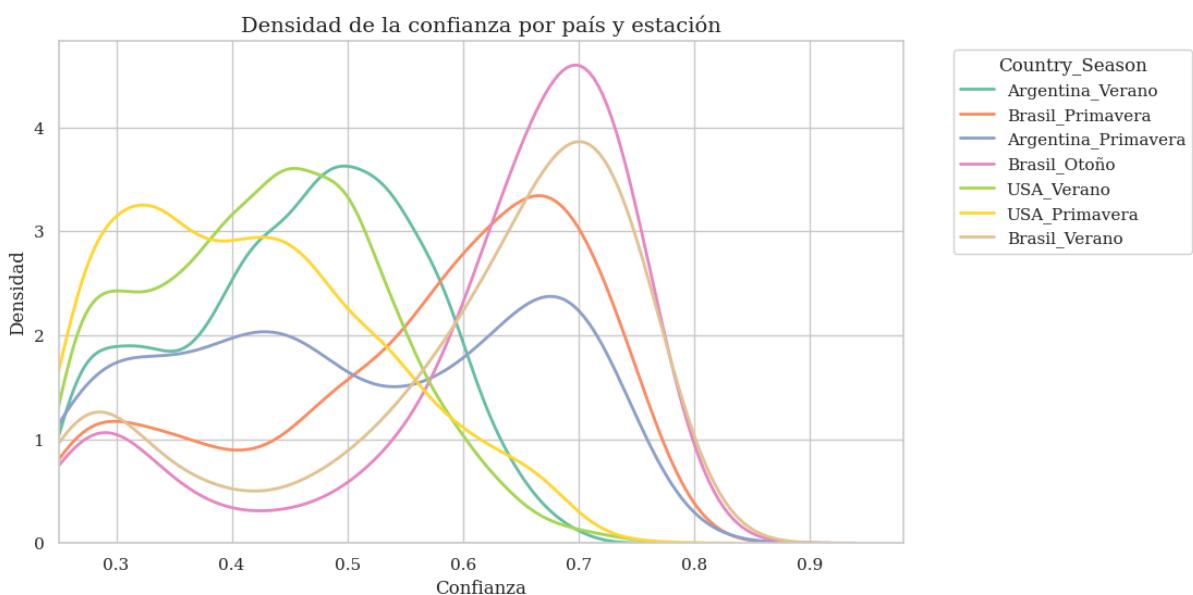


Figura XXXV: Densidad de la confianza del modelo por país y estación

Además se consideraron los puntajes de confianza para las distintas ubicaciones de cada país (véase anexo, Figura a.IV). La confianza del modelo fue significativamente superior en las regiones brasileras en su gran mayoría, mientras que las regiones de Estados Unidos y Argentina ocupan los últimos puestos en función a la media de la confianza estimada.

Por otro lado, otra hipótesis que cuestionaba el funcionamiento del modelo era la existencia de fotos tomadas de noche, en momentos donde la ausencia de luz solar dificulta la identificación de plantas a simple vista. Para evaluar si esto fue determinante para el funcionamiento del modelo se realizó un análisis de resultados para los distintos horarios en los que se tomaron las fotografías -véase Figura XXXVI-.

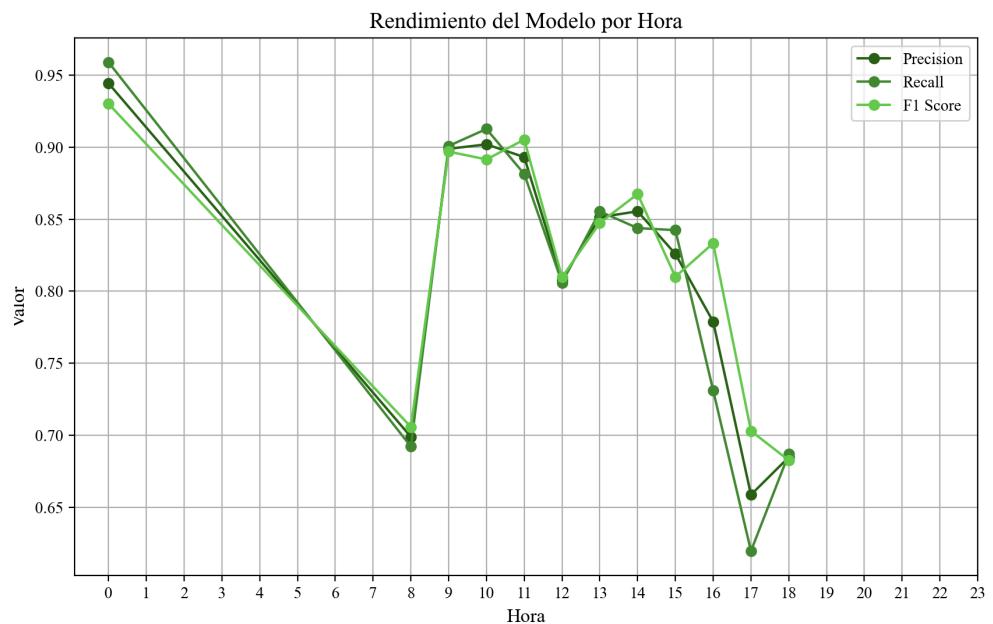


Figura XXXVI: Rendimiento del modelo por hora del dia

La hipótesis no se confirma dado a que, a pesar de que para el ojo humano resulte difícil identificar las plantas durante la noche, el modelo fue capaz de identificarlas, obteniendo hasta mejores resultados que en cualquier otro horario. Una posible explicación para esto es que al seleccionar las imágenes para la validación, se priorizó la inclusión de todas las etiquetas en el conjunto. Esto significó que las imágenes utilizadas para la validación durante el entrenamiento eran aquellas bajo las condiciones más adversas. Como resultado, al seleccionar los mejores pesos para estas imágenes desafiantes, se consiguió que el modelo tuviera un rendimiento sobresaliente en todas las imágenes, incluyendo aquellas tomadas en condiciones no óptimas como la oscuridad.

Por otro lado, en los horarios de las 8, las 17 y las 18 las métricas resultaron ser más bajas que en el resto del conjunto de testeо, por lo que el horario del amanecer o atardecer, cuando la luz es distinta y con perspectiva lateral el funcionamiento del modelo resultó ser peor. Esto puede a su vez verse reflejado en la Tabla a.I. del anexo.

Otra información disponible sobre el dataset son las etiquetas o “tags” marcadas en algunas de las fotos de forma manual cuando se observaba alguno de los fenómenos que se encuentran en la columna significado de la Tabla XIX.

Tag	Significado	F1-Score	Recall	Precision	Confidence score	cantidad_fotos
1	Suelo oscuro	0,83	0,81	0,85	0,43	94
2	Suelo rojo	0,88	0,89	0,88	0,56	188
3	Alta densidad de plantas	0,79	0,78	0,81	0,44	101
4	Presencia de sombras	0,86	0,85	0,87	0,48	256
5	Rastrojo	0,85	0,85	0,85	0,50	301
6	Maleza	0,84	0,84	0,85	0,49	164
7	Imagen borrosa	0,90	0,92	0,89	0,51	84
8	Imagen obscura	0,90	0,93	0,88	0,55	62
9	Imagen clara	0,89	0,88	0,90	0,51	40
10	Plantas crecidas	0,77	0,76	0,79	0,44	95
11	Plantas dobles	0,83	0,83	0,83	0,49	200

Tabla XIX: Métricas obtenidas para cada tag en el conjunto de testeo

En el análisis descriptivo se había considerado la posibilidad de que algunas de estas etiquetas, como plantas dobles o imágenes borrosas, presentaran un desafío. De acuerdo con los resultados obtenidos se evidencia que el modelo tuvo un rendimiento similar al general en plantas dobles y mejor al general en fotos borrosas. A su vez, y así como se ha mencionado previamente, se corrobora que el exceso o ausencia de luminosidad no perjudica en lo absoluto el rendimiento del modelo.

c. Conclusiones de la sección

Durante el proceso de evaluación, se identificaron aspectos significativos que influyeron en el rendimiento del modelo, como las variaciones en el etiquetado manual y el impacto del tamaño de las bounding boxes en la precisión de las detecciones. Además, se exploraron diferentes aspectos del rendimiento del modelo, como su desempeño en diferentes países, estaciones del año y horarios de captura de las imágenes. Se observó que el ancho de las fotos no generó un problema significativo

para el modelo, y se identificaron patrones en el puntaje de confianza del modelo en función de estos factores.

26. Data Augmentation

El concepto de data augmentation en el contexto de procesamiento de imágenes se centra en la generación sistemática de nuevas instancias de datos a partir de conjuntos de datos existentes, mediante la aplicación de transformaciones controladas a las muestras originales. El objetivo fundamental es aumentar la diversidad y la cantidad de datos de entrenamiento disponibles para mejorar la capacidad de generalización y adaptación de los modelos de aprendizaje automático.

En la arquitectura de YOLO hay un proceso automático de Data Augmentation, sin embargo, se realizó la prueba para evaluar si un proceso manual de Data augmentation optimizado para entrenar este modelo específico mejoraría los resultados.

En este caso, estas transformaciones incluyen ajustes de brillo, contraste y saturación de manera aleatoria -vease Figura XXXVII-, así como otras manipulaciones geométricas y de color. Al exponer el modelo a diferentes variantes de las imágenes originales, se busca enriquecer la representación del conjunto de datos de entrenamiento, permitiendo al modelo aprender características más robustas y generalizables. Además, el data augmentation contribuye a mitigar el sobreajuste al proporcionar al modelo una visión más completa y variada del espacio de características presente en los datos de entrada.



Figura XXXVII: Imágenes con valores máximos y mínimos de contraste, brillo y saturación utilizado para data augmentation.

El ajuste de brillo aleatorio modifica el nivel de luminosidad de una imagen al agregar un factor de brillo aleatorio dentro de un rango definido. Este proceso se realiza seleccionando aleatoriamente un factor de ajuste de brillo para cada píxel de la imagen y sumándolo al valor original del píxel. Posteriormente, se utilizan técnicas de limitación de valores para asegurar que los niveles de brillo ajustados se mantengan dentro del rango válido para imágenes en formato uint8. Los valores uint8 van desde 0 hasta 255, lo que permite representar un total de 256 niveles diferentes. Un valor de 0 representa la ausencia de luz o el color más oscuro (por ejemplo, negro en imágenes en escala de grises o ausencia de cada componente de color en imágenes RGB), mientras que un valor de 255 representa la máxima intensidad luminosa o el color más intenso (por ejemplo, blanco en imágenes en escala de grises o el color más brillante en imágenes RGB).

Por otro lado, el ajuste de contraste aleatorio altera la diferencia entre los valores de intensidad de píxeles en una imagen, influyendo en la gama de tonos de grises presentes. Este procedimiento comienza calculando la media de la imagen para determinar el nivel de gris medio.

Luego, se aplica un factor de contraste aleatorio a cada píxel, multiplicando su valor original por este factor y sumando la media calculada. Al igual que con el brillo, se utiliza una técnica de limitación de valores para garantizar que los niveles de contraste ajustados se mantengan dentro de los límites válidos.

La saturación de una imagen se refiere a la intensidad de los colores presentes. El ajuste de saturación aleatorio opera en el espacio de color HSV (matiz, saturación, valor) y consiste en modificar el canal de saturación (S) de la imagen. Después de convertir la imagen al espacio de color HSV, se selecciona aleatoriamente un factor de saturación para cada píxel y se utiliza para ajustar el valor de saturación. Al finalizar la transformación, se aplican técnicas de limitación de valores para asegurar que los niveles de saturación ajustados se encuentren dentro del rango válido.

Se entrenó al modelo con el dataset reducido al que se le aplicó Data Augmentation y se compararon los resultados de entrenar el mismo modelo con el dataset reducido sin este proceso y se concluyó que la aumentación manual de datos no agregó valor, por lo que no se aplicó en el momento de entrenar el modelo para el dataset completo.

	Precision	Recall	F1 - Score
Modelo sin pre procesamiento	0,8002	0,7201	0,7580
Modelo con preprocessamiento	0,7945	0,6607	0,7215

Tabla XX: Comparativa de modelos sin/con data augmentation

27. Posprocesamiento

Analizando las imágenes con bajo performance se encontró que algunas bounding boxes predichas por el modelo contienen maleza, plantas que crecen en lugares donde no se desean.

La imagen que se presenta a continuación -véase Figura XXXVIII- ilustra la capacidad del modelo para distinguir las plantas basándose en su forma. En esta imagen, se observan dos malezas situadas fuera de la línea principal de plantas; una de ellas es incorrectamente identificada como una planta, mientras que la otra no. Este error de identificación puede atribuirse a la similitud en forma entre la maleza detectada y las plantas circundantes. Aunque ambas malezas comparten el color verde característico de las plantas, lo que realmente determina su detección es su forma. Esto subraya la importancia de la forma en la precisión del modelo al diferenciar entre plantas y otros elementos verdes en el entorno. Sin embargo, se detectó que hay casos en los que la maleza puede identificarse por su posición respecto al resto de las plantas.

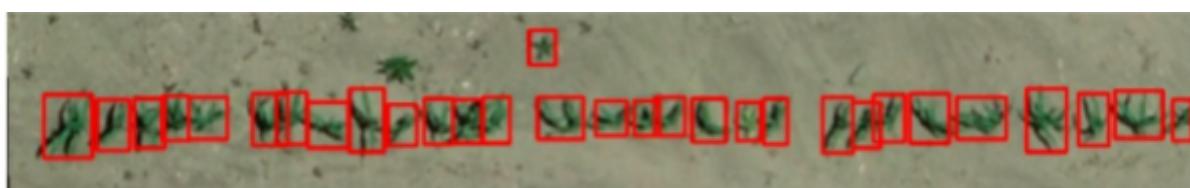


Figura XXXVIII: predicción que detecta maleza como parte de la plantación de maíz

Para poder identificar objetos con posibles anomalías en sus ubicaciones dentro de las imágenes, se realizó una regresión lineal de las coordenadas de los centros de las bounding boxes

detectadas en una imagen. Luego, se calcula la distancia perpendicular de cada punto de centro (x,y) a esta línea de regresión.

Si la distancia perpendicular de un punto al ajuste de la línea excede un umbral predefinido de 50 píxeles, se considera que dicho punto no corresponde a una planta, o es una planta que no se desea que sea incluida en el conteo. Existen únicamente 3 casos en el conjunto de testeo donde existe al menos una predicción que se aleje lo suficiente de la línea de tendencia delimitada. Algunos de esos casos se trata de maleza -así como es posible visualizar en la imagen superior de la Figura XXXIX-, mientras que en otros simplemente se trata de imágenes desalineadas, captando dos líneas centrales simultáneamente o simplemente casos de detecciones cuyos centroides no forman una línea recta.

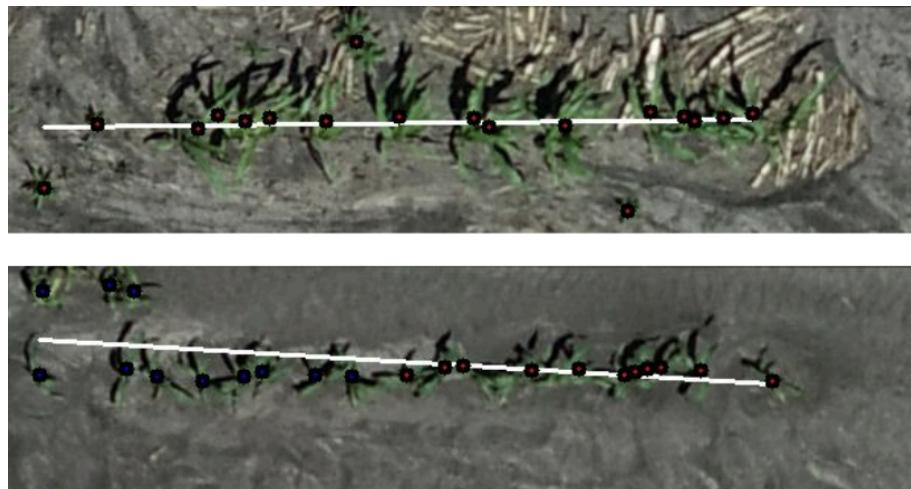


Figura XXXIX: Casos en los que hay predicciones consideradas outliers según regresión lineal

Dado a que los casos en los que se encuentran outliers son pocos y que su condición de planta o maleza pueden variar en cada caso, se optó por imprimir las imágenes con la regresión lineal implementada para que la empresa pueda interpretar y determinar en cada caso deben ser tenidas en cuenta.

28. Implementación

Una vez obtenido el modelo optimizado, se ha integrado en el sistema con el objetivo de hacer accesible y utilizable dicho modelo para la detección de objetos en imágenes nuevas. Este proceso de integración implica la implementación del modelo pre entrenado en el backend del sistema, permitiendo que el mismo sea invocado y utilizado mediante solicitudes HTTP desde el frontend.

El backend, desarrollado sobre Flask, funciona como una interfaz entre el frontend y el modelo de detección de objetos. Cuando un usuario carga una imagen a través del frontend, la API recibe esta solicitud y procesa la imagen. Posteriormente, la imagen es enviada al modelo pre entrenado (en este caso, el modelo YOLOv5) para llevar a cabo la inferencia y realizar la detección de objetos, en este caso, plantas de maíz. Una vez que el modelo completa la inferencia, el backend retorna los resultados de detección al frontend en formato JSON. Esta comunicación entre el frontend y el backend es esencial para el funcionamiento fluido y eficiente del sistema, permitiendo la interacción del usuario con la aplicación de manera transparente y receptiva.

Por otro lado, el frontend, implementado con tecnologías web estándar como HTML, CSS y JavaScript, ofrece una experiencia de usuario interactiva. Permite a los usuarios cargar imágenes desde su dispositivo y visualizar los resultados de la detección de objetos de forma intuitiva. Cuando se inicia el proceso de detección de objetos al enviar una imagen al backend, el frontend muestra indicadores visuales del progreso de la inferencia, como mensajes de carga o barras de progreso. Una vez que se obtienen los resultados del backend, el frontend actualiza dinámicamente la página web para mostrar la imagen original junto con las plantas de maíz identificadas, resaltadas con bounding boxes o marcos delimitadores.

El backend actúa como un puente entre la interfaz de usuario y el modelo de aprendizaje profundo, mientras que el frontend proporciona una experiencia visualmente atractiva y funcional para interactuar con la detección de objetos en tiempo real. Dado que la implementación no se ha subido a un servidor, en el siguiente [video](#) se muestra como es su funcionamiento, ejecutado localmente.

29. Impacto de los resultados en el business case

Tal como se mencionó en la Sección I apartado 5.d *Costos y Escenarios*, los resultados del modelo pertenecen al escenario moderado por lo que, según los estándares de la empresa, deberá realizarse una verificación manual en el 25% de las imágenes que haya recibido menor puntaje de confianza.

A continuación se presentan los impactos económicos del modelo en Eiwa. En el año 2024, los ingresos proyectados para la empresa, considerando el porcentaje de verificación manual, se estiman con un cargo de \$500 pesos argentinos por parcela al cliente. Se anticipa una demanda de 300.000 parcelas para dicho año. El costo de la verificación manual posterior a la ejecución del modelo es equivalente al costo actual por parcela del conteo manual. Siendo:

- $Ingreso\ potencial\ anual = 300.000 * \$500 - (300.000 * (\text{costo de conteo automático por parcela} + \text{costo de conteo manual} * \text{porcentaje de parcelas que deberán contarse manualmente}))$

En donde el *costo de conteo automático por parcela* es \$0,92 pesos argentinos utilizando la instancia EC2:p3 y de \$1,16 pesos argentinos usando la instancia EC2:g5. Por otro lado el *costo de conteo manual* es de \$92,72 pesos argentinos por parcela.

- $\text{Variación porcentual de ingresos} = \frac{\text{ingreso potencial del escenario}}{\text{ingreso potencial escenario actual (manual)}} - 1$
- $\text{Horas que se ahorran respecto al conteo manual} = \frac{\frac{\text{tiempo de conteo manual en segundos por parcela}*300.000}{60*60}}{\frac{\text{tiempo de conteo del escenario resultante en segundos por parcela}*300.000}{60*60}}$

Resulta de crucial importancia la evaluación de distintos escenarios y su escalabilidad correspondiente -véase Tabla XXI-. Al comparar los tres escenarios, se observa que tanto EC2:p3 como EC2:g5 ofrecen mejoras significativas en términos de eficiencia y costos en comparación con el método manual. Ambos escenarios reducen el tiempo de conteo por parcela a aproximadamente una

cuarta parte del tiempo requerido en el método actual, y disminuyen los costos a aproximadamente una cuarta parte del costo actual. Aunque las diferencias entre los escenarios EC2:p3 y EC2:g5 son mínimas, el EC2 resulta ligeramente más eficiente tanto en términos de costo como de tiempo de conteo. Además, ambos escenarios proporcionan un incremento significativo en los ingresos potenciales anuales, superando en más del 16% los ingresos del método manual. En términos de horas ahorradas, ambos escenarios propuestos ofrecen ahorros sustanciales de tiempo, lo que puede traducirse en una mayor productividad y una utilización más eficiente de los recursos disponibles.

Escenario	Tiempo de conteo en segundos (por parcela)	Costo de conteo en pesos (por parcela)	Ingreso potencial anual	Aumento porcentual de ingresos respecto al escenario actual	Demora total en el conteo hs (considerando el conteo manual)	Horas ahorrradas respecto al escenario actual
Actual (manual)	48	92,72	\$ 122.184.000	N/A	4.000	N/A
EC2:p3	12,81	24,1	\$ 142.770.000	16,85%	1.067,50	2.932,5
EC2:g5	12,89	24,34	\$ 142.698.000	16,79%	1.074,16	2.925,83

Tabla XXI: Costos de distintos escenarios de procesamiento considerando el 25% de conteo manual que exige la empresa para un escenario moderado.

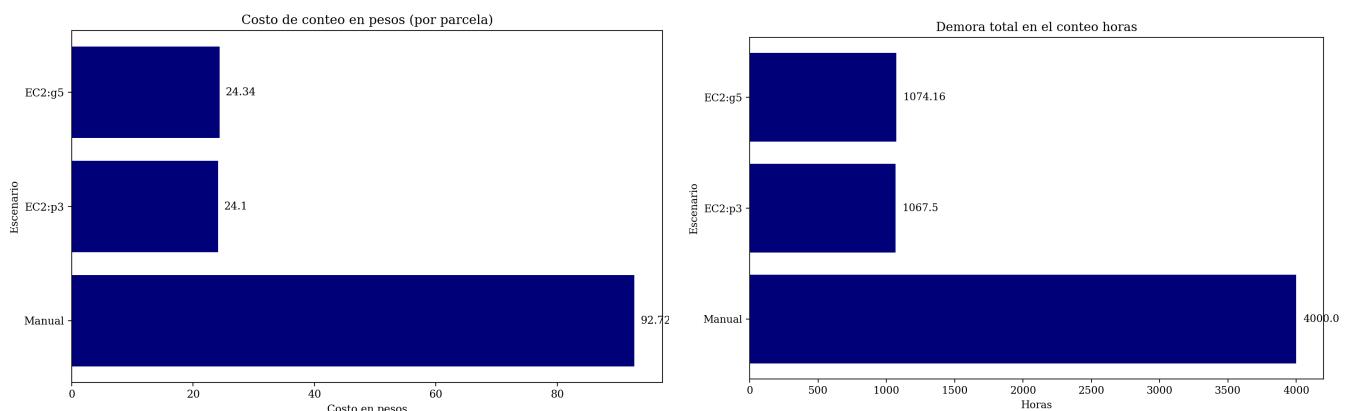


Figura XL: Costo y demora de los distintos escenarios con un 25% de conteo manual post modelo predictivo

Posteriormente, se realiza la comparación sin considerar el 25% de conteo manual exigido por la empresa para un escenario moderado -véase Tabla XXII-. La comparación de estos dos conjuntos de datos revela que la eliminación del 25% de verificación manual resulta en mejoras significativas en términos de costos, ingresos y eficiencia temporal. Mientras que la verificación manual puede ser una medida de control de calidad, su eliminación conduce a una optimización considerable de los recursos y maximización de los ingresos potenciales, rondando el 6% en promedio su mejora en la variación del ingreso potencial respecto al escenario actual. La empresa puede optar por descartar la verificación manual, asumiendo un mayor margen de error, si así lo desea.

Escenario	Tiempo de conteo en segundos (por parcela)	Costo de conteo en pesos (por parcela)	Ingreso potencial anual total	Variación del ingreso potencial respecto al escenario actual	Demora en el conteo en hs	Horas que se ahorran respecto al escenario actual
Actual (manual)	48	92,72	\$ 122.184.000	N/A	4.000	N/A
EC2:p3	1,08	0,918	\$ 149.724.600	22,54%	90	3.910
EC2:g5	1,188	1,155	\$ 149.653.500	22,48%	99	3.901

Tabla XXII: Costos de distintos escenarios de procesamiento sin considerar el 25% de conteo manual que exige la empresa para un escenario moderado

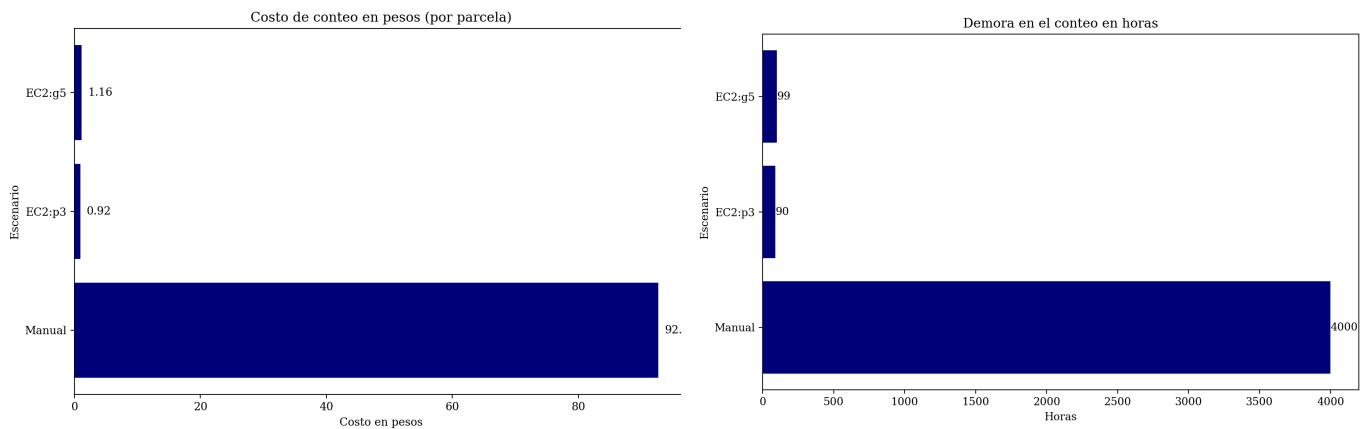


Figura XLI: Costo y demora de los distintos escenarios sin conteo manual post modelo predictivo.

30. Conclusion

El presente análisis ofrece una mirada en la búsqueda de un modelo óptimo para la detección de plantas de maíz (incluyendo su respectivo proceso de de limpieza y preprocesamiento como así también pos procesamiento). A su vez, se analiza cómo el mismo responde a condiciones ambientales variables y utiliza información contextual para mejorar sus predicciones. En primer lugar, se observa que el modelo muestra una notable adaptabilidad a condiciones adversas, como baja luminosidad o visibilidad reducida debido a la borrosidad de las imágenes. Esto sugiere que ha aprendido a generalizar características clave más allá de las condiciones ideales de captura, lo cual es esencial para su aplicación en entornos prácticos donde las condiciones pueden ser impredecibles.

Además, la capacidad del modelo para identificar plantas basándose en características específicas como la forma y la presencia de sombras indica un nivel de interpretación avanzada de características visuales. Esto sugiere que el modelo ha desarrollado la capacidad de discernir entre objetos verdes basándose en criterios más complejos que simplemente el color, lo cual es crucial para una detección precisa en entornos naturales.

El análisis estratégico realizado durante este estudio, incluyendo ajustes como la reducción del umbral de Intersection over Union y el uso del puntaje de confianza para filtrar predicciones, demuestra un enfoque adaptativo para mejorar el rendimiento del modelo en función de los desafíos específicos identificados durante la evaluación, resultando en una performance satisfactoria del modelo.

A pesar de obtener resultados moderados en términos de métricas estándar como precisión, recall y F1-Score, el modelo muestra una consistencia notable en diferentes contextos, incluyendo países diversos, estaciones del año y horarios de captura de imágenes. Esto refleja una robustez y capacidad de generalización que son esenciales para aplicaciones del mundo real donde las condiciones de captura pueden variar significativamente. Por otro lado, la performance del modelo impacta en un aumento en las ganancias de la empresa, habiendo un aumento porcentual en las mismas superior al 16% en caso de realizar un verificación manual en un 25% de las imágenes.

Alternativamente, podría haber un aumento porcentual de las ganancias superior al 22% en caso de que la empresa decida no hacer una verificación manual adicional para un cuarto de la totalidad de las imágenes. A su vez, el aumento en la eficiencia y rapidez permite el ahorro de al menos 2.900 horas de verificadores en ambos escenarios, lo cual permite la alocación de ese recurso en proyectos alternativos.

Feedback del cliente

El proyecto desarrollado en colaboración con Eiwa ha recibido una valoración altamente positiva por parte del cliente. A continuación, se transcribe el feedback proporcionado por Leonardo Maestri, líder del área de Data Science de Eiwa, quien realizó seguimientos del proyecto a lo largo de su duración:

"La experiencia con las chicas fue muy positiva para Eiwa. Tuvimos la oportunidad de probar un nuevo modelo para resolver el problema de conteo y analizar el problema desde otro punto de vista. La facilidad de las chicas para aprender y acelerar el proceso hizo que tengamos un prototipo útil muy rápido, cuando con ese tipo de modelos podemos tardar meses en tenerlos listos."

"El modelo salió tan bien que lo reemplazamos por el actual método de conteo para resolver varias familias de imágenes que abarcaba. Vamos a seguir iterando este nuevo modelo para que abarque todos nuestros casos de conteo y sabemos que tiene el potencial de ser suficientemente robusto y durar en el tiempo."

"La experiencia fue muy enriquecedora para nosotros y anhelamos volver a trabajar con estudiantes de la facultad en el futuro. Esperamos haber aportado a las chicas nuevos conocimientos y experiencia."

El éxito del modelo fue significativo tal que Eiwa optó por sustituir su método de conteo vigente por el nuevo modelo desarrollado. Esto confirma la utilidad práctica del modelo y su capacidad para abordar eficazmente diversos escenarios de conteo de imágenes. Además, se observa un potencial prometedor en términos de durabilidad y solidez del modelo a largo plazo, lo que asegura su pertinencia continua y su efectividad en el tiempo.

Asimismo, se destaca que la experiencia resultó enriquecedora tanto para Eiwa como para los estudiantes participantes. Esta colaboración entre la academia y la industria resalta la importancia y los beneficios de tales asociaciones.

Líneas de Investigación Futura

A lo largo del proyecto se enfrentaron numerosos desafíos que moldearon el resultado final. Inicialmente, la calidad de los datos presentó un reto considerable, ya que las imágenes tenían variabilidad en términos de calidad del etiquetado manual. Esto requirió una limpieza exhaustiva de las etiquetas para asegurar la precisión del modelo.

Durante la exploración de modelos, se emplearon distintas técnicas y metodologías como YOLOv5 para optimizar la detección y conteo de plantas, ajustando hiperparámetros y realizando múltiples iteraciones para mejorar la precisión del modelo. Los resultados fueron validados mediante métricas como IoU, recall, precisión y F1-Score.

Se realizaron análisis visuales cualitativos para asegurar la robustez del modelo frente a condiciones diversas. Finalmente, la implementación en un entorno de producción incluyó la integración con sistemas backend y frontend, asegurando una experiencia de usuario fluida y eficiente. Estos esfuerzos culminaron en un modelo que alcanzó un desempeño notable con una precisión del 0,8414, un recall del 0,8301 y un F1-Score de 0.8357. Estos resultados colocaron al modelo en el escenario conservador, donde la variación del ingreso porcentual anual se proyecta entre un 16.79% y un 16.85%, dependiendo de la instancia de AWS utilizada. Además, el tiempo ahorrado es del 73.3% en comparación con el cotejo manual. No solo mejoró significativamente la eficiencia y reducción de costos, sino que también recibió una valoración positiva del cliente, demostrando su aplicabilidad práctica y potencial de durabilidad.

Si bien la mayoría de los desafíos fueron superados y concluyeron en un modelo satisfactorio, se pueden definir próximos pasos a realizar que no se han podido implementar actualmente por falta de tiempo o recursos de computación. A continuación, se detallan futuras mejoras y próximos pasos:

1. Calidad de datos: Se propone realizar una capacitación a los verificadores para enseñarles cómo poner las cajas respetando los tamaños de las plantas. Además, las imágenes que fueron eliminadas por falta de etiquetas de bounding boxes podrían ser modificadas en vez de ser eliminadas, corrigiendo las etiquetas manualmente para preservar la información útil.
2. Optimización de hiperparámetros: Se propone llevar a cabo una búsqueda más exhaustiva de hiperparámetros utilizando técnicas avanzadas como grid search o la búsqueda bayesiana (Bayesian optimization) para identificar las combinaciones óptimas que mejoren aún más el rendimiento del modelo. Esto puede incluir ajustes más finos en los parámetros de aprendizaje, regularización y arquitectura del modelo.
3. Entrenamiento con nuevas imágenes: Se sugiere expandir el dataset incluyendo nuevas imágenes capturadas en diferentes condiciones climáticas, estaciones del año y regiones geográficas adicionales. Esto permitirá mejorar la generalización del modelo y su robustez frente a variaciones en los datos de entrada, asegurando un desempeño consistente en diversos escenarios y ampliando su aplicabilidad a una mayor variedad de contextos agrícolas.
4. Nuevos Modelos: Se identificó la posibilidad de crear modelos específicos para cada país. Se tiene la hipótesis de que con un modelo específico por país, la detección mejoraría significativamente. La justificación para esta hipótesis radica en las variaciones climáticas, tipos de suelo y prácticas agrícolas que existen entre diferentes países. Estas variaciones influyen en las características de las imágenes, por lo que un modelo ajustado a las condiciones locales tendría una mayor precisión en la identificación y conteo de las plantas.

Estas propuestas tienen el potencial de incrementar aún más la precisión y eficiencia del modelo, asegurando su relevancia y eficacia en entornos de producción reales.

Bibliografía

- D., D. (n.d.). Basics of bounding boxes. Medium. Retrieved from <https://medium.com/analytics-vidhya/basics-of-bounding-boxes-94e583b5e16c>
- Evidently AI. (s. f.). Cómo usar el umbral de clasificación para equilibrar la precisión y la exhaustividad [Página web]. Recuperado de <https://www.evidentlyai.com/classification-metrics/classification-threshold>
- Tøndering, C. (n.d.). Color models and color spaces. Programming Design Systems. Recuperado de <https://programmingdesignsystems.com/color/color-models-and-color-spaces/index.html>
- Pardo, C. J. (2018, April 22). Pirámides de imágenes con Python y OpenCV. Recuperado de <https://carlosjuliopardoblog.wordpress.com/2018/04/22/piramides-de-imagenes-con-python-y-opencv/>
- Mallick, S. (n.d.). Fully convolutional image classification on arbitrary sized image. LearnOpenCV. Recuperado de <https://learnopencv.com/fully-convolutional-image-classification-on-arbitrary-sized-image/>
- Moeller, T., Katija, K., Sherlock, R. E., & Robison, B. H. (2023). Demystifying image-based machine learning: a practical guide to automated analysis of field imagery using modern machine learning tools. *Frontiers in Marine Science*, 10. Recuperado de <https://www.frontiersin.org/articles/10.3389/fmars.2023.1157370/full>
- Fatih Cagatay Akyon. (2021, January 30). SAHI: A vision library for large-scale object detection & instance segmentation. Codable. <https://codablemag.com/2021/01/30/sahi-a-vision-library-for-large-scale-object-detection-instance-segmentation/>
- Kumar, A. (2021). Apply Data Augmentation on YOLOv5/YOLOv8 Dataset. Medium. Recuperado de <https://medium.com/red-buffer/apply-data-augmentation-on-yolov5-yolov8-dataset-958e89d4bc5d>
- YOLO (You Only Look Once): Detección de Objetos en Tiempo Real. <https://pythondiario.com/2018/09/yolo-you-only-look-once-detectio.html>.
- YOLOv5 - Ultralytics YOLOv8 Documentos. <https://docs.ultralytics.com/es/models/yolov5/>
- Ultralytics. (2024). *Configuration - Ultralytics YOLO Docs*. Recuperado de <https://docs.ultralytics.com/es/usage/cfg/#export-settings>
- Li, Y., Du, H., Wang, X., Wang, L., Zhang, Y., & Zhao, H. (2018). A review on YOLO object detection. *Journal of Signal and Information Processing*, 9(4), 139-147. <https://www.scirp.org/journal/paperinformation?paperid=88545>
- Mu Zhu (26 de agosto, 2004) “*Recall, Precision and Average Precision*”. University of Waterloo. Department of Statistics & Actuarial Science.. <https://datascience-intro.github.io/1MS041-2022/Files/AveragePrecision.pdf>

Anexo

Relación entre cantidad de bounding boxes y width de la imagen

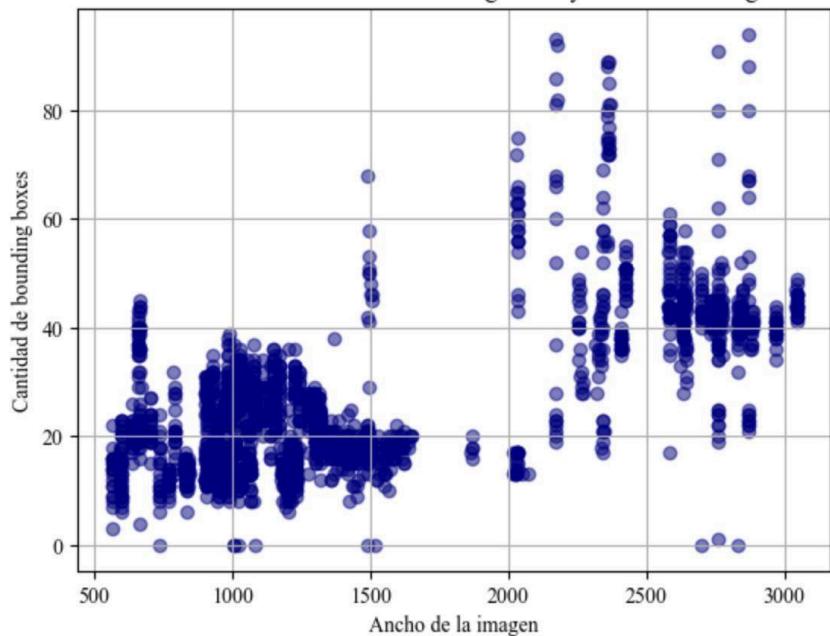
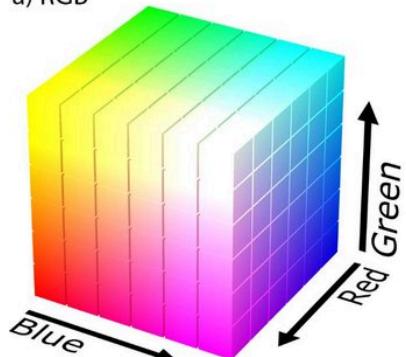


Figura a.I: Scatterplot de la cantidad de bounding boxes y ancho de la imagen

a) RGB



b) HSV

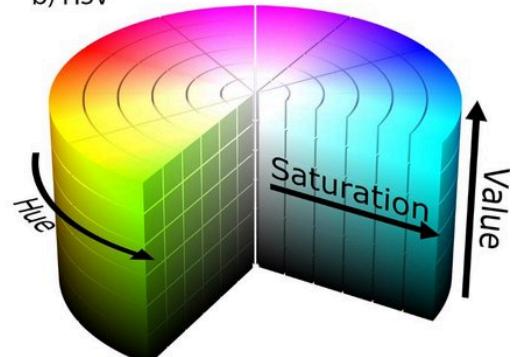


Figura a.II: Distinción entre espacios de color RGB y HSV

Porcentaje de casos en los que al menos una bounding box excede los límites de la foto

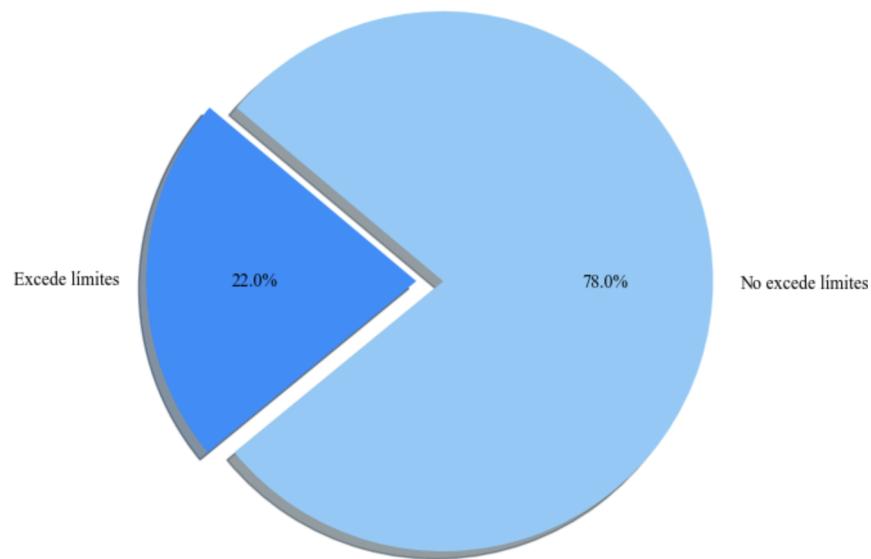


Figura a.III: Porcentaje de casos en los que al menos una bounding box excede los límites de la imagen

Hora de vuelo	F1-Score	recall	precision	Confidence score	Cantidad de fotos
0	0.94	0.96	0.93	0.57	4
8	0.70	0.69	0.71	0.60	3
9	0.90	0.90	0.90	0.48	58
10	0.90	0.91	0.89	0.50	45
11	0.89	0.88	0.91	0.53	40
12	0.81	0.81	0.81	0.47	78
13	0.85	0.86	0.85	0.45	26
14	0.86	0.84	0.87	0.52	56
15	0.83	0.84	0.81	0.49	41
16	0.78	0.73	0.83	0.48	27
17	0.66	0.62	0.70	0.45	10
18	0.68	0.69	0.68	0.39	8

Tabla a.I: Métricas obtenidas para cada horario de vuelo en el conjunto de testeo.

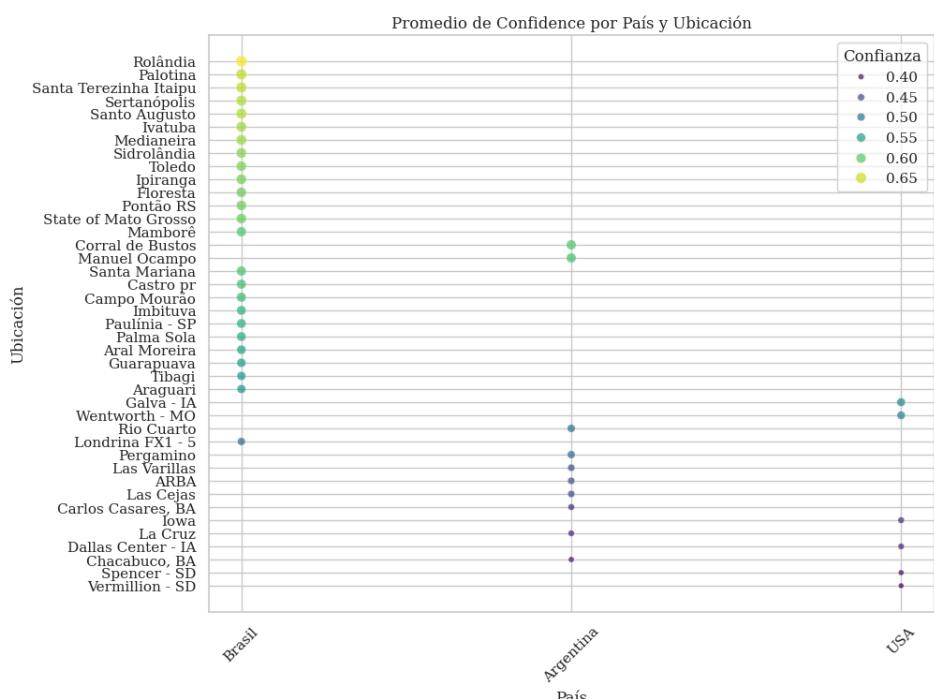


Figura a.IV: Promedio de confianza por país y región